Doctoral Thesis

Chromatin remodeling during mouse spermatogenesis: functional and evolutionary implications

Covadonga Vara González

July 2020

Institut de Biotecnologia i Biomedicina (IBB)
Departament de Biologia Cel·lular, Fisiologia i Immunologia
Universitat Autònoma de Barcelona

Colophon

This document was typeset with the help of KOMA‑Script and LaTeX using the kaobook class.

The source code of this book is available at:
https://github.com/fmarotta/kaobook

Chromatin remodeling during mouse spermatogenesis: functional and evolutionary implications

A thesis presented for the degree of

Doctor of Philosophy in Cell Biology

PhD candidate :

Covadonga Vara González

Supervisor:

Dra. Aurora Ruiz-Herrera Moreno

It's not the size of the woman in the fight, it's the size of the fight inside the woman

— Rachael Newsham, adapted from Mark Twain

Chromosome, broken, awaits sounds of strands pairing, preserving life's thread

— Jim Haber

Funding

Abstract

The germline holds the past and the future of a species, as parental genetic information is recombined through meiosis and transmitted to the offspring. Thus, understanding how the genome is organized and regulated in the nuclear space during the formation of germ cells is essential to comprehend the bases of fertility and its impact on genetic diversity. In the last twenty years, studies on the higher-order structure of the genome in somatic cells have revealed a hierarchical organization: chromosome territories are formed by chromatin compartments, which in their turn are folded into topological associated domains (TADs) and DNA loops. However, little is known about how the genome is organized in the germline and how chromosomal reorganizations modulate genome architecture.

In this context, this thesis aims to: (i) understand the three-dimensional organization of the genome during mouse spermatogenesis and its interplay with gene function and occupancy of insulator proteins (CTCF and cohesins), (ii) investigate the implications of Robertsonian (Rb) fusions in genome folding and meiotic recombination, and (iii) characterize the variability of PRDM9 (which determines recombination hotspots in most mammals) in natural house mouse populations with Rb fusions: the Madeira Rb system and the Barcelona Rb system (BRbS). We applied a chromosomics approach, combining cytological analysis with next generation sequencing technologies including Hi-C, ChIP-seq, RNA-seq and SNP genotyping. For that, we first developed an efficient cell sorting protocol to obtain enriched cell fractions from mouse testis including spermatogonia, primary spermatocytes at early (leptonema/zygonema) and late (pachynema/diplonema) prophase I, round spermatids and sperm.

Our results revealed that the higher-order structure of the genome is extremely dynamic during spermatogenesis, where spermatogonia presents somatic-like compartments and TADs, that disappear during meiosis to be re-established later on in post-meiotic cells. Moreover, transcription correlates with A compartments throughout spermatogenesis, with cell-specific active genes involved in spermatogenesis progression, fertilization and embryonic development. In addition, we found a correlation between cohesin occupancy and active transcription in both meiotic and post-meiotic cells. Our results suggest that in addition to their canonical structural role mediating chromatin cohesion, meiotic cohesins might also play a transcription-regulating role.

Although germ cells with Rb fusions presented the main features of genome architecture (lack of compartments and TADs in primary spermatocytes and a re-establishment of these structures in round spermatids), Rb fusions reorganize the spatial chromosome occupancy. In primary spermatocytes, Rb fusions increase heterologous interactions, promoting the

formation of novel regulatory environments. In round spermatids, Rb fusions reduce inter-chromosomal interactions as a result of mechanistic constrains. The cytological data shows that the increase in heterologous interactions is concomitant with the presence of asynapsed heterozygous metacentrics, which induce the full heterochromatinization of the sex body. Furthermore, Rb fusions affect the number and chromosomal distribution of crossovers in primary spermatocytes, especially in the case of fused metacentrics in homozygosis. The reduction in recombination was also observed in the analysis of linkage disequilibrium based on SNP genotyping, which translated into high levels of genetic divergence in Rb populations when compared to standard mice. In addition, our characterization of PRDM9 variability detected an unprecedented variability in natural house mouse populations, being especially high (54 different alleles) in the insular Madeira Rb system when compared to the continental BRbS (13 alleles). Such differences could be attributed to the combination of different factors: (i) the evolutionary history of each Rb system, (ii) the prevalence of Rb fusions affecting genetic diversity, and to a lesser extent (iii) meiotic functional constrains (i.e., recombination hotspot asymmetry).

Taken together, this thesis shows that chromatin undergoes profound remodeling during spermatogenesis in a cell-specific way, where transcriptional activity correlates with the chromatin state and cohesin occupancy. In addition, Rb fusions alter genome organization in the germline, having an impact on meiotic recombination and genetic diversity.

Publications

During the development of this thesis, I have contributed as a co-author in the following publications. Articles 1, 2 and 3 are an integral part of this thesis, corresponding to sections 4.2, 4.3 and 4.4, respectively. Articles 4 and 5 are the result of collaborations with other members of our research group.

1 **Vara, C.***, Paytuví-Gallart, A.*, Cuartero, Y.*, Le Dily, F., Garcia, F., Salvà-Castro, J., Gómez-H, L., Julià, E., Moutinho, C., Aiese Cigliano, R., Sanseverino, W., Fornas, O., Pendás, A. M., Heyn, H., Waters, P. D., Marti-Renom, M. A., & Ruiz-Herrera, A. (2019). Three-dimensional genomic structure and cohesin occupancy correlate with transcriptional activity during spermatogenesis. Cell Reports, 28(2), 352-367.e9. https://doi.org/10.1016/j.celrep.2019.06.037

2 **Vara, C.**, Paytuví-Gallart, A., Cuartero, Y., Álvarez-González, L., Marín-Gual, L., Garcia, F., Florit-Sabater, B., Capilla, L., Sánchez-Guillén, R. A., Sarrate, Z., Aiese Cigliano, R., Sanseverino, W., Ventura, J., Marti-Renom, M. A., Le Dily, F., & Ruiz-Herrera, A. The impact of chromosomal fusions on 3D genome folding and recombination in the germ line (*under revision*).

3 **Vara, C.**, Capilla, L., Ferretti, L., Ledda, A., Sánchez-Guillén, R. A., Gabriel, S. I., Albert-Lizandra, G., Florit-Sabater, B., Bello-Rodríguez, J., Ventura, J., Searle, J. B., Mathias, M. L., & Ruiz-Herrera, A. (2019). PRDM9 diversity at fine geographical scale reveals contrasting evolutionary patterns and functional constraints in natural populations of house mice. Molecular Biology and Evolution, 36(8), 1686–1700. https://doi.org/10.1093/molbev/msz091

4 Cappelletti, E., Piras, F. M., Badiale, C., Bambi, M., Santagostino, M., **Vara, C.**, Masterson, T. A., Sullivan, K. F., Nergadze, S. G., Ruiz-Herrera, A., & Giulotto, E. (2019). CENP-A binding domains and recombination patterns in horse spermatocytes. Scientific Reports, 9(1), 15800. https://doi.org/10.1038/s41598-019-52153-1

5 Ruiz-Herrera, A., Vozdova, M., Fernández, J., Sebestova, H., Capilla, L., Frohlich, J., **Vara, C.**, Hernández-Marsal, A., Sipek, J., Robinson, T. J., & Rubes, J. (2017). Recombination correlates with synaptonemal complex length and chromatin loop size in bovids—insights into mammalian meiotic chromosomal organization. Chromosoma, 126(5), 615–631. https://doi.org/10.1007/s00412-016-0624-3

*The authors contributed equally to the work.

Main contributions

1 Here we describe the dynamics of the three-dimensional genome organization during mouse spermatogesis and its relation to gene expression and cohesin occupancy. I developed the FACS protocol required to isolate enriched germ cell populations. This included animal handling, cell culture, cell sorting and immunofluorescence analysis. I also performed the Hi-C and ChIP-seq experimental procedures in collaboration with Yasmina Cuartero. I conducted the 3D-FISH experimental work and the ChIP-seq data analysis. I was also involved in the discussion and development of the Hi-C data analysis along with Andreu Paytuví.

2 In this article we studied the impact of Rb fusions in the three-dimensional genome architecture of germ cells using a wild population of house mouse. I was involved in field work and animal processing; I performed the cytological experiments along with its analysis, as well as FACS and the Hi-C experiments. I also analyzed the data derived from the SNPs array, including population structure, F_{ST} and recombination rates. I was also involved in the discussion and development of the Hi-C data analysis, which was mainly conducted by Lucía Álvarez and Andreu Paytuví.

3 In this article we studied the allelic variability of *Prdm9* in two wild Robertsonian (Rb) mouse systems, the Barcelona and Madeira Rb systems. I was involved in field work and animal processing from the Barcelona Rb system. I also conducted the genotyping, including the single molecule PCR methodology. I analyzed *Prdm9* sequences for all genotyped animals, determining the variation in ZnF, alleles and heterozygous allele combinations. Dr. Lucca Ferretti developed the phylogeographic analysis.

4 This manuscript is focused on the study of recombination patterns and centromere binding domain in horse spermatocytes. I contributed to the optimization of the MLH1 immuno-detection in horse spermatocyte spreads and was involved in the cytological experiments.

5 In this article we studied the correlation between the synaptonemal complex length and chromatin loop size in bovids. I collaborated in the cytological experiments.

Acronyms and Abbreviations

3C Chromosome Conformation Capture

3D Three-dimensional

4C Chromosome Conformation Capture on-Chip

5C Chromosome Conformation Capture Carbon Copy

5hmC 5-Hydroxymethylcytosine

aa amino acid

AcNa sodium acetate

APC Allophycocyanin

APS Ammonium Persulfate

ATM Ataxia Telangiectasia Mutated protein

ATR Ataxia Telangiectasia and Rad3 related protein

ME mercaptoethanol

BAC Bacterial Artificial Chromosome

BB Binding Buffer

BLAST Basic Local Alignment Search Tool

bp base pair

BRbS Barcelona Robertsonian System

BSA Bovine Serum Albumin

BTB Blood-Testis Barrier

C2H2 Cys2-His2

CD90.2 Cluster of Differentiation 90.2

CDYL Chromodomain Y Like protein

CEEAH Ethics Committee on Animal and Human experimentation

ChIP-seq Chromatin Immunoprecipitation with sequencing

ChIA-PET Chromatin Interaction Analysis with Paired-End Tag sequencing

COs Crossovers

CT Chromosomal Territory

CTCF CCCTC-Binding factor

CREST Complication of Raynaud's phenomenon, Esophageal dysfunction, Sclerodactily and Telangiectasia

Cryo-EM Cryogenic Electron Microscopy

CXXC1 CXXC-type zinc finger protein 1

Cy3 Cyanine 3

Cy3 Cyanine 5

Da Daltons

DAPI 4 ',6-diamino-2-fenilindol

DDR DNA Damage Response

DEG Differentially Expressed Genes

DIG Digoxigenin

DMC1 Meiotic recombination protein DMC1/LIM15 homolog

DMEM	Dulbecco's Modified Eagle Medium	**GnRH**	Gonadotropin-Releasing Hormone
DMSO	Dimethyl Sulfoxide	**GOEA**	Gene Ontology Enrichment Analysis
DNA	Deoxyribonucleic Acid	**h**	hour
dNTP	Deoxynucleotide Triphosphate	**H2AX**	Phosphorylation of Histone H2AX on serine 139
DPBS	Dulbecco's Phosphate Buffered Saline	**H2A**	Histone 2A
dpc	days post-coitum	**H3K9Ac**	Histone 3 acetylated at lysine 9
dpp	days post-partum	**H3K4Me3**	Histone 3 trimethylated at lysine 4
DSBs	Double Strand Breaks	**H3K9Me3**	Histone 3 trimethylated at lysine 9
dTTP	Deoxythymidine Triphosphate		
dUTP	Deoxyuridine Triphosphate	**H3K27Me1/3**	Histone 3 (tri)methylated at lysine 27
EDTA	Ethylenediamine-Tetraacetic Acid	**H4K16ac**	Histone 4 acetylated at lysine 16
EHMT2	Euchromatic Histone Lysine Methyltransferase 2	**HCl**	Hydrochloric acid
EtOH	Ethanol	**HELLS**	Lymphoid-specific helicase
EWSR1	EWS RNA Binding Protein 1	**HP1**	Heterochromatin Protein 1
ES	Elongated Spermatids	**Ho**	Hoechst 33342
FACS	Fluorescence-Activated Cell Sorting	**HORMAD1**	HORMA domain-containing protein 1
FBS	Fetal Bovine Serum	**Hst1**	Hybrid sterility locus 1
FISH	Fluorescence In Situ hybridization	**ICE**	Iterative Correction and Eigenvector decomposition
FITC	Fluorescein Isothiocianate	**ID**	Identifier
FPKM	Fragments Per Kilobase of transcript per Million mapped reads	**ID4**	Inhibitor Of DNA Binding 4
		IF	Immunofluorescence
FSC	Forward-Scattered light	**IHO1**	Interactor or Hormad protein 1
FSH	Follicle-Stimulating Hormone	**IP**	Immunoprecipitation
GBSS	Gey's Balanced Salt Solution	**kb**	kilobase pair
GFRA1	GDNF Family Receptor Alpha 1		

KCl	Potassium chloride	**P/D**	Pachynema/Diplonema
Kda	kilodalton	**PAA**	Polyacrylamide
KRAB	Krueppel-Associated Box	**PAR**	Pseudoautosomal Region
L/Z	Leptonema/Zygonema	**PAS**	Periodic Acid-Schiff
LADs	Lamina-Associated Domains	**PAX7**	Paired box protein 7
LD	Linkage Disequilibrium	**PBS**	Phosphate-Buffered Saline
LH	Luteinizing Hormone	**PBST**	Phosphate-Buffered Saline with Tween-20
lncRNA	long non-coding RNA	**PCA**	Principal Component Analysis
M	million	**PCR**	Polymerase Chain Reaction
MACS	Magnetic-Activated Cell Sorting	**PFA**	Paraformaldehyde
Mb	Megabase pair	**PGCs**	Primordial Germ Cells
MCD	MEI4-containing DSB-promoting	**PI**	Propidium Iodide
MEI4	Meiotic Double-Stranded Break Formation Protein 4	**piRNA**	piwi-interacting RNA
min	minute(s)	**PMSC**	Post-Meiotic Sex Chromatin
MLH1/3/4	MuTL Homolog 1/3/4	**POP**	Population
mRNA	messenger RNA	**PRDM9**	PR domain containing protein 9
MSCI	Meiotic Sex Chromosome Inactivation	**PDS5**	Sister chromatid cohesion protein pds5
MSH4	MutS Homolog 4	**PVDF**	Polyvinylidene fluoride
MSUC	Meiotic Silencing of Unsynapsed Chromatin	**PWD**	PWD/PhJ mouse strain
NATs	Natural Antisense Transcripts	**RAD21**	RAD21 cohesin complex component
NCOs	Non Crossovers	**RAD21L**	RAD21-like protein
NDR	Nucleosome Depleted Region	**RAD51**	Radiation sensitive protein 51
NGS	Next Generation Sequencing	**Rb**	Robertsonian
NIPBL	Nipped-B-like protein	**REC8/114**	Recombinant protein 8/114
NT	Nick Translation	**RIPA**	Radioimmunoprecipitation Assay
O/N	overnight	**RPA**	Replication Protein A

RNA Ribonucleic Acid

RNAPII Ribonucleic Acid Polymerase II

RNA-seq RNA sequencing

RNF8/212 Ring Finger Protein 8/212

RS Round Spermatids

RT Room Temperature

SA Stromal Antigen protein

SC Synaptonemal Complex

SDS Sodium Dodecyl Sulfate

sec second(s)

SET Su(var)3-9, Enhancer-of-zeste and Trithorax

Sg Spermatogonia

SIX6OS1 Six6 opposite strand transcript 1

SMC1/3 Structural Maintenance of Chromosomes 1/3

SNP Single Nucleotide Polymorphism

SSC Side-Scattered light

SPATA22 Spermatogenesis Associated protein 22

SPO11 Meiotic recombination protein SPO11

SSXRD SSX Repression Domain

STA-PUT Cell sorting by velocity sedimentation

STAG3 Stromal Antigen 3

SYCP1 Synaptonemal Complex Protein 1

SYCP2 Synaptonemal Complex Protein 2

SYCP3 Synaptonemal Complex Protein 3

TAD Topologically Associated Domain

TE Tris EDTA buffer

TEX11/12 Testis-Expressed protein 11/12

TF Transcription Factor

Thy1 Thy1 cell surface antigen

TOPOVIB Type 2 DNA topoisomerase 6 subunit B

TP1/2/4 Transition Protein 1/2/4

TSS Transcription Start Sites

UAB Universitat Autònoma de Barcelona

UV ultraviolet

WAPL Wings Apart-Like Homolog

ZnF Zinc Finger

Contents

List of Figures

List of Tables

GENERAL INTRODUCTION

1.1 Genome Architecture

Understanding how the genome is organized and regulated within the cell nucleus has been one of the main questions in biology. The spatial conformation of the genome arises from a close interplay between gene function and chromatin organization, contextualizing the overall cell function. How different levels of chromatin organization interact among themselves in different species underpin the potential of genome plasticity. This, together with the study of its dynamics during the cell cycle and cell differentiation are both areas of research that are still in their infancy and largely unexplored.

1.1.1 A historical perspective of chromosomal territoriality

Initial descriptions of the cell and its nucleus date back to the 17th and 18th centuries and were provided by relevant figures such as Robert Hooke (1635-1703), Marcello Malpighi (1628-1629) and Antoni van Leeuwenhoek (1632-1723). However, it was not until the 19th century when the first representations of individual structures within the nucleus were made by Schleiden (1804-1881) and Schwann (1810-1882). Later on, Walther Flemming (1843–1905) referred to such structures as mitotic chromosomes [reviewed in (Ribatti 2018)]. During that period, Carl Rabl (1853-1917) proposed the territorial organization of chromosomes in interphase, known as the "Rabl configuration" (Rabl and Gegenbauer 1885), which was later confirmed by Theodor Boveri, who introduced the term **'chromosome territory'** (CT) in his seminal studies of blastomere stages of the horse roundworm (Boveri 1909). Boveri already predicted that: (i) CT order is maintained during interphase, (ii) chromosome neighborhood patterns change from prophase to metaphase, and (iii) new chromosome neighborhood arrangements established in the metaphase plate are conserved to a considerable extent throughout cell division [reviewed in (Cremer and Cremer 2010)].

Although the hypothesis of territoriality was still supported during the 1940s, the development of electron microscopy in the 1950s together with the discovery of DNA's double-helix structure by Franklin, Crick, and Watson (Franklin and Gosling 1953; Watson and Crick 1953), dismissed the concept of CTs for nearly half a century [reviewed in (Cremer and Cremer 2006; Cremer and Cremer 2010)]. Most of the studies focused on nuclear organization developed between the 1970s and 1980s considered that DNA organized into chromatin fibers between 10-30 nm in size, intermingling with each other during interphase, with no retention of individual chromosomes (Wischnitzer 1973). However, and contrary to the accepted view, Stack and collaborators recovered the concept of CTs (Stack et al. 1977), further supported by laser-UV-microirradiation experiments on Chinese hamster cells (Cremer et al. 1982). In their influential study, Cremer and colleagues showed that post-UV damage resulted in a localized damage label, a pattern expected if chromosomes would occupy specific regions throughout the nuclear space, thus providing the first experimental evidence of CTs.

The subsequent development of **Fluorescence In Situ Hybridization (FISH)** techniques (Manuelidis 1985; Pinkel et al. 1986) allowed for the visualization of CTs using fluorescence microscopy. That was the case of initial chromosome painting experiments, which facilitated the visualization of individual chromosomes in human-hamster hybrid cell lines (Lichter et al. 1988). Since then, FISH-based experiments have permitted the study of the nuclear architecture in different taxa including yeast (Bystricky et al. 2005; Molnar and Kleckner 2008), plants (Berr et al. 2006; Pecinka et al. 2004; Shaw et al. 2002), and animals [reviewed in (Cremer and Cremer 2001)]. This has provided evidence

of the evolutionary conservation of CTs between species, such as primates (Mayer et al. 2005), cattle (Koehler et al. 2009), and birds (Habermann et al. 2001). The combined use of chromosomal probes and oligopaints (Beliveau et al. 2015) in three-dimensional (3D) FISH approaches (Cremer et al. 2008) together with confocal and high-resolution microscopy has allowed to delineate patterns of nonrandom distribution of CTs and chromosomal subregions within nuclei.

One of the main factors influencing territoriality is chromosome size, as large chromosomes tend to occupy peripheral positions whereas small chromosomes are orientated towards the center of the nucleus (Parada and Misteli 2002). This radial configuration is concomitant with gene density and has been observed in several species, not only in mammals (Grasser et al. 2008; Koehler et al. 2009; Mayer et al. 2005; Shopland et al. 2006) but also in *Drosophila melanogaster* (Boutanaev et al. 2005) and plants (Tiang et al. 2012). In this manner, gene-dense chromosomes preferentially locate in the interior of the nucleus while gene-poor chromosomes are mostly found in the nuclear periphery (Boyle et al. 2001; Cremer and Cremer 2001; Croft et al. 1999). This radial disposition is further influenced by replication dynamics, as peripheral chromatin replicates later in S-phase than internal chromatin (O'Keefe et al. 1992) and presents lower gene expression levels (Jackson et al. 1993). In fact, most genes located in 'lamina associated domains' (LADs) are transcriptionally silent, suggesting that interactions with nuclear lamina are involved in gene expression regulation [reviewed in (Peric-Hupkes and van Steensel 2010)]. As such, gene transcriptional activity is one of the main factors influencing chromatin dynamics.

Notwithstanding major advances in the field provided by conventional and super-resolution microscopy, the analysis of the spatial conformation of the genome presented limitations in resolution (100-200 nm), which prevented the definition of higher-order genome structures. This was fulfilled by the development of the 3C-based techniques (Dekker et al. 2002) and its derivatives. These approaches quantitatively measure frequencies of close spatial proximity (10-100 nm) between genomic loci averaged genome-wide over a large population of fixed cells. All 3C techniques are based on the digestion and subsequent re-ligation of crosslinked chromatin within nuclei, detecting the spatial proximity between DNA sequences and their probability of interaction (McCord et al. 2020) (**Box 1**). Since their development almost two decades ago, 3C approaches have unveiled fundamental principles of the hierarchical 3D organization of genomes based on high-throughput sequencing (Dekker et al. 2013; Gibcus and Dekker 2013; Lieberman-Aiden et al. 2009; Rao et al. 2014). The initial application of 3C approaches to human cells and later on to mouse, *Drosophila*, yeasts and plants has provided a detailed description of the compartmental distribution of chromosomes into hierarchical folding levels, from the kilobase (kb) scale to the mega-base (Mb) scale (Dixon et al. 2012; Hou et al. 2012; Lieberman-Aiden et al. 2009; Nora et al. 2012; Rao et al. 2014; Sexton et al. 2012b).

1.1.2 The hierarchical three-dimensional organization of the genome

The classical chromatin-folding model assumes the presence of a compacted '30-nm fiber' structure in heterochromatin and the metaphase chromosomes while existing as a relaxed '11-nm fiber' (the nucleosome diameter) in euchromatin and interphase chromosomes (Finch and Klug 1976; Horn and Peterson 2002). However, the integration of multidisciplinary approaches such as Hi-C, super-resolution microscopy, cryo-EM and X-ray scattering has challenged this classical model (Eagen 2018; Maeshima et al. 2014; Moraru and Schalch 2019). Thus, the current view is that mammalian genomes are packaged into a hierarchical chromatin structure, the regulation of which depends on several superimposed layers of organization: (i) chemical modifications of the DNA (i.e., methylation and acetylation), (ii) nucleosomes that wrap the DNA around four core histones (H2A, H2B, H3, and H4) which can suffer modifications as well, and (iii) the 3D high-order organization of chromatin compartments inside the nucleus (**Figure 1.2**).

Box 1

3C-based techniques. Chromosome Conformation Capture (3C)-based methods include: 3C, 4C, 5C, ChIA-PET and Hi-C (**Figure 1.1**). Both 3C and 4C use specific primers for individual genomic loci, yielding single interaction profiles. While 3C generates long-range interaction profiles of the selected loci versus surrounding chromatin, 4C generates a genome-wide interaction profile of the selected loci. In the case of 5C, universal primers are combined with multiplex PCR amplification and sequencing, allowing for the detection of millions of interactions by using thousands of primers in a single experiment. ChIA-PET combines chromatin immunoprecipitation (ChIP) with 3C, allowing the study of genome-wide long-range chromatin interactions bound to a protein of interest. Finally, Hi-C does not depend on specific primers and yields genome-wide contact maps thanks to high-throughput sequencing (Dekker et al. 2013; Han et al. 2018). A novel variant [micro-C; (Krietenstein et al. 2020)] provides chromosome folding patterns at the nucleosome resolution, detecting thousands of new looping interactions.
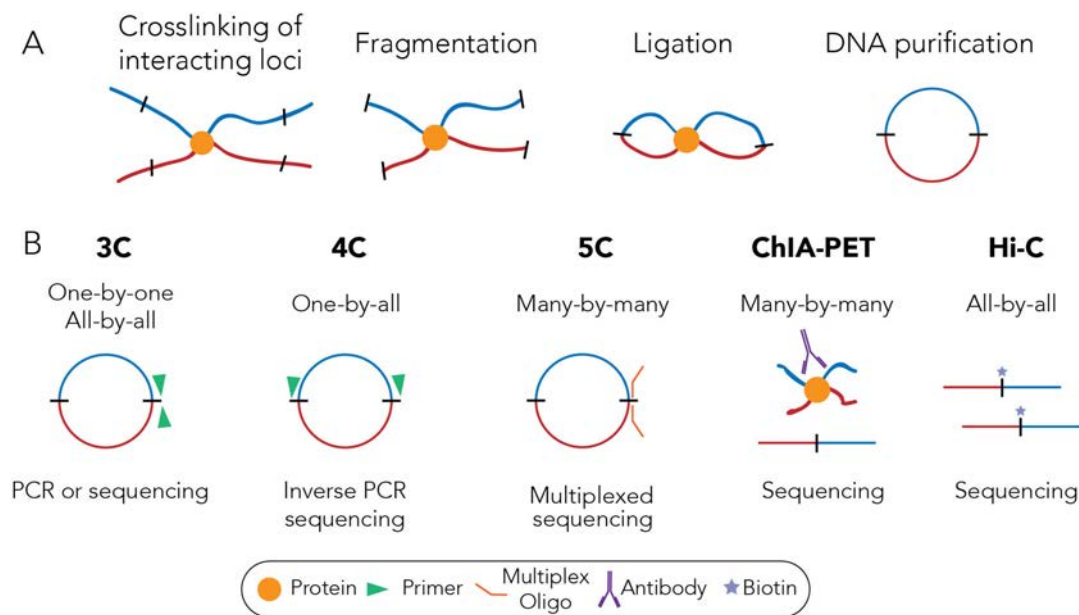
Figure 1.1: Summary of 3C-based techniques (A) Schematic workflow for 3C-based methods. Any chromatin segments that are spatially close in the cell nucleus are initially linked. Chromatin is then fragmented either by enzymatic digestion or by sonication. Then, crosslinked fragments are ligated forming chimeric DNA molecules that are finally purified, sequenced and analyzed. **(B)** Representation of the main 3C-based techniques. Adapted from (Dekker et al. 2013).

It is generally accepted that the 2-meter DNA polymer wraps around histones forming nucleosomes (nucleosomal scale of organization), giving rise to the 'beads on a string' structure (**Figure 1.2**) that form the '11 nm fiber' (Kumar et al. 2020). This chromatin fiber is tightly regulated by post-transcriptional histone modifications (i.e., methylation and acetylation), the incorporation of variant histones, and protein-binding factors involved in DNA replication, transcription and repair (Cosgrove and Wolberger 2005). Further folding of chromatin fibers gives rise to the supranucleosomal scale of genome organization, which represents the **3D high-order organization**. It includes different levels of genome folding from chromatin loops to the transcription-dependent compartmental domains within chromosomes, each occupying specific CTs within the nucleus (**Figure 1.2**) (Dekker et al. 2013; Ea et al. 2015; Gibcus and Dekker 2013; Hansen et al. 2018; Lajoie et al. 2015). This hierarchical organization of the genome has been reported in different mammalian species, including humans (Lieberman-Aiden et al. 2009; Rao et al. 2014), primates (Eres et al. 2019; Vietri Rudan et al. 2015; Yang et al. 2019), mouse (Zhang et al. 2012), dog and rabbit (Vietri Rudan et al. 2015). In *Drosophila*, physical domains correlated with active/repressive histone marks (Eagen 2018; Hou et al. 2012; Rowley et al. 2017; Sexton et al. 2012b). This link between genome organization and its transcriptional
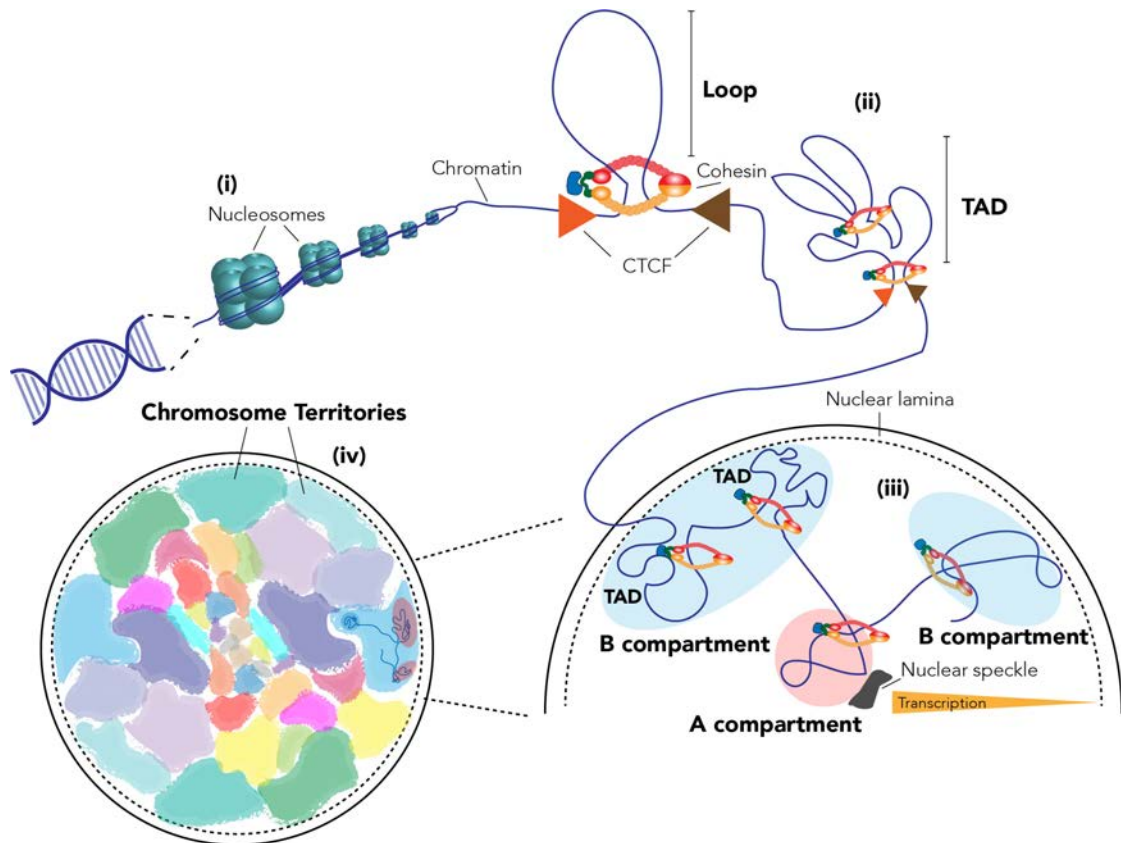
Figure 1.2: Hierarchical 3D organization of the genome. (i) The DNA polymer wraps around histones forming nucleosomes, conforming the chromatin fiber. **(ii)** Chromatin fibers fold into chromatin loops that highly interact with other neighboring loops, conforming topologically associated domains (TADs). Loop boundaries are determined by cohesin complexes in-between CTCF bound convergent motifs. **(iii)** TADs are organized into A or B compartments, according to chromatin accessibility and transcriptional activity. Transcriptionally repressed regions (B) are mostly associated with the nuclear lamina at the nucleus periphery. In contrast, transcriptionally active regions (A) are located towards the center of the nucleus and associated with, e.g., nuclear speckles. **(iv)** Compartments are found within CTs within nuclei.

state has also been reported in yeast (Duan et al. 2010; Mizuguchi et al. 2014), plants (Dong et al. 2017; Grob et al. 2014) and fungi (Galazka et al. 2016), although they are less clear and may depend on different factors from those described in mammals.

Altogether, extensive evidence derived from Hi-C studies has confirmed that the hierarchical 3D organization of the genome and the correlation between transcriptional state and folding of genomic domains is a fundamental and conserved principle across eukaryotes (Duan et al. 2010; Feng et al. 2014; Hou et al. 2012; Mizuguchi et al. 2014; Nagano et al. 2013; Rao et al. 2014; Rowley et al. 2017; Sexton et al. 2012b; Wang et al. 2015a; Zhang et al. 2012). This 3D organization is characterized by different genomic architectural features: (i) *cis/trans* interaction ratio, (ii) distance-dependent interaction frequency, (iii) genomic compartments, and (iv) topological associated domains (TADs) and DNA loops.

1.1.2.1 Genome topology revealed by interaction probabilities

The strongest interaction patterns detected by Hi-C experiments reflect average genome-wide trends and include (i) **cis/trans chromosomal interactions** and (ii) **distance-dependent interaction frequencies** (Lajoie et al. 2015). On average, pairs of loci residing within the same chromosome (*cis* interactions), present higher interaction frequencies that those residing in different chromosomes (*trans* interactions). At the genome-wide level, this *cis/trans* interaction pattern is a reflection of the presence of CTs (**Figure 1.3 A**). Also, as the genomic distance between loci increases, the interaction

between nearby loci gradually decreases as loci are further away from each other (**Figure 1.3 B**). This reduction in interactions can be represented as a genomic distance-dependent interaction curve [P(s)], representing the level of compaction of chromosomes (Dekker et al. 2013).

The following hierarchical level of the 3D organization includes **genomic compartments**, which are provided by the first principal component of Hi-C interaction matrices and captured by the correspondent eigenvector, which discriminates between interaction frequencies (Lajoie et al. 2015; Lieberman-Aiden et al. 2009; Rao et al. 2014; Wang et al. 2016a). In this manner, **A compartments** are defined as genomic regions of low interactions, whereas **B compartments** are genomic regions with high interaction frequencies (**Figure 1.3 B**). Thus, chromosomes are organized in A/B compartments that can vary between 1 to 10 Mb in mammals (Dekker et al. 2013) down to 15 kb in *Drosophila* (Rowley et al. 2017). Importantly, this genomic organization into compartments is linked to the transcriptional state of the chromatin following the classical description of euchromatin (i.e., A compartment) and heterochromatin (i.e., B compartment) (Kumar et al. 2020). Therefore, A compartments are correlated with the presence of active histone modifications ('open' chromatin) such as H3K9ac, H3K27ac, H3K36me3, H3K79me2, H3K4me1 or H3K4Me3, and therefore regions with active transcription (Eagen 2018; Nagano et al. 2013; Nurick et al. 2018; Rao et al. 2014). Moreover, it is known that 'open' chromatin conformations are rich in genes and CpG islands (Terrenoire et al. 2010), more accessible to DNAse I (Eagen 2018), and are characterized by low levels of DNA methylation (Bird 1986; Gilbert 2005). Conversely, B compartments are correlated with the presence of inactive histone modifications ('closed' chromatin) such as H3K27Me3 and H3K9Me2/3 (Rao et al. 2014; Sexton et al. 2012a). Thus, B compartments correspond to chromatin with a transcriptionally repressed state, including regions linked to heterochromatin or tethered to the nuclear lamina (Rao et al. 2014; Rowley and Corces 2018; Szabo et al. 2019).
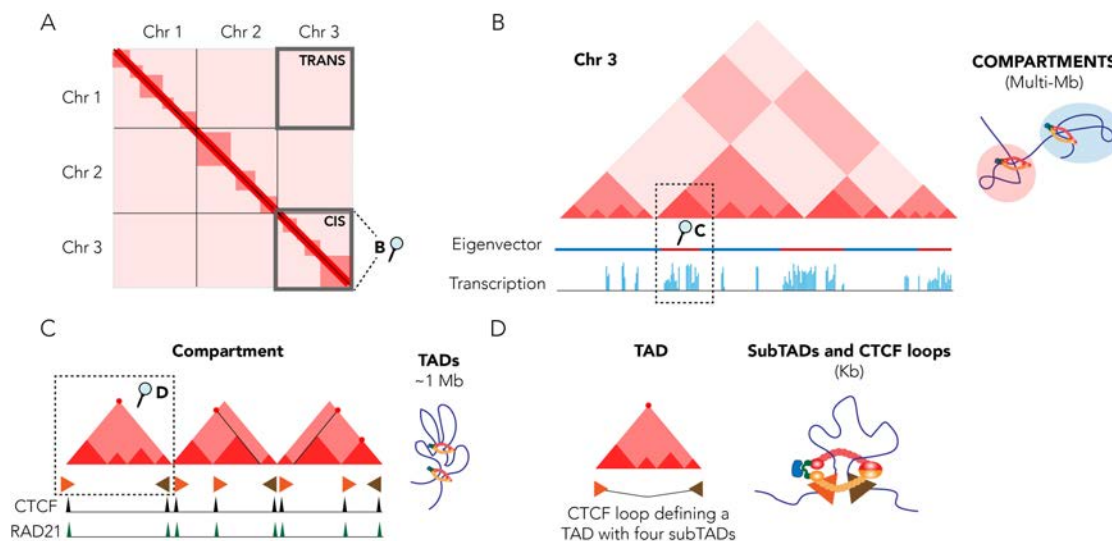
Figure 1.3: Hi-C interaction patterns. (A) Schematic representation of a Hi-C contact probability matrix in which interaction signals from just three chromosomes are shown for clarification. The *trans* interaction pattern of chromosomes 1 and 3 is squared in grey together with the *cis* interaction pattern of chromosome 3. The *cis* interaction pattern is zoomed in panel B. **(B)** Representation of the interactions within chromosome 3 (*cis*). The eigenvector interpretation identifies A (red) and B (blue) compartments, which correlate with transcriptionally active and inactive genomic regions, respectively. The dotted line marks a specific compartment, zoomed in panel C. **(C)** Representation of several TADs within the squared compartment in panel B, where also regions of high interaction can be detected (sub-TADs, represented as darker shades of red). CTCF loops are shown as punctate signals corresponding to strong interactions between CTCF sites. **(D)** Representation of a single TAD from panel C, containing four sub-TADs and flanked by CTCF-cohesin sites forming a loop. Adapted from (Lajoie et al. 2015; Rowley and Corces 2018).

1.1.2.2 Topological associated domains and chromatin loops

Hi-C studies have also revealed the existence of sub-domains within A/B compartments that can range from 40 kb to 3 Mb in size [median size of 186 kb in human cells (Rao et al. 2014)], known as

topologically associated domains (TADs) (**Figure 1.3 C and D**). In this case, TADs are not defined by their chromatin state, but by their interaction pattern instead. Thus, TADs represent genomic loci that preferentially interact with the neighboring *cis* chromatin domains rather than with other regions (Eagen 2018; Hansen et al. 2018; Lajoie et al. 2015; Lieberman-Aiden et al. 2009; Rao et al. 2014), conforming functional chromatin domains (Dixon et al. 2016).

TADs are structures that are generally conserved between different cell types and species (Dixon et al. 2012; Vietri Rudan et al. 2015) and can be classified upon the chromatin signatures present in the highly-interacting domains that define them. These signatures include active/inactive histone marks, polycomb group proteins, and also the lack of specific marks (void or null TADs) (Del Prete et al. 2015; Szabo et al. 2019). In mammals and yeast, TADs correlate with active/inactive transcriptional state (Hou et al. 2012; Imakaev et al. 2012; Lieberman-Aiden et al. 2009; Naumova et al. 2013; Rao et al. 2014; Szabo et al. 2019). Nevertheless, in plants and fungi, where heterochromatic regions appear to dominate genome organization, TADs are less obvious (Duan et al. 2010; Feng et al. 2014; Galazka et al. 2016; Mizuguchi et al. 2014). Studies in *Drosophila* have shown that TADs can also be organized upon polycomb bodies (Boettiger et al. 2016; Sexton et al. 2012b) and this observation extended to plants (Feng et al. 2014). Remarkably, studies in *C. elegans* described a very different pattern of TAD organization, with TADs being relevantly less frequent in somatic chromosomes and only clearly present in the X chromosome (Crane et al. 2015). However, despite the general acceptance of TADs existence, its definition derived from Hi-C studies is not without complexities. In fact, TADs contain multilayered hierarchies of overlapping structures, suggesting additional interaction patterns (Beagan and Phillips-Cremins 2020; Lajoie et al. 2015; Sikorska and Sexton 2019). Furthermore, TAD detection strongly depends on Hi-C data resolution and the methodology used for TAD annotation (Paytuví-Gallart 2019; Serra et al. 2017). So, by increasing the sequencing depth and the resolution, the presence of two fundamental properties for TAD organization has been suggested: (i) the self-association of regions within the TAD and, (ii) the insulation between neighboring TADs (Dixon et al. 2016; Szabo et al. 2019).

TADs are further organized into individual **chromatin loops** or insulation neighborhoods (Dowen et al. 2014; Ji et al. 2016; Nagano et al. 2017; Phillips-Cremins et al. 2013; Rao et al. 2014). These sub-TAD structures also present self-associative and insulation properties, posing the question of whether these structures are reflecting different scales of the same genomic organization level. Regarding this matter, studies show that TADs are conserved between cell types (Dixon et al. 2012; Nora et al. 2012), whereas sub-TADs, loops and insulation neighborhoods are related to cell-specific regulatory events (Dowen et al. 2014; Ji et al. 2016; Phillips-Cremins et al. 2013; Rao et al. 2014). Also, TADs have been suggested to be equivalent to replication domains, while sub-TADs have a different function in regulating DNA replication (Pope et al. 2014).

Hence, TADs are considered chromatin units that define regulatory landscapes, as they are directly implicated in the regulation of enhancer-promoter interactions (Hansen et al. 2018; Lajoie et al. 2015; Merkenschlager and Nora 2016; Szabo et al. 2019). In this manner, TADs partition genomes into "regulatory neighborhoods", in which co-regulation of genes has been observed across different cell types and tissues (Flavahan et al. 2016; Nora et al. 2012; Shen et al. 2012). Furthermore, boundaries between TADs have been implicated in the blocking of the spread of activity between neighboring TADs (Dixon et al. 2016).

1.1.2.3 Architectural proteins insulating TADs and loops

Evidence from several mammalian species has shown that TAD boundaries can be defined by the presence of insulator proteins, such as **CTCF** (also known as 11-zinc finger protein or CCCTC-binding factor) along with **cohesins** (Barrington et al. 2019; Busslinger et al. 2017; Cattoglio et al. 2019; de Wit et al. 2015; Hansen et al. 2018; Krivega and Dean 2017; Merkenschlager and Nora 2016). Initial studies in human and mouse somatic cell lines revealed that TAD boundaries are enriched in CTCF (detected at 76% of all boundaries), active transcription marks (such as H3K4me3 and H3K36me3),

nascent transcripts, housekeeping genes (present in 34% of TAD boundaries) and repeat elements (Bonev and Cavalli 2016; Dixon et al. 2012).

CTCF is a relevant insulator protein that contributes to genome folding (Phillips and Corces 2009; Splinter et al. 2006) (**Box 2**). It is conserved in most bilaterian metazoans (Acemel et al. 2017; Heger et al. 2012), and it is broadly expressed across tissues (Bastiaan Holwerda and de Laat 2013), being essential for cellular function (Bailey et al. 2018). It was initially described as the only known insulator protein in vertebrates (Kim et al. 2007; Ohlsson et al. 2001), binding to target sites across the genome in a sequence-specific manner.

CTCF, together with **cohesins** (**Box 3**), can generate stable chromatin loops (**Figure 1.2, Figure 1.3, Figure 1.6**) that were initially deduced from high-resolution Hi-C data as string punctate signals (Rao et al. 2014; Rowley and Corces 2018) . For a CTCF loop to occur, CTCF has to bind to motifs that are convergently oriented (forward and reverse), serving as loop anchors (Guo et al. 2015; Rao et al. 2014). While almost all TAD borders co-localize with CTCF, only around one-third of CTCF ChIP-seq detected peaks flank TADs (Hansen et al. 2018; Merkenschlager and Nora 2016). Therefore, not all CTCF potential sites are actively engaged in forming chromatin loops in a given moment inside the cell. However, CTCF cannot stabilize loops without the active involvement of cohesins (Rao et al. 2017).

Box 2

CTCF structure. The CCCTC-binding factor (CTCF) is a protein with three conserved domains: the N-terminal, the C-terminal, and the particular 11-zinc-finger domain **(Figure 1.4)**, which recognizes and binds to a variety of DNA sequences (Ohlsson et al. 2001; Zlatanova and Caiafa 2009). These three domains facilitate the interaction with other proteins and with CTCF itself, being highly relevant for establishing links between sites within (looping) and between (bridging) chromosomes (Zlatanova and Caiafa 2009). Moreover, CTCF has been described to bind to the nuclear matrix, suggesting a functional connection between CTCF-dependent insulator elements and the nuclear matrix (Dunn et al. 2003). The combinatorial use of its zinc fingers allows for CTCF's great flexibility in binding-site recognition, which mostly flanks transcriptionally co-regulated genes, confirming the role of CTCF-binding sites as insulators.
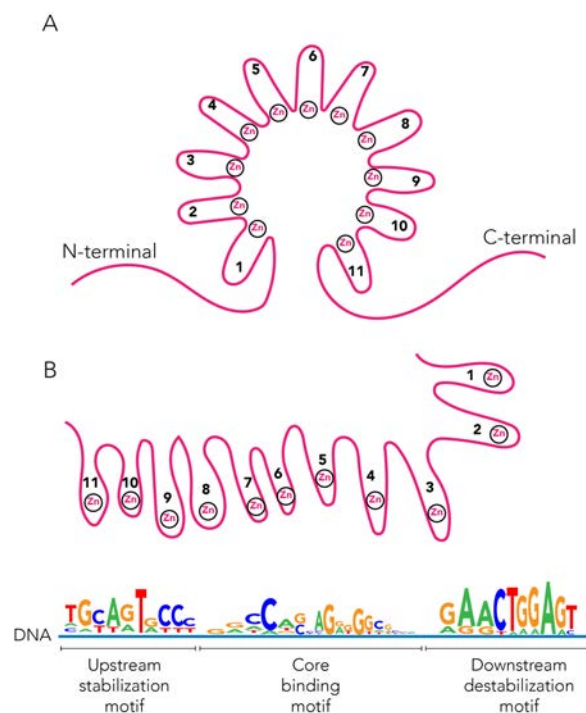
Figure 1.4: CTCF structure. **(A)** CTCF is composed by an N-terminal domain, an 11-zinc-finger-domain, and a C-terminal domain. **(B)** The zinc-finger domain-specific binding to DNA is determined by the combinatorial arrangement of the zinc fingers, in which fingers #3 to #8 bind to the core binding motif. In contrast, the other fingers are involved in either stabilizing or destabilizing binding. Adapted from (Arzate-Mejía et al. 2018; Norton and Phillips-Cremins 2017).

Box 3

Somatic cohesins. Cohesins are ring-shaped protein complexes that are essential for sister chromatid cohesion. Additionally, they have a determinant role in the assembly of DNA replication factories, DSB repair, chromosome condensation, and mitotic spindle assembly, among others (McNicoll et al. 2013). The cohesin complex is formed by the structural maintenance of chromosome (SMC) proteins 1 and 3, an α-kleisin and a stromal antigen protein (SA). In mammalian somatic cells, the α-kleisin is RAD21 and two different SA proteins have been described, SA1 and SA2 **(Figure 1.5)**. SA1 contributes to the stabilization of TAD boundaries, and SA2 promotes cell-type-specific contacts between enhancers and promoters within TADs independently of CTCF (Kojic et al. 2017). Thus, RAD21 depletion results in a general loss of preferential contacts and a general relaxation of chromatin structure (Rao et al. 2017; Sofueva et al. 2013).
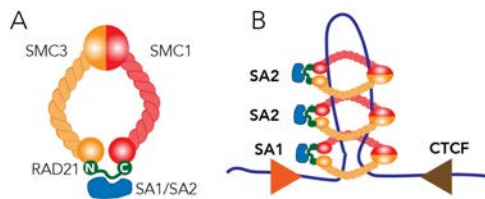
Figure 1.5: Schematic representation of somatic cohesins. (A) Structure of somatic cohesin, formed by SMC proteins 1 and 3, RAD21 and either stromal antigen protein 1 or 2. **(B)** Representation of the role of SA1-cohesin complex in stabilizing TAD boundaries by staking at the base of DNA loops while SA2-cohesin complex is involved in cell-specific loops. Adapted from (Kojic et al. 2017).

1.1.2.4 Loop-extrusion model

The fact that chromatin loops are preferentially formed in-between convergently oriented CTCF-bound sites with a strong bias against the opposite orientation (Guo et al. 2015; Rao et al. 2014), suggests that CTCF binding results from a linear tracking process that can read the directionality of DNA sequences (Merkenschlager and Nora 2016). In this context, the **Loop Extrusion Model** has been proposed to explain how DNA loops are formed during a process mediated by CTCF and cohesins (Fudenberg et al. 2016; Sanborn et al. 2015). Under this model, a cohesin ring extrudes the chromatin fiber progressively until it is blocked by bound CTCF **(Figure 1.6)** (Fudenberg et al. 2016; Sanborn et al. 2015). Briefly, structural maintenance of chromosomes (SMC) proteins that form the cohesin ring together with RAD21 can topologically entrap and move along the DNA until they meet a blocking obstacle such as CTCF (Davidson et al. 2016; Kanke et al. 2016; Skibbens 2016; Stigler et al. 2016). Then, the cohesin complex is driven by SA1 to CTCF-bound sites **(Figure 1.6 A)** and loaded to the chromatin fiber by NIPBL (Nipped-B-like protein) **(Figure 1.6 B)**. From here, loop extrusion **(Figure 1.6 C)** can be accomplished via three main processes: (i) extrusion via cohesin diffusion, in which cohesins are constantly loading to chromatin and generating a diffusion gradient **(Figure 1.6 Ci)**; (ii) extrusion by cohesin motor activity driven by ATP hydrolysis **(Figure 1.6 Cii)** and (iii) extrusion by pushing of RNA polymerase II (RNAPII) **(Figure 1.6 Ciii)**. During the extrusion process, some cohesin complexes are released from the chromatin fiber by the action of WAPL (Wings APart-Like protein homolog) and PDS5 (sister chromatid cohesin protein pds5) **(Figure 1.6 E)** (Rowley and Corces 2018). Recent evidence further shows that cohesins can be translocated to the CTCF binding sites by the transcriptional activity both in mammals and yeast (Busslinger et al. 2017), and also by other factors such as transcription-induced chromatin super-coiling [reviewed in (Racko et al. 2019)].

The loop extrusion model has been recently demonstrated *in vitro* (Fudenberg et al. 2016; Krietenstein et al. 2020). As such, this model explains how TADs are formed, putting into relevance the role of CTCF and cohesins in conforming these chromatin domains. CTCF and cohesins would facilitate enhancer-promoter interactions while also restricting those interactions between sequences located inside and outside CTCF loops (Dowen et al. 2014; Hnisz et al. 2016a), thus conferring TADs their defining features. Moreover, the extrusion process itself would also explain why enhancers and promoters located on opposite sides of a loop anchor are less likely to interact (Rowley and Corces

2018). Hence, CTCF and cohesins are intimately implicated in gene expression regulation through genome folding (Merkenschlager and Nora 2016).
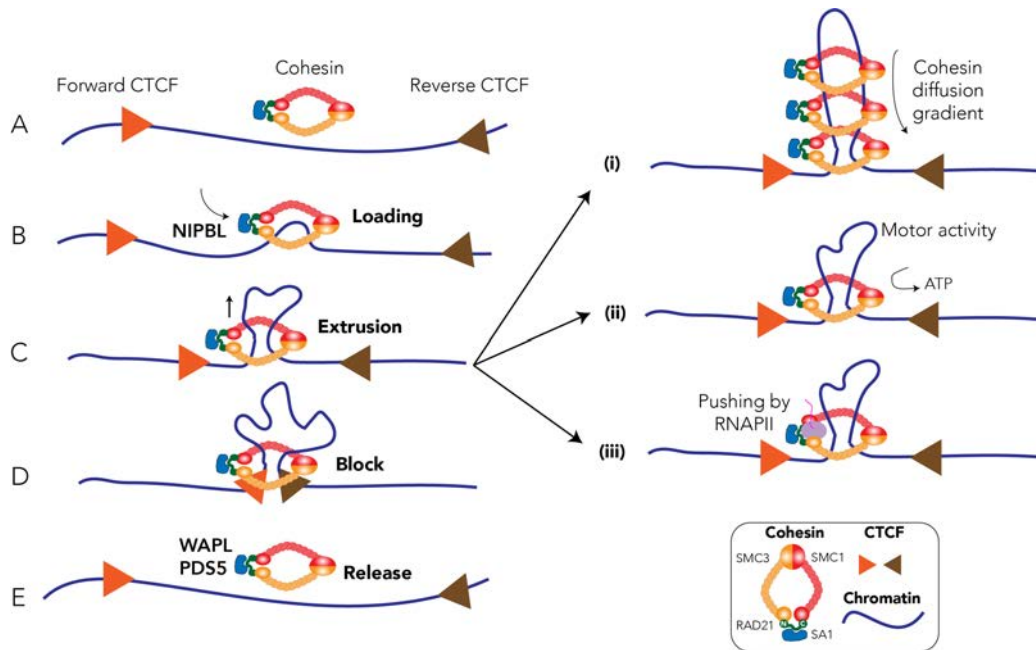
Figure 1.6: Representation of the loop extrusion model. The chromatin fiber is entrapped by a RAD21 cohesin complex loaded to the fiber by NIPBL (A, B, and C). The chromatin loop extrudes through the cohesin either by cohesin diffusion gradient (Ci), by ATP-driven cohesin motor activity (Cii) or pushed by RNAPII (Ciii). The loop extrusion process continues until inversely oriented CTCF blocks the process (D). Some cohesins are released during the extrusion process by WAPL and PDS5. Adapted from (Rowley and Corces 2018).

1.1.2.5 Dynamics of genome organization during the cell cycle

We have seen how the structural organization of the genome is highly hierarchical and modulated by multiple factors in interphase nuclei, accounting for the transcriptional needs of the cell. Moreover, recent single-cell Hi-C studies are providing a new genome-wide glimpse of cell-to-cell variation, revealing that while individual chromosomes maintain a similar organization in the mega-base scale, they present different conformations at larger scales (Beagan and Phillips-Cremins 2020; Krietenstein et al. 2020; Nagano et al. 2013). However, several questions still remain. What are the topological details of cohesin-mediated chromosomal interactions, and what are their dynamics over time? Is the 3D genome folding set once after mitosis and then lasts for the remaining interphase, or is it continuously shaped during the cell cycle?

Naumova and colleagues were the first to describe the specific organization of the genome during mitosis in human cells based on Hi-C data (Naumova et al. 2013). Based on fiber modeling, they inferred the presence of two differential folding states during the cell cycle. In mitotic cells, the characteristic interphase plaid-pattern observed in Hi-C contact maps disappears, and compartments and TADs are not longer detectable (Nagano et al. 2017). Further studies revealed that the loss of interphase organization is mediated by condensins, which also mediate the formation of consecutive loops during mitosis (Gibcus et al. 2018). In the case of metaphase chromosomes, modeling approaches suggest that the organization of the chromatin emerges from linear compaction by consecutive chromatin loops, followed by axial compression (Naumova et al. 2013). On the contrary, inter-chromosomal contacts present the opposite pattern, with an increased local contact during interphase and increased long-range mitotic contacts during mitosis, consistent with the chromosome reorganization into the metaphase plate (Barrington et al. 2017; Nagano et al. 2017). The description of the organization of the mitotic chromosome posits the question whether this specific chromatin

configuration is maintained during the formation of germ cells and its transmission to the next generation.

1.2 Mammalian gametogenesis

Gametogenesis is the biological process by which haploid gametes are generated. This process is divided into two main stages: (i) proliferation and differentiation of gonium and (ii) two consecutive meiotic divisions followed by a specific differentiation program **(Figure 1.7)**. Male and female gametogenesis present fundamental differences, which include timing, cell morphology and overall cell cycle regulation. While mammalian oogenesis is asynchronous, spermatogenesis is continuous throughout the life of the individual. This continuity allows gaining a deeper understanding of germ cell generation and their regulatory processes.

1.2.1 Overview of mouse spermatogenesis

In the house mouse (*Mus musculus domesticus*), the spermatogenic cycle initiates when the first pool of spermatogonial stem cells is established (Ginsburg et al. 1990; McLaren 2003; Phillips et al. 2010). Spermatogonial stem cells originate from primordial germ cells (PGCs) within the primitive streak of embryos at the epiblast stage [7-7.25 days *post coitum* (dpc)] (Chiquoine 1954; Ginsburg et al. 1990; Richardson and Lehmann 2010). It is during the formation of the allantois (between 8.5 and 12.5 dpc) when PGCs migrate and colonize the genital ridges (Bendel–Stenzel et al. 1998). At 12.5 dpc, PGCs are already considered gonocytes, and after two or three rounds of mitosis, they enter a premeiotic stage of mitotic arrest at G0/G1 as pro-spermatogonia (McLaren 1984; McLaren 2003). From 13.5 dpc onwards, pro-spermatogonia are confined in the testicular cords formed by Sertoli precursor cells and myoid cells, where they resume proliferation the first week after birth (McLaren 2003; Phillips et al. 2010). Between 1 and 4 days postpartum (dpp), pro-spermatogonia migrate from the testicular cord to the basal membrane of the seminiferous tubules (Clermont and Perey 1957). During the pre-pubertal period (1-6 dpp), the seminiferous tubules only contain Sertoli cells and spermatogonia. These cells initiate the first round of spermatogenesis, establishing the initial pool of spermatogonial stem cells around eight days after birth and progressively generating germ cells during the entire lifespan of the individual (Bellvé et al. 1977; Nebel et al. 1961).
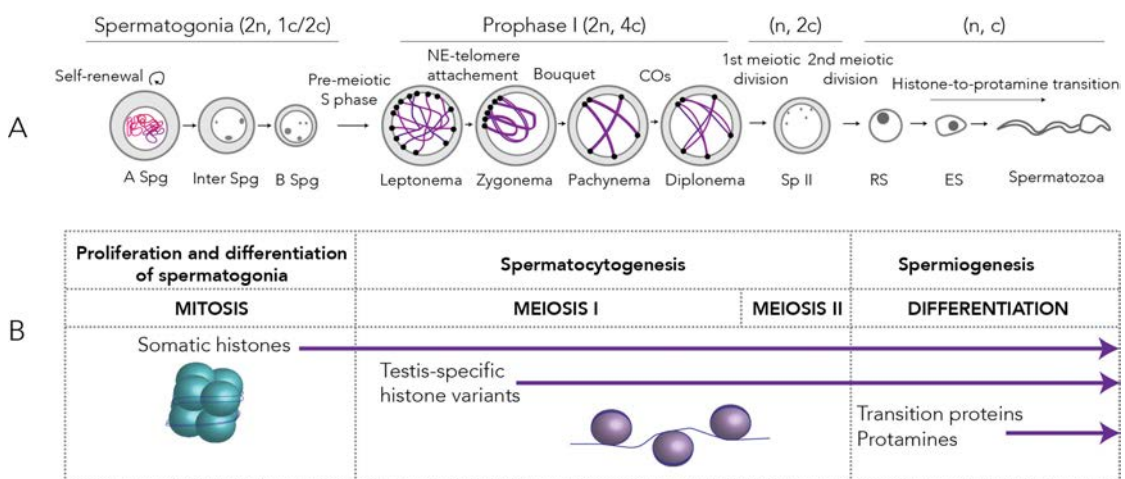
Figure 1.7: Schematic representation of spermatogenesis. (A) Summary of the spermatogenic process. Self-renewing A spermatogonia commit to meiosis and subsequently differentiate into intermediate and B spermatogonia. After DNA duplication, primary spermatocytes are generated, which undergo meiotic prophase I: leptonema, zygonema, pachynema and diplonema. The first meiotic division results in secondary spermatocytes (SpII) that, in their turn, divide in round spermatids (RS). Round spermatids undergo spermiogenesis, differentiating into elongated spermatids (ES) and finally spermatozoa. Adapted from (Reig-Viader et al. 2016). **(B)** Histone changes per spermatogenesis phase. Adapted from (Sharma and Agarwal 2011).

In adult mice, spermatogenesis takes place within the testes in the seminiferous tubules (Sharma and Agarwal 2011; Walker 2011) in a cycle that spans every 8.6 days (Clermont 1972; Oakberg 1956a; Phillips et al. 2010). Seminiferous tubules constitute a multi-layer germinative epithelium that represents nearly 90% of the testis volume (Romrell et al. 1976; Wistuba et al. 2007) **(Figure 1.8)**. Each tubule is divided into the basal and the adluminal compartments, separated by the blood-testis barrier (BTB) (Cheng and Mruk 2002; Phillips et al. 2010; Wistuba et al. 2007), which is formed by tight junctions between Sertoli cells that attach to the basal lamina of the epithelium (Cheng and Mruk 2002; Wistuba et al. 2007). Spermatogonia and early primary spermatocytes are found in the basal compartment, while the rest of the spermatocytes and spermatids are located in the adluminal compartment **(Figure 1.8)**. The BTB insulates germ cells residing in the adluminal compartment from the circulatory and lymphatic systems, generating an immunologically privileged microenvironment that allows for the completion of meiosis (Mruk and Cheng 2015; Sharma and Agarwal 2011).

Spermatogenesis in adult mice is continuous throughout the seminiferous tubules and it can be divided into twelve stages (I-XII) (Oakberg 1956a), each one presenting different combinations of spermatogonia, spermatocytes and spermatids that synchronously proceed through spermatogenesis (Meistrich and Hess 2013; Phillips et al. 2010; Wistuba et al. 2007). This synchronicity is facilitated by the fact that as germ cells divide and differentiate, they remain joined by inter-cellular bridges **(Figure 1.8)**.
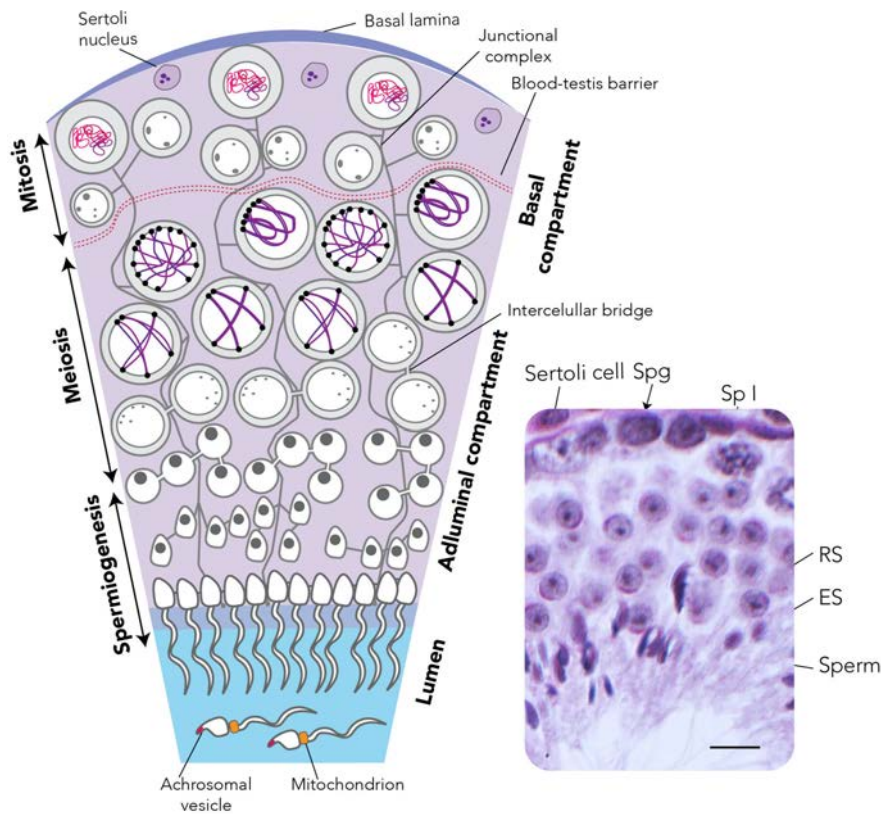
Figure 1.8: The histological organization of the testes. Left panel: schematic representation of the different germ cells and their organization within the seminiferous tubules. Right panel: microscopy image of a histological section of mouse testis stained with Periodic Acid-Schiff (PAS)-hematoxylin showing the different cell types detectable in that specific section: Sertoli cells, spermatogonia, primary spermatocytes, round and elongated spermatid and sperm in the lumen. Scale bar: 20 µm. Spg-spermatogonia, SpI-primary spermatocyte, RS-round spermatid, ES-elongated spermatid.

Spermatogenesis can be divided into three stages: (i) proliferation and differentiation of spermatogonia, (ii) meiosis: two rounds of DNA division, where the genome is reduced from diploid (2n) to haploid (n) DNA content, and (iii) spermiogenesis: the process of spermatid maturation to spermatozoa. All three stages of spermatogenesis are characterized by dramatic and highly

regulated chromatin remodeling and cell-type-specific gene expression.

1.2.1.1 Proliferation and differentiation of spermatogonia

Spermatogenesis initiates with the generation of a population of self-renewing spermatogonia that either mitotically multiply or commit to meiosis and differentiate (de Rooij and Russell 2000; O'Donnell et al. 2000; Wistuba et al. 2007). Traditionally, three main types of spermatogonia have been described upon their nuclear morphology (Clermont and Leblond 1953; Monesi 1962; Roosen-Runge and Giesel 1950): (i) type A, (ii) intermediate, and (iii) type B spermatogonia. While type A spermatogonia are defined by the absence of heterochromatin (a characteristic of undifferentiated cells), intermediate spermatogonia contain small proportions of heterochromatin regions that are increased in type B spermatogonia (Phillips et al. 2010).

Several models have been proposed to explain the proliferation and differentiation of spermatogonia: (i) A_s model (Huckins 1971; Oakberg 1971), (ii) the **fragmentation model** (Hara et al. 2014; Nakagawa et al. 2010), and (iii) the **hierarchical A_s model** (Aloisio et al. 2014; Chan et al. 2014; Helsel and Oatley 2017; Helsel et al. 2017; Komai et al. 2014; Sun et al. 2015a). Irrespectively of which model prevails, self-renewing spermatogonial stem cells can be distinguished from those committed to meiosis by the presence of specific markers. Spermatogonial stem cells presenting Thy1 (also known as CD90.2) are considered self-renewing spermatogonial stem cells, while cells positive for Kit are spermatogonial stem cells committed to the differentiation pathway, remaining Kit positive throughout spermatogonial differentiation (Hammoud et al. 2014; Tseng et al. 2015). Additionally, several studies have shown that self-renewing spermatogonial stem cells express specific proteins such as ID4, PAX7, and GFRA1 (Aloisio et al. 2014; Gassei et al. 2010; Helsel et al. 2017; Tseng et al. 2015) that are not found in differentiating spermatogonial stem cells. Relevantly, both spermatogonial stem cell types present bivalency and transcriptional silencing at specific promoters (e.g., promoters of embryonic transcription factors) and also poised upstream enhancers (Hammoud et al. 2014). This **poised state** is attained by particular methylation patterns that also involve atypical DNA methylated promoters present through all spermatogenesis (Hammoud et al. 2014). Spermatogonial stem cells would establish a poised state of specific promoters and enhancers to maintain a competent pluripotency program (needed for the development if the future embryo) during spermatogenesis, by silencing developmental genes (e.g., *Nanog*), while enforcing others involved in the meiotic process *per se*.

1.2.1.2 Meiosis

Meiosis-committed spermatogonia migrate out of their niche in the basal compartment of the seminiferous tubule to follow the differentiation pathway, generating intermediate spermatogonia that differentiate into type B spermatogonia. As type B spermatogonia lose contact with the basal membrane and mitotically divide and differentiate into primary spermatocytes at pre-leptonema (pre-meiotic phase). Spermatocytes then transit across the BTB into the adluminal compartment, where they can progress through meiosis **(Figure 1.7)** (Mruk and Cheng 2015). During the prophase of the first meiotic division, four different stages can be distinguished: leptonema, zygonema, pachynema, and diplonema **(Figure 1.7)**.

Following the meiotic replication phase, pre-leptonema spermatocytes initiate DNA condensation and assembly of the synaptonemal complex (SC) (see **section 1.2.2**), with telomeres tethering to the nuclear membrane while transiently pairing between homologous telomeres (Bellvé et al. 1977; Boateng et al. 2013; Scherthan et al. 1996). At **leptonema**, double-strand breaks (DSBs) take place as homologous chromosomes start to condense, pair and synapse, clustering their telomeres in the *bouquet* structure [reviewed in (Reig-Viader et al. 2016)]. At this stage, the proteinaceous structure of the SC starts to assemble along with meiotic cohesins (REC8 and RAD21L) (**Box 4**). Elements of the SC include the lateral element proteins SYCP2 and SYCP3, the transverse filament protein

SYCP1, the central element proteins SYCE1, SYCE2, SYCE3, TEX12 along with SIX6OS1 (Dapper and Payseur 2019; Fraune et al. 2012; Gómez-H et al. 2016; Henderson and Keeney 2005; Keeney et al. 2014; Scherthan et al. 1996; Winkel et al. 2009; Zickler and Kleckner 1999; Zickler and Kleckner 2015). At **zygonema**, homologous chromosomes are entirely paired, allowing for synapsis to begin [reviewed in (Reig-Viader et al. 2016)]. It is during leptonema and zygonema when resected strands from DSBs undergo homology search and strand invasion (Dapper and Payseur 2019), which will be repaired either as crossovers (COs) or non-crossovers (NCOs). Homologous chromosome synapsis and recombination are completed at **pachynema**, with a subset of recombination sites maturing into COs (between 20-26 COs per cell in mouse) (Baier et al. 2014) (see **section 1.2.2**). Meiosis I ends at **diplonema** with the segregation of homologous chromosomes, which remain physically attached at sites where recombination took place (Handel and Schimenti 2010). As such, successful progression through prophase I is dependent on faithful completion of two interconnected processes: the assembly of chromatin loops into chromosomal axes, and the formation and repair of DSBs (Keeney et al. 1997; Longhese et al. 2009; Romanienko and Camerini-Otero 2000). The first meiotic division results in **secondary spermatocytes**, which undergo a short secondary meiotic division to produce **round spermatids (Figure 1.7 and Figure 1.8)** (Handel and Schimenti 2010).

Box 4

Meiotic cohesins. During meiosis, somatic cohesins are replaced by two meiotic-specific cohesin complexes: REC8 and RAD21L, both containing the SA protein STAG3 **(Figure 1.9)**. It is during the pre-meiotic DNA replication of spermatogonia that RAD21 is no longer detected, and REC8 appears, persisting throughout the first meiotic division. RAD21L appears after DNA replication, reaching its peak in early prophase I (Lee and Hirano 2011) **(Figure 1.9)**. Meiotic cohesins are essential for chromosome synapsis between homologs, having a pivotal role in the formation of the chromosomal axis, holding chromatin loops to the axes, and determining the number and location of DNA loops **(Figure 1.11)** (McNicoll et al. 2013). Importantly, in the absence of REC8, the chromosome axis is separated in determined regions, giving rise to the assembly of illegitimate SC between sister chromatids. This phenomenon has also been observed in mice STAG3 mutants and KO for SMC1β, confirming the pivotal role of cohesin complexes in axis formation and SC assembly (Agostinho et al. 2016; Ishiguro and Watanabe 2016).
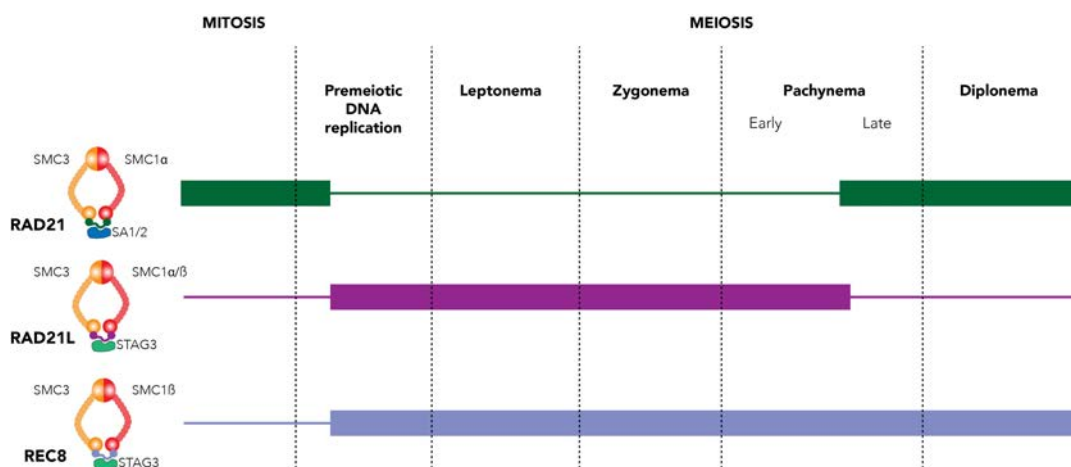
Figure 1.9: Temporal distribution of cohesin complexes during meiotic prophase I in mammals. During DNA replication previous to meiosis, the somatic cohesin-RAD21 is substituted by meiotic cohesins RAD21L and REC8, which are expressed until the end of prophase I. Adapted from (Lee and Hirano 2011).

Meiosis progression is accompanied by atypical chromatin-transcriptional relationships that involve transcription from DNA methylated promoters with 5hMC, already established in spermatogonial

stem cells (Hammoud et al. 2014). It is generally accepted that there are two waves of active transcription during meiosis: one before entering meiosis, and a second in the transition from round spermatids to sperm (da Cruz et al. 2016; Sassone-Corsi 2002). Most of the genes that are upregulated during leptonema and zygonema are soon down-regulated before pachynema (da Cruz et al. 2016). Moreover, there is a significant change in gene expression signatures at pachynema with the transcription of specific piRNAs that differ from those pre-meiotically expressed **(Figure 1.10)** (Hammoud et al. 2014). It is in pachynema that a large proportion of genes related to spermiogenesis and sperm functionality are expressed, with the highest number of expressed during diplonema (da Cruz et al. 2016; Ernst et al. 2019).
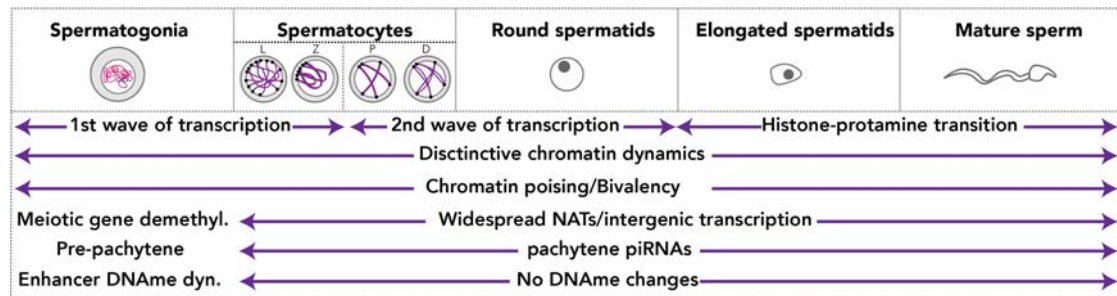
Figure 1.10: Summary of transcription and chromatin transitions during spermatogenesis. NATs-non-annotated transcription units, DNAme-DNA methylation. Adapted from (Hammoud et al. 2014). L-Leptonema; Z-Zygonema;P-Pachynema; D-Diplonema.

Finally, it is important to take into account that sex chromosomes behave differently than autosomes during male meiosis. As the X and the Y chromosomes only present a small region of homology (the pseudo-autosomal region or PAR) (Burgoyne 1982; Burgoyne et al. 2009), the presence of unsynapsed large regions during prophase I activates the silencing of sex chromosomes – a phenomenon called **meiotic sex chromosome inactivation (MSCI)** (Burgoyne et al. 2009; Margolin et al. 2014; Soumillon et al. 2013; Turner 2015). The transcriptional silencing of the chromosomes X and Y is detectable in pachynema spermatocytes as the 'sex body' and it is characterized by the presence of several histone modifications including γH2AX, H3K9me3/2, H2A ubiquitylation, HP1β (Namekawa et al. 2006; Turner et al. 2006) and absence of active RNA polymerase II, H3K27m1/3, H3K9ac and H4K16ac (Namekawa et al. 2006).

1.2.1.3 Spermiogenesis

Through **spermiogenesis**, round spermatids elongate to finally differentiate into sperm **(Figure 1.7)** (Brinkley et al. 1986; Govin et al. 2004; Oakberg 1956b). This maturation process can be divided into sixteen steps (1-16), a classification that in rodents is based on the profound morphological changes that take place (Oakberg 1956b). Briefly, the acrosome begins its formation between steps 1-8 concomitant with a high active transcriptional activity present round spermatids. Transcription decreases between steps 9-11 when changes in spermatid morphology are detectable. In these stages the acrosomic material is found in the dorsal, caudal angle of the nucleus. It is between steps 12-14 that spermatids nucleus condensate, finally presenting the characteristic hook-type spermatozoa morphology in steps 15-16 (Bao and Bedford 2016; Oakberg 1956b). Spermiogenesis ends with the release of spermatozoa to the lumen of the seminiferous tubule (Meistrich and Hess 2013). Sperm within the testis have little mobility, and they start maturation in the epididymis. Capacitation includes cellular or physiological changes in the sperm that are completed during the ejaculation, acquiring hypermotility and the ability to perform the acrosome reaction, binding to the zona pellucida of the female oocyte and accomplishing fertilization (Sharma and Agarwal 2011).

Thus, spermatids undergo dramatic morphological changes due to chromatin remodeling that results in a highly-compacted nucleus in mature sperm (Ward and Coffey 1991). This condensation is achieved by the replacement of the vast majority (99%) of nucleosomal histones by protamines

(Figure 1.7) (Balhorn 2007; Braun 2001; Dadoune 2003; Miller et al. 2010; Rathke et al. 2014; Steger 1999). For a successful histone-to-protamine transition to occur, mechanisms to access the DNA are involved, such as histone modifications (McPherson and Longo 1993), histone replacement by transitional proteins (Rathke et al. 2014) and the transient appearance of DNA breaks (Cavé et al. 2019; McPherson and Longo 1993). In mammals, histones are firstly substituted by the transition proteins TP1, TP2, and TP4, prompting chromatin condensation (Balhorn 2007, 2011), which are consequently replaced by protamines (Braun 2001; Carrell et al. 2007). Protamines stabilize the sperm chromatin, conferring a specialized and highly condensed toroid structure (Bao and Bedford 2016). This condensation facilitates the generation of the hydrodynamic shape that characterize sperm. Moreover, the super-coiling of chromatin enables the sperm to shed most of the cytoplasm, conferring the sperm with rapid mobility, while also protecting it from physical and chemical damage (Balhorn 2011; Bao and Bedford 2016; Braun 2001; Carrell et al. 2007; Oliva 2006; Sharma and Agarwal 2011).

Furthermore, the transcriptional silencing of sex chromosomes is maintained throughout spermiogenesis in a distinct compartment known as **postmeiotic sex chromatin (PMSC)** (Namekawa et al. 2006; Sin et al. 2015). Interestingly, 13% of sex-linked genes escape silencing. Some of these active genes are regulated by RNF8 (ring finger protein 8), directing histone modifications characteristics of PMSC such as H2A ubiquitination and H3K4me3 (Sin et al. 2012). Multicopy genes sitting in the X chromosome also escape postmeiotic silencing, although unclear, they may acquire permissive chromatin marks due to unusual DNA configurations risen from multicopy palindromes (Turner 2015).

Regarding genome-wide transcription, somatic/progenitor genes are silenced during spermatogenesis, while reproductive genes are activated (Sin et al. 2015). However, somatic/progenitor genes have to be successfully re-activated after fertilization. This re-activation is attained by establishing bivalent domains that are pre-programmed from spermatogonial stem cells to mature sperm **(Figure 1.10)** with poised chromatin, with the presence atypical 5hmC in transcriptional start sites (TSS) or proximal enhancers of progenitor genes involved in embryogenesis, allowing for the recovery the somatic/progenitor program (Hammoud et al. 2014; Sin et al. 2015).

1.2.2 Organization of meiotic chromosomes

1.2.2.1 The importance of chromosomal axes and DNA loops

During meiotic prophase I, homologous chromosomes are organized into DNA loops anchored to the chromosomal axes formed by the **SC** a proteinaceous structure with a zipper-like morphology that mediates the synapsis of homologous chromosomes (Fraune et al. 2012; Syrjänen et al. 2017; Zickler and Kleckner 2015). The SC establishes the context in which synapsis and recombination between homologs take place as it prompts strand invasion and the subsequent genetic recombination. The SC also has a pivotal role in joining sister chromatids within the lateral components of the SC axis by meiotic cohesin complexes (Rankin 2015), such as REC8 (Xu et al. 2005) and RAD21L (**Box 4**) (Lee and Hirano 2011; Llano et al. 2012). As such, the SC modulates the structure of meiotic chromosomes, arranging chromatin in DNA loops (around 20 loops per micron of axis length) that emerge out of chromosomal axes **(Figure 1.11)** (Kleckner 2006; Zickler and Kleckner 1999).

Importantly, the SC length has direct implications for the number and length of SC-anchored DNA loops (Kleckner et al. 2003; Ruiz-Herrera et al. 2017; Zickler and Kleckner 1999). For instance, in some mammalian species (i.e., humans), females show longer SC axis (and, therefore, shorter DNA loops) than males (Gruhn et al. 2013; Zickler and Kleckner 2016). Also, species with different genome sizes present variations in loop size or total axis length while maintaining the basic axial structure (Kleckner 2006; Ruiz-Herrera et al. 2017). Altogether, these observations highlight the modulation of loop size based on chromosome axis length (Wang et al. 2019a).

1.2.2.2 Formation and repair of DSBs

Successful progression through early stages of prophase I is dependent on the assembly of chromatin loops into chromosomal axes and the formation and repair of DSBs. These two processes are interconnected and tightly regulated (Keeney et al. 1997; Longhese et al. 2009; Romanienko and Camerini-Otero 2000). DSBs are generated during leptonema by SPO11 (Keeney et al. 1997; Romanienko and Camerini-Otero 2000), a highly conserved endonuclease amongst eukaryotes (Malik et al. 2007) **(Figure 1.11)**. The formation of DSBs activates DNA damage response (DDR) mechanism (Baudat et al. 2010; Myers et al. 2010; Parvanov et al. 2010) as an integral part of the meiosis program. Together with its leading role in DSBs generation, SPO11 is also involved in the pairing of homologous chromosomes (Boateng et al. 2013), as a minimum number of DSBs are needed to ensure inter-homolog interactions (Kauppi et al. 2013; Smagulova et al. 2013).
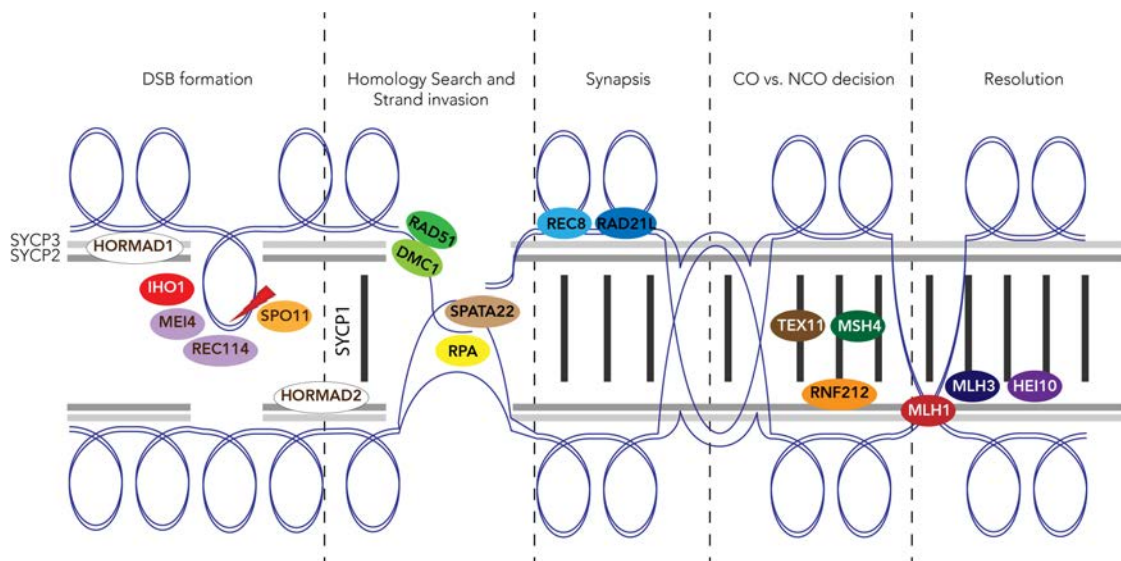
Figure 1.11: Mechanics of recombination. Schematic representation of the main proteins involved in meiotic recombination and the resolution of DSBs into COs. Adapted from (Dapper and Payseur 2019).

Although numbers vary between species (Kauppi et al. 2013; Ruiz-Herrera et al. 2017; Wang et al. 2019a), between 200-300 DSBs occur genome-wide. The formation and repair of DSBs occurs in the context of the SC with the implication of different factors **(Figure 1.11)**. The DSBs machinery is formed by the MCD recombinosome (MEI4-Containing DSB-promoting), a proteinaceous complex that includes MEI4, REC114, TOPOVIB, and IHO1 (interactor of HORMAD protein 1) (Kumar et al. 2015; Robert et al. 2016; Stanzione et al. 2016). The MCD directly activates SPO11 (Dapper and Payseur 2019; Stanzione et al. 2016), which generates the DSBs monitored by both HORMAD1 and HORMAD2. Once the double-stranded DNA is resected, SPO11 is erased and posteriorly depleted from the recombination site, leaving 3' single-stranded overhangs at both sites of DSBs (Lam and Keeney 2015). In an orchestrated manner, the formation of DSBs induces the phosphorylation of histone H2AX on serine 139 (γH2AX) by the ATM/ATR pathway (Burma et al. 2001; Kuo and Yang 2008; Rogakou et al. 1998), which recruits DDR proteins. These include RPA (replication protein A) (He et al. 1995), DMC1 (DNA meiotic recombinase I) (Yoshida et al. 1998), and RAD51 (RAD51 recombinase) (Pittman et al. 1998), among others. As such, the 3' single-stranded overhangs are coated by RPA (Dapper and Payseur 2019; Ribeiro et al. 2016), which recruits recombinases RAD51 and DMC1 that facilitate the homologous chromosome search and establishment of inter-chromosome interactions **(Figure 1.11)**. The formation of the SC stabilizes the RAD51/DMC1 action during leptonema/zygonema.

Hence, DSBs are repaired in the context of the SC, involving the lateral element protein SYCP3 (Henderson and Keeney 2005). The single-stranded end elongates and, by homology search, invades the opposing DNA helix, displacing the complementary DNA strand and generating a D-loop

(displacement loop) (Baudat et al. 2013; Gray and Cohen 2016). The completion of strand invasion is facilitated by SPATA22 (spermatogenesis associated 22) **(Figure 1.11)** (Dapper and Payseur 2019). Depending on the interaction that is established (either with the D-loop or with the elongated invading strand), two different recombination products can be generated: (i) COs or (ii) NCOs (Cole et al. 2010; Moens et al. 2007). Ultimately, COs are resolved with the involvement of the MutL complex, recruited to COs sites by TEX11 and composed by MLH1 and MLH3, which repair mismatches (Dapper and Payseur 2019). HEI10 is also recruited to the SC together with the MutL complex, antagonistically regulating RNF212, thus also involved in COs regulation **(Figure 1.11)**.

The reciprocal exchange of genomic regions between homologous chromosomes generated by the resolution of Holliday junctions results in the formation of **COs** (Kleckner 2006; Moens et al. 2007; Zickler and Kleckner 2015). The number and position of COs along chromosomes are tightly regulated, involving MSH4 (MutS protein homolog 4), which is stabilized RNF212 (Ring finger protein 212) (Dapper and Payseur 2019; Moens et al. 2007). This refined regulation ensures the formation of at least one COs per bivalent (**obligatory CO**) (Bolcun-Filas and Schimenti 2012; Borde and de Massy 2013; Kleckner 2006). The genomic distribution of COs positively correlates with GC content and gene density while negatively correlating with transposable elements (TE), overall relating with certain regulatory elements and histones modifications (Stapley et al. 2017). Additional factors govern the number and distribution of COs per chromosome. These include crossover interference (Muller 1916) and the centromeric effect (Beadle 1932; Mather 1938). **Crossover interference** is a universal phenomenon by which the formation of a CO at one site along the chromosome interferes with the establishment of additional COs nearby [reviewed in (Otto and Payseur 2019; Zickler and Kleckner 2016)]. CO interference is essential to prevent the formation of multiple DSBs in the same chromatin loop, which would be hazardous for genome stability (i.e., may cause chromosomal rearrangements). Under the **centromeric effect**, on the other hand, the formation of COs is reduced at centromere proximities (Beadle 1932; Mather 1938; Talbert and Henikoff 2010). This phenomenon avoids disruption of pericentric sister chromatid cohesion during the first meiotic division.

Moreover, and given that chromosomal axis length is inversely correlated with the size of chromatin loops emerging from the SC (Kleckner et al. 2003; Zickler and Kleckner 1999), the number and distribution of COs per chromosome is linked to the structural organization of the genome during meiosis (Wang et al. 2019a). In fact, CO distribution is influenced by the physical length of the chromosomal axis (micrometers) rather than genomic (Mb) or genetic distance (cM) (Ruiz-Herrera et al. 2017; Zickler and Kleckner 2015). This is exemplified by the PAR region, where the chromatin arranged in numerous and small loops experience a 10-fold increase in DSBs compared to other regions of the genome (Acquaviva et al. 2020; Kauppi et al. 2011; Mu et al. 2020).

1.2.3 PRDM9

1.2.3.1 Structure and function

Early events that determine where DSBs occur across the genome are critical for the formation of DSBs. In this context, **PRDM9** (PR/SET domain-containing protein 9) determines the location of meiotic recombination hotspots in the majority of mammals studied so far to the exclusion of canids (Baker et al. 2017; Baudat et al. 2010; Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014; Muñoz-Fuentes et al. 2011; Myers et al. 2010; Parvanov et al. 2010). The *Prdm9* gene was first described as the hybrid sterility gene *Hst1* (Forejt and Iványi 1974) based on its link with the sterility observed in *M. m. domesticus* x *M. m. musculus* hybrids (Forejt 1996). Further studies showed that *Hst1* (also known as *Meisetz*) was located in chromosome 17 in mice (Forejt et al. 1991; Gregorova and Forejt 2000) and encoded for PRDM9, an essential meiosis-specific protein (Hayashi and Matsui 2006; Mihola et al. 2009).

PRDM9 is a histone (H3) methyltransferase expressed in primary spermatocytes during early prophase I (Parvanov et al. 2017; Sun et al. 2015b). The PRDM9 protein presents four functional domains **(Figure 1.12 A)**: (i) an N-terminal KRAB domain involved in inter-protein interactions (Imai et al. 2017; Parvanov et al. 2017), (ii) an SSXRD signal of nuclear localization, (iii) a SET domain that provides a methyltransferase activity (Eram et al. 2014; Hayashi et al. 2005; Powers et al. 2016; Wu et al. 2013) and (iv) a zinc finger (ZnF) domain containing a highly variable C2H2 [Cys(2)-His(2)] ZnF array (Berg et al. 2010; Parvanov et al. 2010). In the house mouse, the ZnF array is formed by repetitions in tandem of 84 bp units (translated into 28 amino acids). Each ZnF presents hypervariable positions at amino acid coordinates -1, +3, and +6 of the ZnF alpha helix, determining DNA-binding specificity (Baudat et al. 2010) **(Figure 1.12 B)**. The ZnF array binds specifically to a degenerated DNA motif mainly through ZnF domains in positions #3 to #6 (Baker et al. 2017; Patel et al. 2016, 2017), while the remaining fingers stabilize the protein-DNA binding (Billings et al. 2013; Pratto et al. 2014). It is the composition of the terminal ZnF array that determines the identity of PRDM9 variants, which are species-specific **(Figure 1.12 A)**.
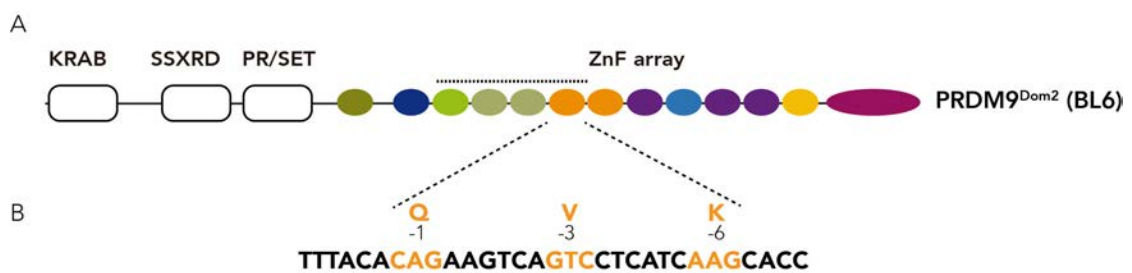
Figure 1.12: PRDM9 structure. (A) Representation of the PRDM9 protein structure and its domains. In this example, the PRDM9 variant found in laboratory mice C57BL/6J (Dom2) is represented. Each circle from the zinc finger array represents one tandem repeat, and the dotted line on top of the ZnF array indicates the ZnF positions significantly involved in DNA binding (ZnF #3 to #6). **(B)** Representation of one tandem repeat (84 bp unit), which its hypervariable positions translate to the aminoacid triplet QVK. Modified from (Capilla 2015).

Catalyzed by its SET domain, PRDM9 adds trimethylation marks in lysines 4 an 36 of histone 3 from close-by nucleosomes (Baker et al. 2014; Eram et al. 2014; Hayashi et al. 2005; Powers et al. 2016; Smagulova et al. 2011; Wu et al. 2013) **(Figure 1.13)**. These histone modifications are epigenetic marks defining open chromatin status, thus allowing DNA accessibility by displacing methylated nucleosomes creating nucleosome depleted regions (NDRs) (Baker et al. 2014). Recently, it has been described that not only H3k9me3 but also H3K4Me1 and H3K9ac marks can be associated with PRDM9-determined hotspots in mouse (Spruce et al. 2020).

According to recent evidence, PRDM9 forms a multi-protein complex in combination with the chromatin remodeler protein HELLS (lymphoid-specific helicase) (Spruce et al. 2020) **(Figure 1.13)**. The PRDM9-HELLS complex has the ability to: (i) recognize partial motifs within nucleosomal DNA, (ii) generate active histone modifications, and (iii) increase chromatin accessibility; thus, providing the required chromatin accessibility for DSBs machinery. Hence, the assembly with HELLS would facilitate the remodeling of chromatin architecture, allowing for correct hotspot localization and subsequent DSB formation and repair (Spruce et al. 2020).

Once specific DNA motifs are recognized, PRDM9 interacts with additional proteins through its KRAB domain. It has been suggested that PRDM9 binds to EHMT2 (euchromatic histone-lysine N-methyltransferase 2) and CDYL (chromodomain Y-like protein) in leptonema (Parvanov et al. 2017). Both proteins limit the extent of nucleosome trimethylation, thus regulating PRDM9 action **(Figure 1.13 A)** (Baker et al. 2014; Parvanov et al. 2017; Powers et al. 2016). At the end of leptonema, EHMT2 and CDYL are removed, and PRDM9 interacts with CXXC1 (CXXC-type zinc finger protein 1), which later recruits IHO1 as part of the MCD recombinosome (Imai et al. 2017). SPO11-resected regions are recruited by the PRDM9-coordinated protein multiplex to the synaptonemal proteinaceous scaffold to facilitate DSB repair (Baudat et al. 2013; Parvanov et al. 2017). The recruitment of resected sections to the SC is accomplished by the formation of the PRDM9-EWSR1 complex, which interacts with axis-bound cohesin REC8 **(Figure 1.13 B)**.
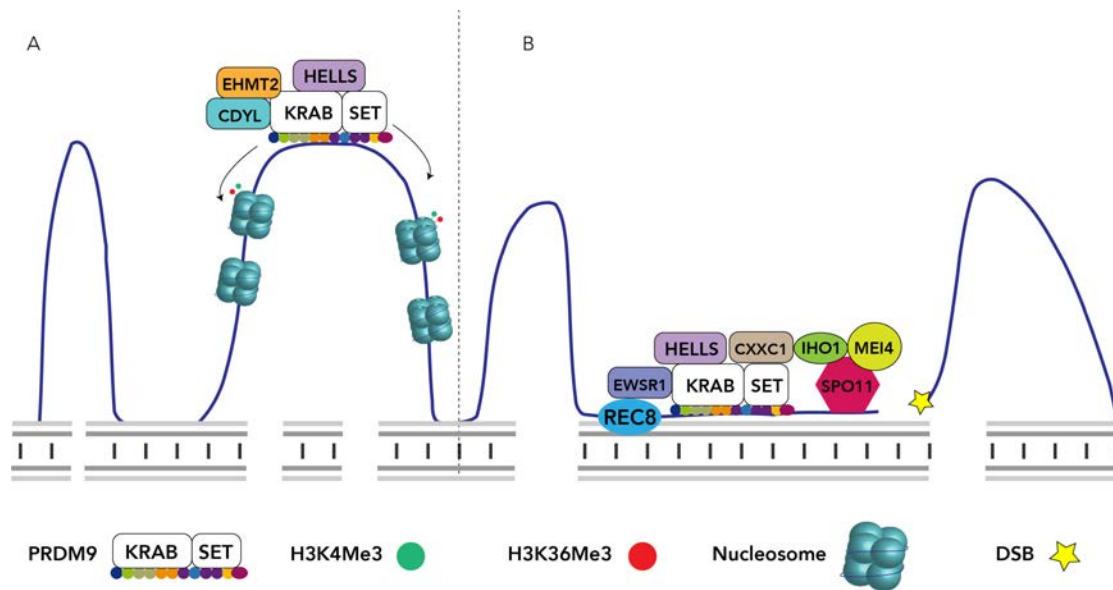
Figure 1.13: PRDM9 determines recombination hotspots and recruits the DSB machinery. (A) PRDM9 forms a complex with HELLS and determines recombination hotspots by specific-sequence recognition through its ZnF. PRDM9's SET domain adds H3K4Me3 and H3K36me3 methylation marks, establishing NDRs, facilitating chromatin accessibility. Methylation is limited by EHMT2 and CDYL, which interact with PRDM9 via its KRAB domain, thus regulating its activity. **(B)** The DSB machinery led by SPO11 is recruited by the interaction of PRDM9 with CXXC1, generating DSBs. Then, PRDM9 recruits the resected regions to the SC by the interaction with EWSR1 and the axis bound REC8, prompting DSB repair.

1.2.3.2 Evolutionary implications of PRDM9

The formation of DSBs does not occur randomly across the genome, but is localized in recombination hotspots (Arnheim et al. 2007; Baudat et al. 2013; Ortiz-Barrientos et al. 2016; Smagulova et al. 2011). Recombination hotspots in eukaryotes can be classified into ancestral hotspots and derived-location hotspots (Stapley et al. 2017). Ancestral hotspots are evolutionarily conserved across organisms and are temporally stable and associated with gene promoter regions (Stapley et al. 2017), while PRDM9-determined hotspots result from the recognition of a species-specific degenerated DNA motif. Closely related species with PRDM9-determined hotspots such as humans and chimpanzees present diverging locations of recombination events while species with PRDM9-independent mechanisms present similar location of recombination events (Baker et al. 2017).

PRDM9 is characterized by being highly polymorphic upon the variability of its ZnF array, presenting inter-specific differences in both its sequence and the number of ZnF conforming the array. This ZnF polymorphism results in a species-specific hotspot distribution, a pattern found in different mammalian species, including humans, non-human primates, cattle, and mice (Berg et al. 2010, 2011; Brick et al. 2012; Groeneveld et al. 2012; Sandor et al. 2012). But, how are recombination hotspots established and maintained if PRDM9 and the DNA motifs that recognizes are so variable? This poses the **hotspot paradox**, by which the PRDM9 binding to specific DNA motifs results in the disruption of those motifs by the recombination process itself, consequently eroding the DNA of the recognized hotspot sequences and leading to a weaker PRDM9 binding in the following generation (Boulton et al. 1997; Coop and Myers 2007; Davies et al. 2016). The paradox is that this has not happened. And this is mainly because PRDM9 is under positive selection, rapidly evolving to bind to new targets and maintaining hotspot activity. In this manner, new PRDM9 variants with different DNA binding specificities are generated in each generation, thus promoting the spread of new alleles within populations (Baudat et al. 2010; Oliver et al. 2009; Parvanov et al. 2010; Ponting 2011; Thomas et al. 2009; Tiemann-Boege et al. 2017). Wild mice populations represent an informative example of PRDM9 variability (Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014). In fact, 28 distinct alleles have been described in *M. m. domesticus,* 34 in *M. m. musculus*, and 37 in *M. m. castaneus* (Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014).

Experimental crossings between mice from different genetic backgrounds carrying different alleles have been informative in understanding the role of PRDM9 in **hybrid sub-fertility** and evolution. This is the case, for instance, of *M. m. musculus* females (e.g., PWD strain) x *M. m. domesticus* males (e.g., C57BL/6J strain) yielding an F1 in which males are infertile, presenting autosomal asynapsis at early pachynema that derives in meiotic arrest and final apoptosis of primary spermatocytes (Bhattacharyya et al. 2013; Dzur-Gejdosova et al. 2012; Flachs et al. 2012; Forejt 1996). The sterility detected in F1 males appears as a consequence of the negative epistasis between different *Prdm9* alleles, and several loci located in the X chromosome (Flachs et al. 2012; Mihola et al. 2009), generating **Dobzhansky-Muller incompatibilities** that contribute to postzygotic isolation (Bateson 1909; Coyne and Orr 2004; Dobzhansky 1936; Muller 1942; Seehausen et al. 2014). A closer inspection of recombination maps of mouse hybrids revealed that *Prdm9* variability could be subjected to functional constraints that facilitate the accumulation of allelic combinations that maintain recombination **hotspot symmetry** (i.e., differences in genomic distribution and number of DSBs between homologous chromosomes). As such, differences in a single ZnF repeat can affect the specificity of over 70% of the meiotic DSBs, leading to the redistribution of recombination sites in a single generation (Smagulova et al. 2016). In this context, DSBs tend to occur in the homolog chromosome that has suffered less DSB erosion (Smagulova et al. 2016), causing the asymmetric distribution of recombination hotspots, mispairing (Bhattacharyya et al. 2013; Gregorova et al. 2018) and eventually sterility (Mihola et al. 2009; Turner et al. 2014).

Altogether, PRDM9 arises as an essential genic factor that can have an influence in the evolution and the diversity of eukaryotic genomes (Yamada et al. 2017), being the only speciation gene hitherto described in mammals (Buard et al. 2014; Mihola et al. 2009; Turner et al. 2014).

1.3 Chromosomal reorganizations

1.3.1 Evolutionary implications of chromosomal reorganizations

Chromosomal rearrangements imply the reshuffling of chromosomal regions and can be classified depending on whether they alter gene dosage (**unbalanced**) or not (**balanced**). Balanced reorganizations include **inversions**, reciprocal **translocations**, **fissions**, and **fusions**, while unbalanced reorganizations include **duplications** and **deletions**.

Chromosomal rearrangements contribute to sequence divergence and eventually, speciation in several ways (Ayala and Coluzzi 2005; Faria et al. 2019; Faria and Navarro 2010; Navarro and Barton 2003; Wellenreuther and Bernatchez 2018). The early chromosomal speciation theory (King 1993; White 1969, 1973, 1978) also referred to as the 'hybrid dysfunction model' relies on the development of chromosomal incompatibility between divergent lineages due to impaired meiosis. Additional hypotheses do not necessarily invoke chromosomal rearrangement to reduce fitness, but rather prevent gene flow between individuals within the same population by suppressing recombination in the hybrid [i.e., the 'suppressed recombination model of speciation', (Ayala and Coluzzi 2005; Brown and O'Neill 2010; Faria and Navarro 2010; Farré et al. 2013; Navarro and Barton 2003)]. This can lead to the disruption of genes located in the reorganized regions, and they are implicated in the evolution of lineage-specific genes (Faria and Navarro 2010). But not only that. As a result of a reduction in recombination, chromosomal rearrangements can also capture new allele combinations or polymorphisms that confer some selective advantage (Guerrero and Kirkpatrick 2014), thus influencing fitness. Altogether, these events highlight the pivotal role of rearrangements in speciation (Capilla et al. 2016; Farré et al. 2015).

The evolutionary implications of chromosomal rearrangements are therefore linked to meiotic recombination (Capilla et al. 2014; Dumas and Britton-Davidian 2002; Faria and Navarro 2010; Farré et al. 2013; Wilch and Morton 2018). Specifically, inversions can delay the pairing and synapsis of homologous chromosomes during prophase I by the formation of inversion loops (Hale

1986; Torgasheva et al. 2013); while also generating other mechanical problems that can inhibit the formation of COs within the reorganized region (Navarro et al. 1997). Thus, recombination landscapes can be modified by chromosomal rearrangements, emphasizing the link between chromosomal rearrangements, meiotic recombination and evolution (Borodin et al. 2008; Capilla et al. 2014; Castiglia and Capanna 2002; Dumas and Britton-Davidian 2002; Faria and Navarro 2010; Farré et al. 2013; Förster et al. 2013; Franchini et al. 2010; Navarro and Barton 2003; Noor et al. 2001; Rieseberg 2001; Seehausen et al. 2014).

1.3.2 Robertsonian fusions

Robertsonian fusions (Rb) involve the centric fusion of two acrocentric chromosomes to form a single metacentric (Robertson 1916; Slijepcevic 1998). Rb fusions represent one of the principal sources of mammalian karyotype variation as they are present in different taxa, including mammals, reptiles, insects, or mollusks (Robinson 1995). These include small mammals such as the common shrew or the house mouse, which present Rb fusions at high rate. Such natural systems represent a valuable tool for the study of the implications of chromosomal rearrangements in genome architecture, fertility and evolution (Borodin et al. 2008; Hauffe et al. 2012; Piálek et al. 2005; White and Searle 2007). Three main mechanisms for Rb formation have been described in the literature: (i) elimination of p-arm telomeres by chromosome breakage before the fusion (Garagna et al. 1995; Nanda et al. 1995), (ii) telomere shortening, leading to telomere loss and high frequencies of Rb fusions (Blasco et al. 1997; Sánchez-Guillén et al. 2015) and (iii) telomere inactivation (Scherthan et al. 1996). Following the formation of the Rb fusion, the telomeric sequences are lost, and a new centromeric region is formed (Garagna et al. 1995; Nanda et al. 1995), maintaining the DNA polarity through the fusion point (Garagna et al. 2001).

In mice, Rb fusions are extensively found in western European *M. m. domesticus* populations **(Figure 1.14)**, with diploid numbers varying from 2n= 40 to 2n=22 (Piálek et al. 2005). These populations are characterized by a particular set of metacentric chromosomes in homozygosis, conforming to what is known as a **metacentric race** (Garagna et al. 2014). Groups of metacentric races can be found in the same geographic area, conforming to what are known as **"metacentric systems"**. One example of a metacentric system is the one found in the Portuguese archipelago of Madeira (Britton-Davidian et al. 2007; Britton-Davidian et al. 2000; Förster et al. 2013; Gündüz et al. 2001a). Metacentric races usually develop within small-sized populations that are isolated from each other by physical barriers such as islands or mountains (Britton-Davidian et al. 2000). There are also areas in which metacentric races are in contact with other metacentric races or with standard populations, generating an area of polymorphisms known as **"hybrid zone"**, such as the one described in the west of Lake Garda in Northern Italy (Hauffe et al. 2011; Hauffe et al. 2012). Nevertheless, there are exceptional situations in which metacentric chromosomes are not fixed within populations, but they are present in polymorphic state. This is the case of the specific area around Barcelona (Spain) known as the **"Barcelona Rb System" (BRbS)** (Capilla et al. 2014; Gündüz et al. 2001a; Gündüz et al. 2010; Medarde et al. 2012, 2015).

1.3.2.1 The Barcelona Rb system

The BRbS is found within an area of about 5,000 km^2 that extends in the vicinity of Barcelona, in the north-east of the Iberian Peninsula (Adolph and Klein 1981; Gündüz et al. 2001a; Medarde et al. 2012; Sans-Fuentes et al. 2005). In this system, different Rb mouse populations with diploid numbers ranging from 27 to 39 are found, surrounded by standard populations (2n=40). Seven specific chromosomes fusions have been described within the area either in homozygosis or heterozygosis (3.8, 4.14, 5.15, 6.10, 7.17, 9.11, and 12.13) and with different geographical distributions (Sans-Fuentes et al. 2005, 2007). Rb populations are defined as chromosomal clines presenting high metacentric frequency, located in a region surrounded by standard mice populations that delineate the staggered distribution of the clines (Gündüz et al. 2001a). Interestingly, no chromosomal race has been detected
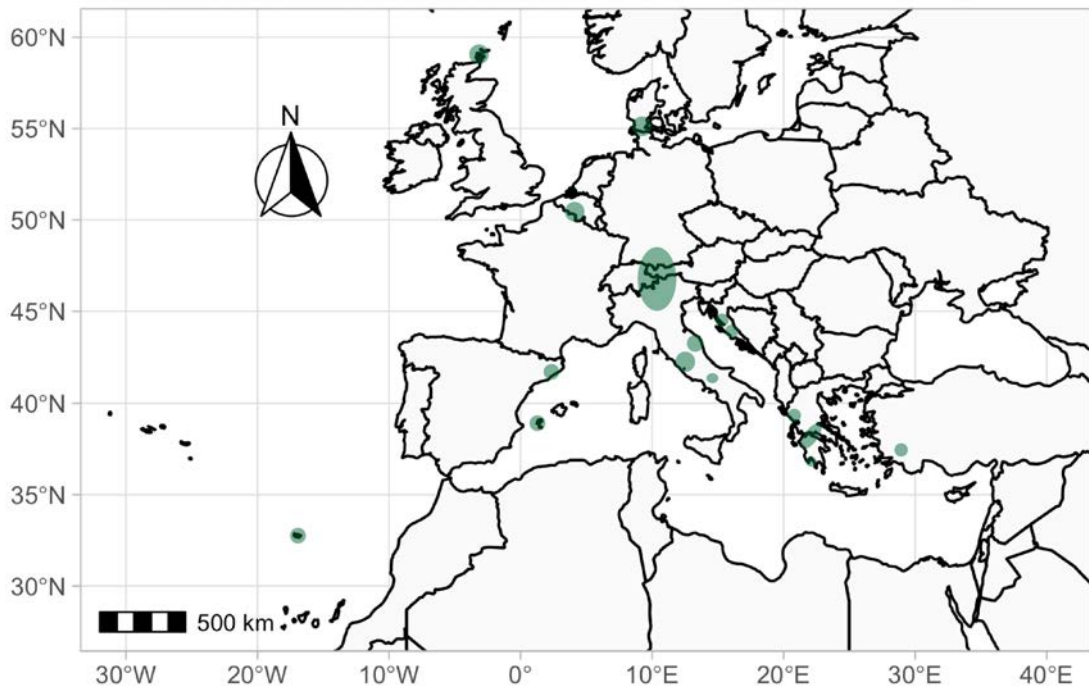
Figure 1.14: Distribution of metacentric wild mice populations in Western Europe. Map showing the locations in which *Mus musculus domesicus* Rb populations have been described, indicated by green circles. Adapted from (Piálek et al. 2005).

in this area. Thus, the BRbS is considered as a polymorphic Rb zone rather than a hybrid zone, where no metacentric race has been yet fixed.

Several studies have shown that Rb fusions can have an impact on the morphology of the BRbS mice, together with structural heterozygosity and local geographical isolation (Martínez-Vargas et al. 2014; Muñoz-Muñoz et al. 2006, 2003). Moreover, mice with Rb fusions present different daily motor activity patterns than standard mice (Sans-Fuentes et al. 2005). Hence, the observed morphological and ethological differentiation of wild populations point out that reduction in gene flow between populations is taking place, highlighting the role of chromosomal fusions in population differentiation. However, the implications of Rb fusions at the genome-wide level remain largely unexplored.

1.3.2.2 The Madeira Rb system

The Madeiran archipelago is located 600 km off the Atlantic coast of North Africa and is of volcanic origin. The archipelago is composed by three island groups: Madeira, Porto Santo, and Ilhas Desertas. The main island, Madeira, presents an extreme topography, with steep mountains that separate narrow valleys, the only habitable areas for humans and their commensals (e.g., mice) (Britton-Davidian et al. 2000; Gündüz et al. 2001a). Evidence suggests that mice was introduced in Madeira by the first Portuguese settlers during the fifteenth century (Mathias and Mira 1992), although molecular studies point to previous colonization in the 9th century most likely brought by Viking expeditions (Förster et al. 2009; Gündüz et al. 2001a).

These human-mediated colonization events followed by the genetic isolation between populations due to the island holography and the genetic drift resulted in high chromosomal diversity (Britton-Davidian et al. 2007; Förster et al. 2013). Hence, six different chromosomal races presenting unique Rb fusions have been described in the main island of Madeira (Britton-Davidian et al. 2000). The location of Rb mice races follows the pattern of human urbanization, where most populations are found in the most western part of the island: PEDC (2n=23-26), PADC (2n=24-28), a contact zone

PEDC/PADC, and PLDB (2n=24). The PSAN race (2n=22) is located at the northeast part of the island, while PPOD (2n=27-28) and PSVI (2n=24) races are found in the center-northern region (Britton-Davidian et al. 2000; Chmátal et al. 2014; Förster et al. 2009; Piálek et al. 2005). A lack of decrease in genetic diversity among Madeiran mice suggests the presence of large founding populations, multiple introductions, and/or high post-colonization expansion rate (Britton-Davidian et al. 2007). Interestingly, standard mice are found in the neighboring island of Porto Santo, which presents less extreme holography, reinforcing the role of geographical barriers in the chromosomal radiation found in Madeira's main island.

1.3.2.3 The impact of Robertsonian fusions on fertility

Robertsonian fusions can have an impact on fertility as they are linked to recurrent miscarriages, infertility and aneuploid offspring in humans (Scriven 2001; Wilch and Morton 2018). This effect has been attributed to the presence of defective chromosome synapsis that can activate meiotic checkpoints resulting in meiosis arrest and even apoptosis (Burgoyne et al. 2009).

Metacentric chromosomes resulting from Rb fusions can be present either in homozygous (both homologous chromosomal pairs are fused) or heterozygous state (i.e., Rb fusion occurs between one homologous of two different chromosomes). This can result in different meiotic configurations (bivalent *versus* trivalents) that can interfere with proper chromosome segregation, impairing efficient gametogenesis and reducing the individual's fertility. It has been suggested that different degrees of synapsis in Rb heterozygous trivalents can have a direct implication in the individual's reproductive fitness in mice (Garagna et al. 2014). Moreover, asynapsed trivalents often associate in a heterologous manner with other trivalents and/or with the sex chromosomes (Berríos et al. 2017; Burgoyne et al. 2009), activating the meiotic silencing of unsynapsed chromosomes checkpoint (MSUC) (Manterola et al. 2009). Also, studies in mice with heterozygous metacentrics have detected infertility related to the number of metacentrics present and the complexity of meiotic chromosomal configurations (Hauffe and Searle 1998; Medarde et al. 2015; Wallace et al. 2002). These metacetrics are known as complex heterozygotes that conform synaptically problematic multivalents (Ribagorda et al. 2019). In the case of the BRbS, the presence of fusions have been related to a reduction in gamete viability leading to hybrid subfertility. This sub-fertility, however, it is not enough to prevent the spread and conservation of Rb fusions in these populations (Medarde et al. 2015, 2013; Sans-Fuentes et al. 2010).

Overall, the contrasting nature of Madeira and Barcelona Rb systems make of these natural populations unique evolutionary models, allowing for the examination of the mechanisms of formation of Rb fusions in nature as well as the implication for genetic diversity and fertility.

OBJECTIVES

Objectives | 2

One of the main interests of our research group is to understand the functional evolutionary plasticity of the higher-order structural organization of the genome and how this is transmitted to the offspring. When this work started, recent advances in the field put into relevance how mammalian genomes were organized within the cell nucleus in interphase. However, how this organization is established and maintained during the formation of germ cells remained largely unexplored.

In this context, the **main objective** of this thesis is to provide an integrated structural and functional framework of the 3D organization of the mouse genome in germ cells, paying special attention to the effect of genomic structural changes in genome folding and recombination. To reach this aim we established the following **specific objectives**:

1. To study the dynamics of the 3D conformation of the genome during spermatogenesis in the house mouse. This objective included two secondary objectives:

   1.1 To develop a reproducible flow cytometry protocol to isolate highly enriched germ cell populations representing all stages of spermatogenesis (pre-meiotic, meiotic and post-meiotic cells). Adressed in section 4.1.

   1.2 To study the chromatin 3D folding during spermatogenesis and its relation to insulator proteins and gene expression. Adressed in section 4.2.

2. To investigate the effect of Rb fusions on the 3D conformation of the genome in the germ line of natural populations of house mice. This objective is adressed in section 4.3 and included two secondary objectives:

   2.1 To analyze chromatin interactions in germ cells of house mice with Rb fusions.

   2.2 To study the effect of Rb fusions on meiotic recombination and genetic diversity.

3. To characterize PRDM9 variability in natural populations of house mice. Adressed in section 4.4.

MATERIALS AND METHODS

3.1 Biological samples

3.1.1 C57BL/6J mice

Regarding the laboratory strain included in this study, 125 male mice (C57BL/6J) aged from 10 to 16 weeks old were obtained from Charles River Laboratories and housed in the animal facility at Universitat Autònoma de Barcelona (UAB) in compliance with the guidelines of the Ethics Committee on Animal and Human experimentation (CEEAH) from UAB (CEEAH 2920).

3.1.2 Wild mice

A total number of 413 wild house mice were live-trapped from two Robertsonian systems: the BRbS and the Madeira system (**Figure 3.1**).
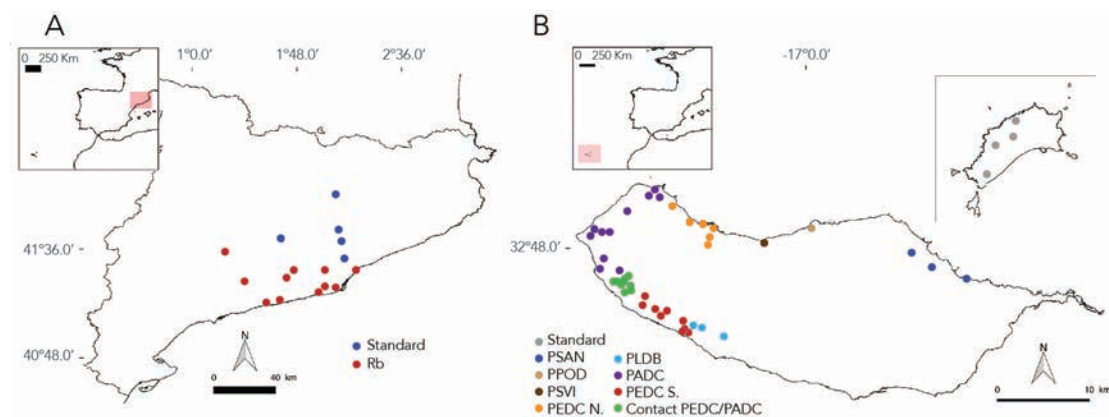
Figure 3.1: Distribution of wild Rb mice populations. (A) Distribution of sampled populations from the BRbS. Blue dots correspond to standard populations (2n=40) whereas red dots correspond to Rb populations (2n=28-39). **(B)** Distribution of sampled populations from the Madeira archipelago. Populations are colored by chromosomal race, as described by (Britton-Davininan et al., 2000). See appendix tables **A.1.4.1** and **A.1.4.2** for detailed information on Barcelona and Madeira samples, respectively. Adapted from (Vara et al. 2019a).

The Barcelona Robertsonian System (BRbS).   Our sample included 201 wild house mice from local farms located in sixteen different populations across the BRbS distribution (**Figure 3.1 A**, **appendix table A.1.4.1**). These included: L'Ametlla de Segarra (2n= 39-40, N= 7), Badalona (2n= 39-40, N=11), Calafell (2n=37-39, N=14), Caldes de Montbuí (2n=40, N=8), Castellfollit del Boix (2n=40, N=18), Olost (2n=40, N=16), Les Pobles (2n=39-40, N=16), El Papiol (2n=37-39, N=7), St. Sadurní d'Anoia (2n=35-39, N=15), Cubelles (2n=32-39, N=11), La Granada (2n=32-37, N=3), El Prat de Llobregat (2n=31-35, N=11), Torreferrussa (2n=40, N=19), Viladecans (2n=31-32, N=24), Vilanova i la Geltr ú (2n=32, N=1) and Castelledefels (2n=28-33, N=22).

In each field trip, around fifty Sherman traps were placed wherever strong evidence (i.e., mice feces) was previously reported by farm owners **(Figure 3.2 A)**. Traps were filled with bread soaked in oil as bait for mice and left overnight. Traps were retrieved the next morning, with an average capture success of 50%. Capture procedure followed the guidelines and ethical approval of the CEEAH from the *Departament d'Agricultura, Ramaderia, Pesca, Alimentació i Medi Natural of the Generalitat de*

Catalunya (SF/934). All captured animals were immediately transferred to the laboratory and culled following protocols from the CEEAH from UAB (CEEAH 2920).

Figure 3.2: Mice sampling. (A) Examples of locations in local farms were Sherman traps were placed. **(B)** Sherman traps with successful captures (left panel) and an example of the captured wild mice (right panel).

From the initial wild-caught animals, 192 of them were used for the PRMD9 study (**appendix table A.1.4.1**). Moreover, a total of 69 animals were used in the recombination study distributed as follows: 45 included in the CO analysis, 34 in the SNPs analysis, 14 in the DSBs analysis, 13 in the SC length and CO/arm analysis and 8 in the centromere analysis (**appendix table A.1.4.1**).

The Madeira Robertsonian System. A total number of 212 wild house mice from the Madeira archipelago (Madeira main island and Porto Santo island) were live-trapped by our collaborators from Dr. Maria L. Mathias' lab, from the Centre for Environmental and Marine Studies (CESAM), in the University of Lisbon (Portugal). Captured animals were handled in compliance with the guidelines and ethical approval of the *Órgão Responsável pelo Bem-Estar dos Animais (ORBEA)*, from the University of Lisbon.

The studied area consisted of thirty-seven different populations that included the six chromosomal races present in the island of Madeira [as described by (Britton-Davidian et al. 2000)] (**Figure 3.1B, appendix table A.1.4.2**): PSAN (2n=22, N=10), PADC (2n=24-28, N=55) PEDC (2n=23-26, N=75), PLDB (2n=24, N=16), PPOD (2n=27-28, N=10) and PSVI (2n=26-27, N=4). Also, standard animals (2n=40, N=11) from 4 localities from the neighbor island of Porto Santo were included in the study (**Figure 3.1 B, appendix table A.1.4.2**).

3.1.3 Cell lines

We used two primary fibroblasts cell lines previously established in our laboratory (Sánchez-Guillén et al. 2015): one derived from a laboratory mouse (C57BL/6J) and another from a Rb mouse from the BRbS (2n=30, Castelldefels).

3.2 Experimental Techniques

All reagents, resources and solution recipes are detailed in the **appendix tables A.1.2 and A.1.3**).

3.2.1 Testicular disaggregation

Mice were dissected and testes placed in a petri dish with Gey's Balanced Salt Solution (GBSS). Once the tunica albuginea was removed, testes were incubated in 10 ml of GBSS with 0.5mg/ml collagenase type II and 1µg/ml of DNase at 33ºC during 15 minutes in constant agitation. Samples were manually mixed 10 times every 5 minutes.

After leaving the seminiferous tubules to deposit for 3-4 minutes, the supernatant was discarded. A second incubation in GBSS with 0.5mg/ml collagenase type II and 1µg/ml of DNase was performed, followed by an incubation in 10 ml GBSS with 0.375 mg/ml of trypsin and 1µg/ml of DNase at 33ºC for 15 minutes in constant agitation. Then, more trypsin was added to the disaggregation solution (final concentration 0.375 mg/ml) and the whole mix was manually homogenized 10 times and incubated another 15 minutes at 33ºC with constant agitation.

After trypsin disaggregation, the mix was pipetted for 3 minutes at RT in order to break cell aggregates and to obtain a homogenized cell suspension. In order to inactivate the trypsin reaction, 500 µl of Fetal Bovine Serum (FBS) were added for each 10 ml of cell suspension and then the solution was filtered through a 70 µm diameter cell strainer. Cells were then centrifuged for 3 minutes at 1,800 x*g* and resuspended in 4 ml GBSS supplemented with 5% of FBS. Once the cell number was estimated using a Neubauer's chamber, cells were centrifuged for 10 minutes at 290 x*g* and resuspended in 1x PBS (1ml/M cells), where formaldehyde was added to a final concentration of 1%. Cells were incubated for 10 minutes in agitation and glycine was added to a final concentration of 0.125 M and incubated with agitation at RT for 5 minutes and then at 4ºC for 15 more minutes. Finally, cells were centrifuged for 10 minutes at 290 x*g* and at 4ºC and resuspended in 3 ml of 1x PBS in case of immediate sorting. Alternatively, cell pellets were stored at -80ºC until used.

3.2.2 Fluorescence-Activated Cell Sorting

A Fluorescence-Activated Cell Sorting (FACS) strategy was used to isolate different germ cell populations from male mice, taking advantage of their different DNA content and level of chromatin compaction.

Briefly, a cell sorter aligns cells in what is known as the fluidic system, generating a flow in which laser beams and cells interact. In our case, the BD Influx^{TM} (BD Biosciences) sorter generates droplets that are coupled with the cells by acoustical coupling in the nozzle assembly (**Figure 3.3**). An excitation light source illuminates cells as they pass through, refracting light in all directions: the forward scatter light (FSC), light diffusion at small angles that is correlated to the relative size of the cells; and the side scatter light (SSC), light diffusion at large angles related with the structural complexity of the cell (Picot et al. 2012). Additionally, cells interact with different laser beams of different wavelengths (e.g. 350 nm, 640 nm), depending on cell population features and the fluorochromes involved in the experiment. Both light diffraction and fluorescence are then analyzed by the electronic system of the flow cytometer, the BD FACS^{TM} in our case, which compiles all the information from the cell flow (number of total events, cell size, morphology, etc). Droplets differently charged according to the cell population analysis are generated, contained each population and sorted accordingly (**Figure 3.3**).

For the purpose of our study, cells obtained in **section 3.2.1** were resuspended in 1x PBS and dyed with 5µg/M of Hoechst 33342 for 30 minutes at 33ºC in constant agitation. Cells were then maintained in ice and filtered through a 30 µm cell strainer just before sorting using the BD Influx^{TM} cell sorter at the Flow Cytometry Core facility from the *Centre de Regulació Genòmica and Universitat Pompeu Fabra*, at the Barcelona Biomedical Research Park. During sorting, FSC and SSC patterns were firstly evaluated in order to discard cell debris (**Figure 3.4 A**). The Hoechst 33342 dye used is excited at around 350 nm and it is detected at two emission wavelengths after UV excitation: Hoechst red (630-650 nm) and Hoechst blue (405-450 nm). Taking advantage of Hoechst properties, from the FSC vs SSC analysis, cells within the gate were then analyzed by 355-DAPI Width and 355-DAPI (laser UV355-460/50), in order to thoroughly assess cell morphology and discard cell aggregates placing a gate accordingly (**Figure 3.4 B**).

Cells within the gate discarding aggregates were then analyzed by 355-DAPI (laser UV355-460/50; discriminating the DNA content of the cells) and 355 (laser UV355-670/30; discriminating cell complexity). Gates were placed accordingly to discriminate cells by DNA content (c, 2c and 4c) and cell complexity (two different populations per DNA content) (**Figure 3.4 Ci**).
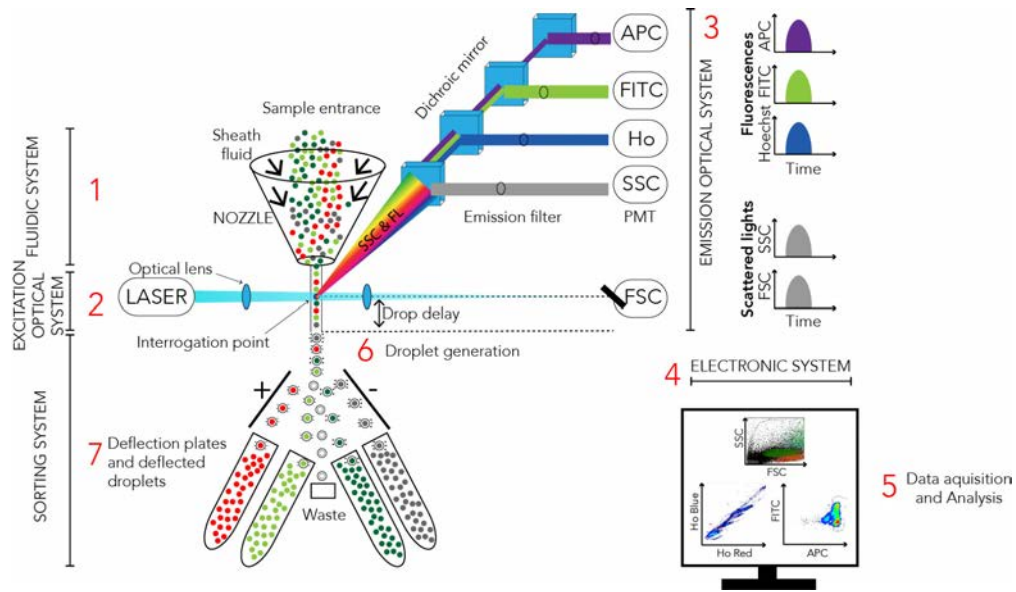
Figure 3.3: Schematic representation of the flow cytometry principle. The sample labeled with different fluorochromes is coupled with the sheath fluid in the nozzle and the fluidic system generates a single-cell stream (1). Then, cells are excited by a laser from the excitation optical system (2) at the interrogation point. Cells scatter the light forward (FSC) and side (SSC) upon their morphology. The side scattered light goes through dichroic mirrors that diffract different wavelengths, so the different fluorochromes used in the experiment (here as an example, Hoechst, FITC and APC are depicted) can be detected by the emission optical system (3). The data goes through the electronic system and cell populations are analyzed based on their light emission patterns and whether are positive or negative for the fluorochromes used (4 and 5) and sorting gates are placed accordingly. Droplets are generated by the principle of drop delay, resulting in individual drops negatively or positively charged, containing the cell of interest (6). Finally, droplets go through an electric field between two deflection plates, sorting cells according to their charge. FL-fluorescence, Ho-Hoechst. Adapted from (Picot et al. 2012).
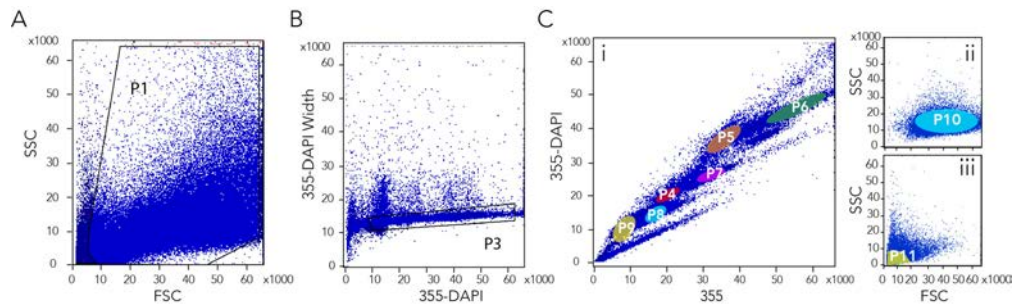
Figure 3.4: Fluorescence-Activated Cell Sorting analysis and gating with the BD FACS^TM Sortware. (A) FSC vs SSC plot that shows cell patterns of light diffraction according cell morphology and complexity. It allows for the distinction between cells with good integrity (gated in P1) and cell debris. **(B)** Hoechst blue emission plot (355-DAPI width vs 355-DAPI) to discard cell aggregates by placing a gate accordingly (P3). **(C)** Sorting of spermatogenesis populations. (i) Hoechst blue (355-DAPI) vs. Hoechst Red (355) plot allows for the distinction of cell populations by their complexity (laser 355, Hoechst red) and DNA content (laser 355-DAPI, Hoechst blue). Gates are placed accordingly to discriminated cell populations: P4-Spermatogonia, P5 and P6-primary spermatocytes, P7-Pairs of round spermatids, P8-Round spermatids and P9-Sperm. (ii) FSC vs SSC plots from gate P8 (round spermatids) and (iii) from gate P19 (sperm). Both populations are re-gated by cell morphology (P10 and P11), to ensure high enrichment and avoid contamination of the neighboring population as they have the same DNA content (c).

3.2.3 Enrichment analysis

For each FACS experiment, we calculated cell enrichment in each isolated population by immunodetection of specific meiotic proteins (see **section 3.2.7**) and by evaluating DAPI morphology. To that aim, a total of 30 µl from each cell population isolated by FACS were placed into a slide. Then, 100 µl of fixative solution I were added, and slides were incubated for at least 20 minutes in a humidified chamber. Next, slides were left to air dry for at least 45 minutes and washed 3 times (2 minutes per wash) with Washing solution I and left to air dry. Immunofluorescence using

antibodies against SYCP3 and γH2AX (as described in **section 3.2.7**) was performed on primary spermatocyte populations, to assess population enrichment by distinguishing the different stages of meiotic prophase I (leptonema, zygonema, pachynema and diplonema). After immunofluorescence, 15 µl of DAPI were added and slides were mounted with a cover slip. Preparations were analyzed under a epifluorescence microscope (Axioskop, Zeiss) coupled with a CCD camera (ProgRes® CS10plus, Jenoptik). At least 100 cells were counted to establish cell type proportions. If enrichment analysis resulted above 70%, the isolated cell fraction was considered of good quality and kept for downstream analyses.

3.2.4 Cell culture and chromosome harvest

Cell lines (**section 3.1.3**) were cultured in DMEM supplemented with 50 µg/ml Gentamycin, 1x Penicillin-Streptomycin, 20% Fetal Bovine Serum and 2 mM L-Glutamine until reaching 70-80% confluence (37 ºC, 5% CO_2).

Modal karyotypes were assessed to ensure that cell lines maintained genome integrity while in culture. For chromosome harvest, 80 µl of colcemid were added for each 10 ml of cell culture and incubated at 37ºC for 4 to 6 hours. Cells were washed with 1x Dulbecco's Phosphate-Buffered Saline (DPBS) and trypsinized with 2 ml of 0.05% trypsin (5 minutes incubation at 37ºC). Trypsin was then neutralized and cells were centrifuged for 5 minutes at 300 x*g*. Subsequently, 5 ml of 0.075 M KCl (previously warmed at 37ºC) were added drop by drop while vortexing the cells. Following, cells were incubated in 0.075 M KCl for 30 minutes at 37ºC in the water bath. After the incubation, cells were centrifuged for 5 minutes at 300 x*g* and 5 ml of fresh carnoy's solution were added drop by drop while vortexing. Both centrifugation and carnoy fixation steps were repeated three times and chromosome suspension was stored at -20ºC until used.

Chromosome spreads were prepared by placing a drop of chromosome suspension on a superfrost slide. Slides were left to air dry and store at -20ºC until used. For the assessment of modal karyotype, Giemsa staining was performed on chromosomal preparations. Slides were aged for 2 hours at 65ºC. Once back at RT, slides were placed in trypsin working solution for 15-30 seconds and then rinsed with DPBS. Slides were then stained with Giemsa (4 ml of Giemsa stain and 50 ml of Sorensen's buffer) for 7-10 minutes and rinsed with tap water. Slides were then analyzed under the optical microscope (model Zeiss Axioskop) by counting the number of chromosomes per metaphase.

3.2.5 *In nuclei* Hi-C

The *in nuclei* Hi-C method is a chromosome conformation capture technique that allows the detection of chromatin interactions in the cell nucleus (Belton et al. 2012; Dekker et al. 2002). As a result, the probability of interaction between close-by genomic loci can be assessed at genome-wide scale, allowing for the reconstruction of genome architecture from a specific cell type. This technique gathers populational information from pooled cells, highlighting the importance of a homogeneous cell pool for this type of experiments. For the study of the dynamics of the 3D genomic structure during spermatogenesis, we adapted the *in situ* Hi-C protocol from (Rao et al. 2014) (**Figure 3.5**).

As the method is overall complex four controls are used in each *in nuclei* Hi-C experiment to ensure high-quality results. These include:

- ► **Non-Digested (ND) sample**: control of chromatin integrity.
- ► **Non-Ligated (NL) sample**: control the digestion efficiency.
- ► **3C sample**: to monitor the efficiency of labelling and ligation, it will be ligated directly after digestion (no end-repair).
- ► **Hi-C sample**.
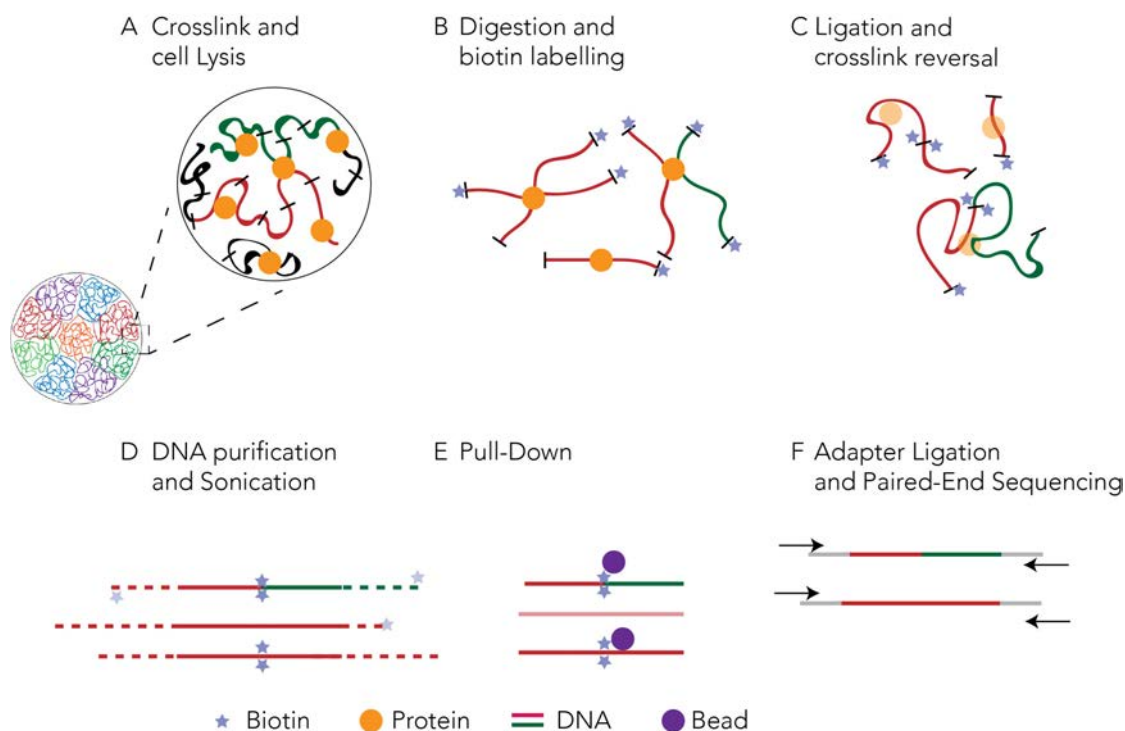
Figure 3.5: *In nuclei* **Hi-C method. (A)** Cells are fixed via crosslink and lysed. **(B)** Cell nuclei are digested with the restriction enzyme MboI and the generated blunt ends labelled with biotin. **(C)** Digestion ends are ligated, and the crosslink reversed with proteinase K. **(D)** The resulting DNA is purified and sonicated. **(E)** During library preparation, DNA fragments with biotin are recovered by a pull-down with streptavidin beads. **(F)** Adapters and indexed primers are ligated to DNA fragments by PCR and resulting libraries are paired-end sequencing.

3.2.5.1 Crosslinking and cell lysis

Cells were fixed in order to preserve the structure as it was *in vivo*, maintaining protein-DNA interactions. In the case of the germ line, cells were crosslinked as described in **section 3.2.1** during testicular disaggregation. As for adherent fibroblast cultures, cells were washed with 1x PBS. Then, cells were fixed with 1% formaldehyde in 1x PBS and incubated for 10 minutes at RT. Glycine was added at a final concentration of 0.125 M and incubated 5 more minutes at RT. Cells were again incubated for 15 minutes at 4ºC and then washed with 5 ml of 1x PBS twice. Following, 2 ml of trypsin 0.05% were added to the culture to ease cell detachment and incubated for 8 minutes at RT. Cells were then washed with 1x PBS twice and 1 ml of 1x PBS was added. Then, cells were collected and centrifuged for 5 minutes at 1,800 x*g*. Cells were snap-frozen in liquid nitrogen and stored at -80°C until used.

Once cells were crosslinked, they were lysed carefully to obtain clean cell nuclei without compromising their integrity. To attain that, Lysis Buffer was added to each crosslinked aliquot and cells were incubated on ice for 30 minutes and then centrifuged for 5 minutes at 1,800 x*g*. After two 1x NEB2 washes, cells were gently resuspended in 1x NEB2 with 0.5% SDS at RT to ensure complete cell lysis and incubated for 10 minutes at 65ºC with rotation. Then, 1x NEB2 with 3% Triton X-100 was added and cells were incubated for 30 minutes at 37ºC. Cells were then centrifuged 5 minutes at 1,800 x*g* (4ºC) and washed with 300 µl of 1x NEB2 twice.

3.2.5.2 Restriction digest and biotin labelling

DNA was digested using the 4-cutter restriction endonuclease MboI, while keeping a ND control that consisted of an aliquot from the main sample (10% of the sample). The ND control was topped

up to 100 µl with 1x NEB2 and incubated at 37ºC together with the digested sample. Both ND control and chromatin samples were digested O/N at 37ºC in rotation.

To check chromatin integrity at this stage of the protocol, 5 µl from the digested samples were taken for digestion controls (**D**) and topped up to 100 µl with 1X NEBuffer 2. Following, 10µl of proteinase K (10 mg/ml) were added to the **ND** and **D** controls and incubated for 45 to 60 minutes at 65ºC. In parallel, 2 µl more of MboI were added to the main sample and left at 37ºC until chromatin integrity control finalized. After proteinase K incubation, 200 µl of H_2O were added and samples were purified with Phenol: Chloroform. The quality of the sample was checked by running it on 0.8% agarose gel, were digested samples presented a homogeneous smear while ND controls presented an integral DNA genomic band.

After digestion, the resulting DNA overhangs are labelled with biotin. For that, first an aliquot from the main sample was taken for the **3C control**, which was kept from being repaired (biotinylated). The main sample was resuspended directly with the Reparation Mix and incubated for 45 minutes at 37ºC, then for 10 minutes at 65ºC and finally centrifuged for 5 minutes at 1,800 x*g*.

3.2.5.3 Ligation and crosslink reversal

Once labelled with biotin, hybrid molecules are ligated. Samples (main **Hi-C** sample and **3C** controls) were resuspended with ligation buffer and incubated at 16ºC, 4 hours or O/N with mixing. Samples were centrifuged for 5 minutes at 1,800 x*g* and resuspended in 400 µl of 1x NEB2. A total of 10 µl of 10 mg/ml RNAse A were added and samples incubated 15 minutes at 37ºC. Then, 20 µl of 10 mg/ml proteinase K were added and samples incubated at 65ºC O/N to reverse the crosslink.

3.2.5.4 DNA purification and sonication

After crosslink reversal, samples were cooled to RT and purified with Phenol: Chloroform. The purified DNA was resuspended in 100 µl of nuclease-free H_2O. Electrophoresis (gel 0.8%) of the ND, D (non-ligated) and 3C controls together with an aliquot of the main Hi-C sample was run to check that the protocol was successful (**Figure 3.6**): the ND control presented as an integral genomic DNA band while NL control appears as an homogeneous smear. The 3C control presented recovery of the integral band or a largely reduced smear, whereas in the Hi-C sample, due to the biotin labelling, ligation is less efficient, and a smear is detected, but reduced from the digested control. Finally, DNA content of the main Hi-C sample was measured with Qubit.
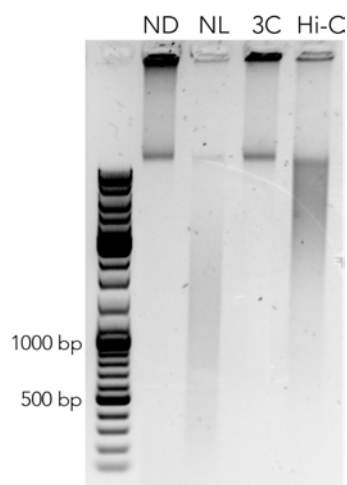
Figure 3.6: Hi-C control gel. Agarose gel (0.8%) were all the controls and the final Hi-C product are loaded. The non-digested control (ND) lane shows an entire genomic DNA band whereas the non-ligated (NL) lane shows a smear of DNA resulting from digestion. The 3C control lane shows a successful ligation, with an entire DNA band. Ligation is less successful in the Hi-C lane, due to the presence of biotin in the ends, but the smear is reduced from the NL control, pointing that effective ligation has taken place also in the Hi-C sample.

Approximately 3µg of sample were used as starting material for next-generation sequencing library preparation. Samples were sonicated per run of sonication in 100 µl of final volume. The sonication program for Biorruptor Pico was: 30″ time ON, 30″ time OFF, 10 cycles. After sonication, the different

tubes containing the same sample were pooled together. Electrophoresis gel (1.2% agarose) was run to check fragment size resulting for sonication, to ensure optimal size for Hi-C libraries (most fragments around 300-500 bp).

3.2.5.5 Pull down

All fragments containing biotin are recovered in this step by using streptavidin magnetic beads that bind biotin molecules. Dynabeads MyOne Streptavidin T1 beads were washed with 400 µl of 1x binding buffer (BB)+Tween. Beads were washed then with 400 µl 1x BB, twice. Following, beads were resuspended with 300 µl of BB 2x and 300 µl of sample are added to the mix. Beads with the sample were incubated for 30 minutes under rotation (rotating wheel), at RT, for the biotin binding to the beads. Then beads were washed with 400 µl of BB, twice and resuspended with 100 µl of 1x NEB T4 DNA ligase.

3.2.5.6 Adapter ligation and paired-end sequencing

Biotin is removed from un-ligated ends, generating blunt ends which are subsequently repaired. Accordingly, the beads containing the Hi-C resulting DNA fragments with biotin, were resuspended with the end repair mix. Samples were incubated for 30 min, at RT. Following, the beads were washed with 400 µl 1x BB, twice and equilibrated with 100 µl of 1x NEBuffer 2. After the end repair, the beads were resuspended with 100 µl dATP attachment master mix and incubated at 37ºC for 30 min; in this step As are added to 3' ends. Beads were then washed with 400 µl of 1x BB, twice and resuspended with 100 µl of 1x NEB Quick ligation buffer. Hi-C DNA chimeric fragments were then ligated with adapters that will be amplified with specific primers in order to be deep-sequenced.

Following the library production steps, beads were resuspended with 50 µl of Adapter ligation mix an incubated for 15 minutes at RT. A total of 3µL of USER enzyme were added and incubated for 15 minutes at 37ºC for adapters to be accessible to primes. Then, the beads were reclaimed and washed with 400 µl of 1x BB, twice. Beads were resuspended in 50 µl MilliQ and kept until the final amplification.

Then, Hi-C fragments were amplified with specific indexed primers that allowed for sequence identification in next-generation sequencing. To that aim, aliquots of 5 µl of beads were used as starting material for PCR. The PCR mix was prepared (5 µl beads library sample + 2.5 µl H_2O, 2.5 µl indexed primer 10 µM, 2.5 µl universal primer 10 µM, 12.5 µl cocktail master mix NEB) and set (98ºC 30 sec, 98ºC 10 sec, 60ºC 30 sec x 8 cycles, 72ºC 30 sec, 72ºC 5 min). PCR product size was checked by running an electrophoresis (agarose 1.2%).

After fragment amplification by PCR, a final purification by fragment size is performed. Accordingly, 0.8X amount of AMPure beads were added to the library and incubated for 10 minutes at RT. Beads were then washed twice with 700 µl of 70% fresh ethanol without mixing and kept for the remaining ethanol to evaporate. Following, beads were eluted with 30 µl of Tris Buffer, gently mixed by pipetting, and incubated at RT for 5 min. Samples were separated on a magnet from the beads and transferred to a fresh tube. A final electrophoresis gel (1.2% agarose) was performed to ensure that the library fragments selected have the expected size. Alternatively, fragment size could also be checked using Bioanalyzer (Agilent). Final DNA yield was measured with Qubit. Hi-C libraries were sent to *Centre d'Anàlisi Genòmica (CNAG)* and Illumina sequenced, paired-end 75 bp each side on HiSeq 2500, v4).

3.2.6 Spermatocyte spreads from frozen testis

Mice testicular biopsies were obtained immediately after animal dissection and maintained at -80°C until used. A cell suspension was obtained by scattering a slice of frozen testis in a petri dish, in 200 µl of cold 1x PBS. The suspension was left on ice for 5 minutes in order for non-dissociated tissue fibers to precipitate. Following, the suspension was disposed onto superfrost slides (20-40 µl per slide) and 90 µl of lypsol 1% were added to the slides, for cells to swell. Slides were incubated for 16 minutes in lypsol in a humidified chamber. Then, 100 µl of Fixative solution I were added and incubated for 2 hours. Afterwards, slides were left to air-dry for at least 45 minutes and then washed three times (2 minutes each) with Washing solution I and left to vertically air-dry. Slides were mounted with DAPI counterstaining. The quality of the spreads (cell number and morphology) was verified with an epifluorescence microscope (Axioskop, Zeiss) and stored at -20°C until used.

3.2.7 Immunofluorescence

The immunofluorescence (IF) technique allows the identification of proteins of interest by using different sets of primary antibodies that recognize protein epitopes. Such epitopes are subsequently revealed with fluorescent secondary antibodies. In this work, different proteins involved in meiotic division process were analyzed on mouse spermatocytes (see **appendix table A.1.1**), using two different approaches: immunofluorescence on slides and "in solution" immunofluorescence.

3.2.7.1 Immunofluorescence on slides

Labelling of specific meiotic proteins by immunofluorescence was used for the post-FACS enrichment analysis and for all cytological studies included in the study. Firstly, slides were washed in milliQ water for 10 minutes in agitation in order to remove the coverslip. In the case of MLH1, a specific step was performed before primary antibody incubation. The antigen retrieval step consisted in incubating the spermatocyte spreads in sodium citrate buffer at 95-100°C for 40 minutes. Afterwards, the slides were left to cool down at room temperature for 20 minutes. Then, slides were blocked for 10 minutes by washing them twice for 5 minutes in agitation with blocking solution (PBST 0.05%). Primary antibodies were diluted in Blocking solution I according to each antibody specific requirements (see **appendix table A.1.1**) and 100 µl of the primary antibody solution were applied to the slides and incubated O/N in a humidified chamber at 4°C. Following, slides were washed again and secondary antibodies (see **appendix table A.1.1**) were all diluted 1:200 in blocking solution and 100 µl of the dilution applied. Then, slides incubated from 45 minutes to 1 hour in a humidified chamber at 37°C. After incubation, slides were washed and mounted with 20 µl of DAPI.

3.2.7.2 'In solution' immunofluorescence

Labelling of specific meiotic proteins by 'in solution' immunofluorescence was used in FACS experiments to distinguish primary spermatocytes in early prophase (leptonema/zygonema) from those in late prophase (pachynema/diplonema). Briefly, the cell suspension obtained in **section 3.2.1** was centrifuged at 1,800 x*g* for 5 minutes and then resuspended in Blocking solution I, with a proportion of 100 µl per million cells. Cells were then incubated at 37°C in agitation in the thermomixer for 5 minutes twice. Following, cells were distributed in different aliquots up to 6-7 M cells per tube (600-700 µl) and primary antibodies were added. The cell suspension was incubated at 4°C O/N. Next, cells were centrifuged at 1,800 x*g* for 5 minutes and resuspended in **block solution I**, 100 µl per million cells. Cells were then incubated at 37°C in agitation in the thermomixer for 5 min. Centrifugation and cell incubation in block solution were repeated and after, cells were resuspended in the according volume of blocking solution (100 µl/M cells) and the secondary antibodies were added. The suspension was then incubated for 35 minutes at 37°C and in agitation

in the thermomixer. Afterwards, cells were centrifuged at 1,800 *xg* for 5 minutes and resuspended in 100 µl of Blocking solution I per million of cells. Cells were incubated at 37 ºC and in agitation for 5 minutes twice. Following, each tube was resuspended in 1 ml of PBS and 5 µl of Hoechst 33342 (1:10 dilution from 10 mg/ml stock) was added to each tube and cells incubated at 33ºC for 15-20 minutes in agitation. Cells were maintained at 4ºC until FACS.

3.2.8 Three-dimensional (3D) Fluorescence in situ hybridization (FISH) on fibroblasts and germline cells

3D FISH was used to experimentally test differences in chromatin condensation state by measuring the physical distances between loci located in A (open accessible chromatin) and B (closed, more condensed chromatin) compartments.

3.2.8.1 Design of *locus*-specific probes

Based on Hi-C interaction maps data (see **section 4.2**), specific pairs of loci from chromosomes 12 and 14 that remained stable in either from A or B compartments in fibroblasts, spermatogonia and round spermatids were chosen. Using the mm10 mouse genome version available at the *Ensembl database* (Zerbino et al. 2017), Bacterial Artificial Chromosomes (BACs) covering specific genomic regions were selected. Sequence homology was analyzed with the Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990) to ensure probe specificity to the region of interest (**Table 3.1**).

Table 3.1: BACs combinations used in 3D FISH to label specific regions of chromosomes 12 and 14 known to be in either A or B compartments.

BAC	Chr	Compartment	Start position	End Position
RP23-218K11	14	A	24,101,846	24,278,172
RP23-157K12	14	A	24,344,637	24,545,440
RP23-93L3	14	A	25,993,682	26,147,325
RP23-118G18	12	A	72,372,756	72,569,214
RP24-92G22	12	A	73,333,062	73,572,222
RP23-318P2	12	B	36,438,632	36,644,697
RP23-351M10	12	B	36,748,530	36,944,925
RP23-34D20	12	B	37,662,161	37,864,768
RP23-382L21	12	B	62,887,606	63,091,331
RP23-304F23	12	B	63,936,423	64,142,896

3.2.8.2 Probe labelling by Nick Translation

BACs culture and plasmid DNA extraction were performed using the Qiagen Midi Plasmid kit. DNA probes were labelled by a nick translation reaction following instructions from the manufacturer: X µl for 1 µg extracted DNA (X=volume depending on DNA concentration), (17.5-X) µl nuclease-free water, 2.5 µl dUTP (Dig or Cy3), 5 µl 0.1 mM dTTP, 10 µl dNTP mix, 5 µl 10X Nick Translation Buffer, 10 µl Nick translation enzyme). The resulting mix was briefly centrifuged and vortexed and then incubated for 8-16 hours at 15ºC. The reaction was stopped by heating it at 70ºC for 10 minutes. Finally, the mix was chilled on ice until further use.

Once the probe is labelled, at least 500 ng were precipitated with cot-1 DNA (1mg/ml) to block repetitive sequences, salmon sperm (10 mg/ml) as carrier DNA, AcNa (0.1 vol) and EtOH 100% (3 vol). Precipitation for single FISH included just one probe whereas for double FISH, two differently labelled probes were precipitated together for at least 2 hours at -20ºC or 1 hour at -80ºC.

The nick translation precipitation product was centrifuged at 16,000 x*g* for 15 minutes and then washed twice with EtOH 70%. The DNA pellet was left to dry for about 5-10 minutes at 37ºC in the thermoblock before adding 14 μl of Hybridization Buffer; then the pellet was resuspended with the Hybridization Buffer and incubated for approximately 10 minutes at 37ºC in agitation.

3.2.8.3 Slide preparation and hybridization

In the case of germ cells, a cell suspension obtained in **section 3.2.1** was placed on a coverslip at 37ºC for 1 h. Cell attachment was checked under the microscope. Following, cells were incubated in 0.3x PBS for 40 sec. Then, cells were fixed in 4% PFA/0.3x PBS for 10 minutes at RT and washed in 1x PBS at RT for 3x 5 min. Cells were then incubated in 0.5% Triton X-100/1x PBS at RT for 20 min. In the next step, cells were transferred to 20% glycerol/1x PBS and incubated at RT for at least 30 min. Afterwards, cells were frozen by dipping the slide into liquid nitrogen (30 sec) and thawed. The freezing/thawing step was repeated four times. Cells were washed once again in 0.05% Triton X-100/1x PBS for 3x5 minutes and incubated in 0.1 N HCl for 5 min. Slides were washed in 2x SSC for 2 minutes and then incubated in 50% formamide (pH 7.0)/2x SSC at RT for at least 1 hour before proceeding with hybridization or optional pepsin digestion.

As for adherent cells, around 50,000 fibroblasts (**section 3.1.3**) were sub-cultured on a sterilized slide, placed in a sterile petri dish and incubated for at least 2 hours at 37ºC, for cells to adhere to the surface. Pre-warmed medium at 37ºC was added, covering all slide surface, and cells were left to grow O/N to desired confluence. Then, slides were washed with 1x PBS three times in the petri dish and then fixed by adding 1ml of 4% PFA in 0.3x PBS on the slide for 10 minutes. Slides were washed 3 times with 1x PBS in order to remove fixative residues. Following, slides were incubated in 0.5% Triton X-100/1x PBS for 15 minutes and then in 20% glycerol/1x PBS for at least 1h. Cells were frozen by immersing the slides in liquid nitrogen for 30 seconds. Once the ice layer disappeared, slides were immersed in 20% glycerol/1x PBS for 1 minute. This step was repeated 4 times. Slides were washed in 1x PBS 3 times for 5 minutes and then incubated in 0.1N HCl for 8 min. Then, slides were washed in 2x SSC 2 times for 3 minutes and then incubated in formamide 50% (pH 7)/2x SSC for at least 1 hour before proceeding with the pepsin treatment and hybridization.

In all cases (germ cells and fibroblasts), slides were equilibrated in 2x SSC at RT for 2 minutes and in 1x PBS at RT for 3 min. Then, slides were incubated in pepsin (0.005% in 0.01 N HCl) for 3 to 5 minutes, depending on the cell type to allow probe penetration. Slides were then incubated in 1x PBS/50 mM MgCl$_2$ to inactivate pepsin, at RT for 2x5 min. Slides were fixed in 1% paraformaldehyde/1x PBS, at RT for 10 minutes and washed in 1x PBS, at RT for 5 minutes and in 2x SSC for 2x5 min, then returned to 50% formamide/2x SSC for at least 1 hour before hybridization.

A total of 14 μl of DNA probe mix were added to the slide and placed on a hot block at 75ºC to denature cellular and probe DNA for 2 min. Hybridization was performed in a humid chamber at 37ºC in the stove for at least overnight or preferably for 2–3 days.

After hybridization, slides were transferred to 2x SSC. Slides were then washed in 2x SSC at 37ºC, shaking for 3x 5 minutes and then washed in 0.1x SSC at 60ºC, shaking for 3x5 min. Then, slides were briefly rinsed in 4x SSC/0.2%Tween (Washing solution II) and blocked in blocking solution II at 37ºC for 10–15 min.

In the case of indirect probe detection by using dUTP-Dig, correspondent antibodies had to be used to detect the probes. The required antibodies (anti-DIG-FITC) were diluted to the appropriate working concentration (1:150) in 4x SSC/0.2%Tween + 1% BSA. Slides were incubated with the antibody in a dark humid chamber at 37ºC for 45 min. Then, slides were washed in 4x SSC/0.2%Tween, shaking for 3x3 min. This last antibody detection was unnecessary in case of direct probe detection with dUTP-Cy3. Finally, slides were mounted with DAPI and stored at -20ºC until use.

3.2.9 Image processing and analysis

Immunofluorescence on spermatocyte preparations were visualized using a Zeiss Axioskop epifluorescence microscope equipped with the appropriate filters and a charged coupled device camera (ProgRes® CS10plus, Jenoptik) and captured with ACO XY (A. Coloma, Open Microscopy). Images were processed using Photoshop and ImageJ.

As for 3D FISH experiments, preparations were visualized with the confocal microscope Leica SP5 and captured with the LAS X Life Science software. Captures were made every 0.34 µm, getting between 20-40 stacked images per cell. Confocal images were then processed and analyzed with the IMARIS software. With IMARIS, for each analyzed cell, a surface was established according the DAPI signal corresponding to the cell nucleus, and a dot was assigned for every probe signal (red or green) (**Figure 3.7**). Following, the physical distance was measured with the "Compute distance between surfaces and spots" MATLAB based function integrated in the software together with the volume of the nucleus from the surface covering the nucleus (**Figure 3.7 B**).
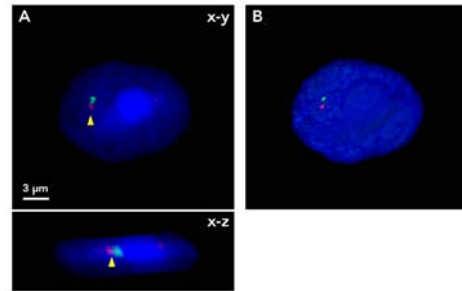
Figure 3.7: Example of round spermatid after double 3D fish with probes RP23-218K11 (green) and RP23-157K12 (red). (A) Confocal image of a round spermatid in the x-y plane (upper panel) and x-z plane (lower plane) were the nucleus is stained in DAPI (blue) and probe signals are detectable in red and green (color respective to each probe used). **(B)** IMARIS recreation of the cell, with a surface representing the cell nucleus (blue) and two dots representing the probes RP23-218K11 (green) and RP23-157K12 (red) signals. Adapted from (Vara et al. 2019b).

Regarding the analysis of COs, the Micromeasure 3.3 software (Reeves 2001) was used to measure the distance between COs (MLH1) in pachynema, together with measures of the synaptonemal complex length (SYCP3 signal). Briefly, each MLH1 focus position was determined relative to the distance from the centromere, accounted as a percentage of the total synaptonemal complex length, as previously described (Ruiz-Herrera et al. 2017; Segura et al. 2013).

3.2.10 Western blotting

The insulator protein CTCF and meiotic cohesins RAD21L and REC8 were detected by western blot using appropriate antibodies (see **appendix table A.1.1**). Western blotting was essential to validate the antibodies' antigen detection efficiency to further use them in for ChIP-seq experiments.

For western blot experiments, protein extracts from fibroblasts were used as controls. Moreover, proteins from mouse testis were extracted, to check antibody specificity for CTCF (around 120 Kda), RAD21L and REC8 (both around 63 Kda).

3.2.10.1 Protein extraction and quantification

In the case of cell lines, cultured cells were centrifuged at 130 x*g* for 10 min and then placed on ice and the supernatant discarded. Cells were then resuspended in 1 ml of ice-cold PBS and centrifuged at 130 x*g* for 10 min. Then, the cell pellet was resuspended in ice-cold RIPA buffer (complemented with 20 µl PIC and 10µl NaF per ml of the buffer) and transferred to a 1.5 ml eppendorf tube. The mix was shaken for 30 minutes in ice and then centrifuged at 16,000 x*g* for 20 minutes at 4ºC. The supernatant containing the proteins was recovered and transferred to a fresh tube.

As for mouse testis, a volume of 300 µL of ice-cold RIPA buffer was added for each 5 mg tissue and homogenized with a pestle. Additional 300-600 µL of lysis buffer were added during homogenization

and the solution was incubated for 2 hours at 4°C in agitation. Tubes were then centrifuged at 16,000 x*g* for 20 minutes at 4°C. The supernatant was collected in a fresh tube and placed on ice.

In all cases, the protein extracts were stored at -80ºC to further use if protein quantification was not performed in the same day. Protein quantification was performed following manufacturer's instructions from Pierce's BCA protein kit assay.

3.2.10.2 Western blotting

First of all, an 8% Running gel mix was prepared and once polymerized, a 4% stacking gel mix was poured on top of the running gel and left to polymerize for 5 min. Parallelly, protein extracts were thawed to RT and mixed as follows: sample volume (around 15 µg of protein) + same volume of 2x Laemmli Buffer (with ß-ME) =30 µl (well total volume). Samples with Laemmli were heated up to 95ºC for 5-10 minutes for them to denaturalize and then left to cool down at RT.

Both the protein marker (10 µl) and samples (30 µl per sample and well) were loaded to the gel and then run in 1x running buffer, firstly at 80 V and then at 100-120 V until samples reached the desired degree of separation. Following, proteins from the gel were transferred to a polyvinylidene fluoride (PVDF) membrane with the Trans-Blot® Turbo system from Bio-Rad (2.5 A constant up to 25 V, for 10 minutes). Protein load transferred to the membrane was checked by incubating the membrane in Ponceau staining solution for 5 min. Ponceau staining was removed from the membrane with TBST solution washes and then placed in blocking solution for an hour with agitation.

In the protein detection step, anti-rabbit CTCF was used at 1:2500 dilution; and both anti-rabbit REC8 and anti-rabbit RAD21L were used at 1:2000 dilution. In parallel, anti-rabbit ß-tubulin was used at 1:5000 dilution as control. The membrane was incubated in blocking solution III containing the diluted primary antibodies at 4ºC in agitation O/N and then washed 3 times for 5 minutes in agitation with TBST.

For the antibody detection step, the membrane was incubated in blocking solution III with anti-rabbit HRP-PO at 1:15000 dilution, at RT for 1 hour and 30 minutes in agitation. Following, the membrane was washed 3 times for 5 minutes in agitation with TBST and finally incubated for 5 minutes with the ClarityTM Western ECL substrate and revealed with Versadoc^{TM}.

3.2.11 ChIP-seq

Once the antibodies were validated, they were used for chromatin immunoprecipitation followed by sequencing (ChIP-seq). ChIP-seq is a method that detects binding events between DNA and the protein of interest, recovering those chromatin fragments that were bound to a specific protein (Furey 2012) (**Figure 3.8**). We applied this technique with the aim to detect the binding regions of CTCF and meiotic cohesins RAD21L and REC8 in mouse germ cells.
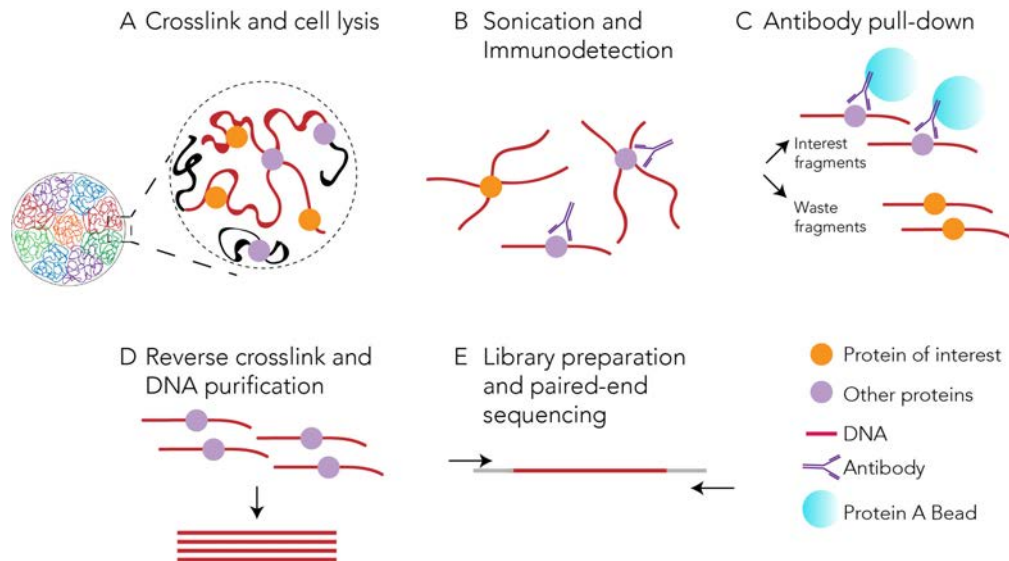
Figure 3.8: ChIP-seq method. (A) Cells are crosslinked and lysed. **(B)** After sonication, DNA fragments are incubated with a specific antibody that will bind to the protein of interest. **(C)** DNA fragments bound to the protein of interest with the antibody are pulled down with protein A beads (immunoprecipitation). **(D)** The crosslink is reversed, and the DNA purified. **(E)** Sequencing libraries are prepared from ChIP resulting fragments to finally be sequenced.

3.2.11.1 ChIP-seq protocol

A cell pellet containing at least 10 million cells obtained from testis disaggregation (**section 3.2.1**) was resuspended in Lysis Buffer I (100µl/ M cells). From this, a 100 µl aliquot (equivalent to 1 M cells) was taken apart to use as input and chromatin integrity check. Samples were incubated 10 minutes on ice and subsequently stored at 4ºC O/N. Proceeding with the chromatin integrity check, the 1 M cell aliquot was centrifuged for 7 minutes at 1,800 x*g* at 4ºC and the cell pellet resuspended in 100 µL/1 M of Lysis Buffer II and incubated for 10 minutes on ice. Then, the input sample was added to a sonication tube and topped up to 100 µL with lysis buffer II and sonicated (Biorruptor Pico): 30" ON, 30" OFF, for 10 cycles to obtain 200-300 bp fragments. After sonication, the solution was left on ice for around 60 min, to allow SDS precipitation.

Lysates were then centrifuged for 10 min, at 16,000 x*g* at 4ºC and the supernatant was recovered and kept on ice. 40 µL of the supernatant were taken and 360 µL of Lysis buffer II+10 µL Proteinase K (10mg/ml) added. The mix was incubated O/N at 65ºC in the thermomixer with shaking. The remaining 60 µl were kept as ChIP-seq input. After the proteinase K incubation, phenol extraction was performed, and samples resuspended in 15 µL of water, quantified by Qubit and sonicated fragments size-checked in a 1.2% agarose gel. Once chromatin integrity control resulted positive, main samples left at 4ºC were processed as described for the control. Following, both the main sample and the input were resuspended in IP buffer up to 1.2 ml and 5-10 µg of the antibody were used (for approximately 25 µg of chromatin) per sample. After pre-cleaning, the antibody was added to the main sample and incubated O/N at 4ºC in a rotating wheel.

A volume of 42µl of beads (Unblocked Protein A beads) for antibody-chromatin pull-down were used for the immunoprecipitation, diluted in 500 µl of 1x cold IP buffer. Beads were centrifuged for 2 minutes at 1,800 x*g* and 4ºC and resuspended with 1x cold IP buffer, adding BSA to 1% final concentration to block the beads. Then, beads were incubated for 15 minutes at 4ºC in a rotating wheel and then centrifuged for 2 minutes at 1,800 x*g* and 4ºC. Following, beads were resuspended in 1x cold IP buffer and then centrifuged again for 2 minutes at 1,800 x*g* and 4ºC to finally be transferred to the tube containing the main sample. The mix was incubated for 3 hours at 4ºC in the rotating wheel and beads were washed three times with 1x cold IP buffer and twice with 1x cold TE buffer (pH=8.0).

The antibody-chromatin fragments were then eluted using 400 µL of freshly prepared elution buffer and incubated for 25 minutes at RT in a rotation wheel. Beads were centrifuged for 3 minutes at 1,800 x*g* at RT, recovering the supernatant containing the antibody-chromatin fragments. The supernatant was then transferred to a new tube and NaCl was added to a final concentration of 200 mM (16 µl NaCl 5M/400 µl eluate), to precipitate proteins. The precipitation mix was incubated O/N at 65ºC.

The protein digestion solution was added to each precipitated sample and incubated for 1 hour at 45ºC. Then, phenol extraction was performed, and samples were resuspended in 20 µL of Tris pH=8.0. To ensure DNA dilution, samples were incubated for 10 minutes at 37ºC.

3.2.11.2 Library preparation

Ends from pulled-down fragments were repaired by adding the end repair solution (final volume of 50 µl) and incubated for 30 minutes at RT. Samples were purified with 1x AMPure beads, as described in **section 3.2.5.6** and eluded with a final volume of 25 µl. Then, the A-tailing solution (final volume of 50 µl) was added to the samples and incubated for 30 minutes at 37ºC. Again, samples were purified with 1X AMPure beads, as described in **section 3.2.5.6** and eluded with a final volume of 25 µl. In the next step, the adapters ligation mix (final volume of 50 µl) was added to the samples and incubated for 15 minutes at RT. 3 µl of the USER enzyme were added to the mix and incubated 15 minutes more at 37ºC. As in the other steps, samples were purified with 0.8x Ampure beads, as described in **section 3.2.5.6** and eluded with a final volume of 25 µl.

Finally, the ChIP-seq library product was amplified by PCR, being the PCR mix the following: 20 µl sample, 2.5 µl primer i5 (universal primer), 2.5 µl primer i7 (indexed primer), 25 µl NEBNext High-Fidelity 2X PCR Master Mix. Then the PCR was performed using the following program: 98ºC for 30 sec, 98ºC for 10 sec, 65ºC for 30 sec, 72ºC for 30 sec (for 14 cycles) and 72ºC for 5 min. Samples were purified with 0.8X Ampure beads, as described in **section 3.2.5.6** and eluded with a final volume of 20 µl. Final DNA amount of library products was measured in Qubit and the fragment size checked either by an electrophoresis gel (1.2% agarose) or with BioAnalyzer. Libraries were sent to *Centre d'Anàlisi Genòmica* (CNAG) and high-throughput Illumina sequenced (Hi-seq 2500).

3.2.12 RNA-seq

RNA-seq experiments where performed by our collaborators from CNAG (*Centre d'Anàlisi Genòmica*) Dr. Holger Heyn and Dr. Catia Moutinho.

3.2.13 *Prdm9* amplification

In order to study the natural variability of the *Prdm9* alleles in the Madeira and Barcelona Rb system, exon 12 of the *Prdm9* gene, from repeat #2 to the C-terminal domain, was amplified and sequenced by Sanger sequencing. Exon 12 contains the ZnF domain array that recognizes and methylates specific DNA sequences (Parvanov et al. 2010).

3.2.13.1 Genomic DNA extraction

Genomic DNA was obtained from mouse tails (**section 3.1.2**). Briefly, a tail fragment (0.5-1 cm) was incubated in Lysis Buffer III with 100 µg/ml proteinase K O/N at 56ºC in agitation. Then, the lysate was centrifuged for 20 minutes at 800 x*g* and the supernatant recovered and mixed with 500 µl of isopropanol. The lysate-isopropanol mix was thoroughly mixed until DNA started to precipitate, then the mix was left 30 minutes at -80ºC to facilitate DNA precipitation. Samples were then centrifuged for 20 minutes at 800 x*g* and 4ºC. The supernatant was decanted and 200 µl of cold

ethanol 70% were added to the DNA pellet. Samples were centrifuged again for 20 minutes at 800 x*g* at 4ºC and then, the supernatant decanted. DNA pellets were left to air-dry at 37ºC for 5-10 min. DNA pellets were resuspended in MilliQ water O/N at 4ºC in agitation or at 60ºC for 10 min. Once DNA solutions were obtained, they were quantified in Nanodrop and stored at -20ºC until use.

3.2.13.2 Whole-genome PCR and DNA retrieval

Samples were genotyped as heterozygous or homozygous with an initial PCR, which also allowed determining allele size. In the case of heterozygous samples, around 20 amplification reactions of the same sample were performed in order to obtain enough DNA for Sanger-sequencing. For homozygous samples, two or three reactions were sufficient.

The PCR master mix was prepared calculating to use 100 ng of initial genomic DNA per PCR reaction: 4 µl of 10x PCR buffer (Takara) (final concentration x1), 3.2 µl of dNTPs (2.5 mM each, final concentration 0.2 mM), 1 µl of forward and reverse primers (**Table 3.2**) at 10 µm (final concentration 0.25 µM), 1.6 µl of DMSO (4% final concentration) and 0.25 µl of Taq (5U/µl, final concentration 0.03 U/µl). The PCR Program used was the following: 95ºC 3 min, 30 cycles of 95ºC 30 sec, and 64ºC for 30 sec followed by 72ºC 90 sec and ending by 10 minutes at 72ºC.

Table 3.2: Primer sequences for *Prdm9* amplification.

Primers	Source
PRDM9F-GAAAGTAAGAGAACTGTGGAAGAGCTCAGAA	Life Sciences
PRDM9R-GAGATGTGGTTTTATTGCTGTTGGCTTTCTC	Life Sciences

A 1% agarose gel was run for 2 hours at 75V and analyzed in the GelDoc XR. In case of heterozygous mice amplification bands were purified following manufacturer's instructions from Nucleospin and PCR Clean-up kit (Cultek). During purification, same-allele bands were pooled together. Homozygous samples PCR products and purified bands from heterozygous samples were then precipitated with 1/10 volume of 3M sodium acetate, pH 5.2 and 2-3 volumes of 100% ethanol and frozen O/N at -20 ºC. Following, samples were spun at 16,000 x*g* at 4 ºC for 30 minutes and the DNA was rinsed with 70% ethanol. The pellet was air-dried for 15 minutes and 25µl of TE buffer were added. Samples were left resuspending either O/N at 4ºC or 10 minutes at 60ºC.

Samples were sent to Macrogen Spain (https://dna.macrogen.com/esp/) for Sanger-sequencing. Chromatograms were analyzed with the Bioedit sequence alignment editor (version 7.2.5).

3.2.13.3 Single-molecule PCR and DNA retrieval

Single-molecule PCR is a highly efficient technique for the amplification of tandem repeat sequences such as the *Prdm9* ZnF array. Based in sperm typing studies (Striedner et al. 2017; Tiemann-Boege et al. 2006), the main premise is to extremely dilute the DNA sample in order to amplify just one copy of the *Prdm9* allele, avoiding unspecific artifactual amplifications.

Molecule Number Calculation.   This calculation is based on the assumption that the average weight of a base pair (bp) is 650 Da. This means that one mole of a bp weights 650 g and that the molecular weight of any double stranded DNA template can be estimated by taking the product of its length (in bp) and 650. The inverse of the molecular weight is the number of moles of template present in one gram of material. Using Avogadro's number, 6.022×10^{23} molecules/mole, the number of molecules of the template per gram can be calculated:

$$molecules/g = mol/g*molecules/mol$$

The number of molecules or number of copies of template in a sample can be estimated by multiplying by $1*10^3$ to convert to ng and then multiplying by the amount of template (in ng). In this manner, it is possible to calculate the number of molecules in 10 ng of mouse DNA, taking into account that the mouse genome size is 2.8×10^9 bp.

$$N \text{ molecules} = \frac{conc(ng) * N \text{ Avogadro}}{genome \text{ size (bp)} * 1 * 10^9 * average \text{ weight of a bp}}$$

Serial sample DNA dilutions.   To test the amplification of single molecules, serial dilutions of 3000-300-30-3 molecules of the DNA sample were performed.

In order for the DNA polymerase to work, some background DNA that won't be amplified with the primers in use is needed. This background DNA is a bulk of DNA where the polymerase can bind but the primers cannot.

For *E. Coli*:

- ▶ 1 ng/µl → 1 ng = 200,000 copies
- ▶ 0.1 ng/µl → 0.1 ng = 20,000 copies
- ▶ 0.01 ng/µl → 0.01 ng = 2,000 copies

PCR preparation and amplification.   A 5 µM primer solution was previously prepared, containing both forward and reverse primers (**table 3.2**). Then a PCR master mix was prepared in order to perform at least 20 single-molecule PCR reactions per sample, always including enough mix for a negative control and taking into account pipetting errors: 14.05 µl H_2O, 4 µl of 5x GC Buffer (1x final concentration), 0.4 µl dNTPs 10 mM (200 µM final concentration), 0.2 µl SybrGreen 10x (final concentration 1x), 0.25 µl primer mix 5 µM (0.0625 µM final concentration), 0.1 µl Polymerase (2U/µl, final concentration 0.1 U/10 µl) and 1 µl of DNA. DNA samples were quantified, and the number of molecules determined in order to adjust the amount of DNA to each PCR reaction. Samples were then amplified by RT-PCR: 94ºC for 2 min, 70 cycles of: 94ºC for 15 sec, 62ºC for 5 sec, 72ºC for 30 sec and 72ºC for 7 min; followed by 65ºC for 5 sec and 95ºC until the PCR finished. **Figure 3.9** shows an example of *Prdm9* amplification by RT-PCR and **Figure 3.10** an example of RT-PCR from seriated dilutions.
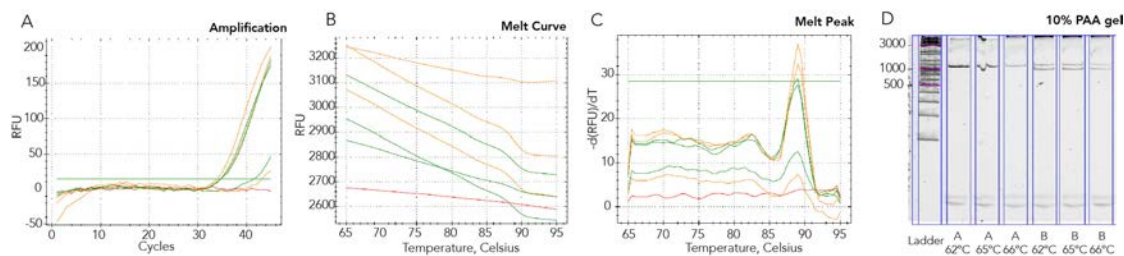
Figure 3.9: Example of *Prdm9* amplification by RT-PCR with temperature gradient using DNA from C57BL6/J mice (sample A) and from wild Rb mouse from the Rb Barcelona population (Sample B). (A) Amplification curves representing the relative fluorescent units (RFU) per amplification cycle. **(B)** Fusion peaks representing the RFU per temperature (ºC). **(C)** Melt peaks representing primer specifity, determined by plotting the negative first derivative of fluorescence versus the temperature (ºC). **(D)** 10% polyacrylamide gel showing *Prdm9* amplification (in bp), were the number of zinc fingers can be estimated by the PCR product size. The A sample is homozygous for a *Prdm9* allele with 12 zinc fingers and the B sample is heterozygous for a 10-zinc finger and 13-zinc finger alleles. PAA-polyacrilamide.
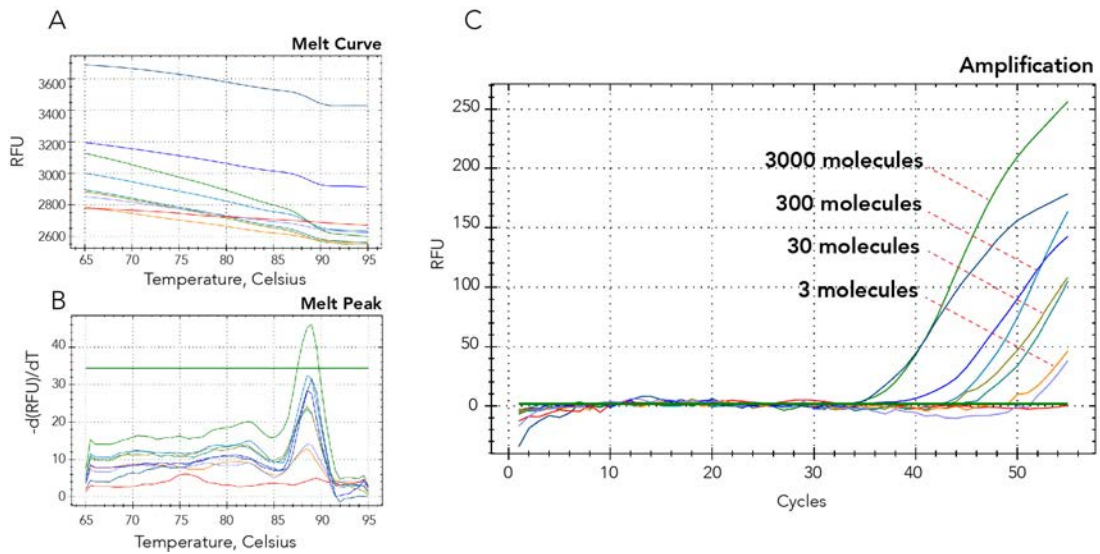
Figure 3.10: Example of RT-PCR from seriated DNA dilutions (A) and **(B)** show the curves and the fusion peaks respectively, as described in **Figure 3.9**. **(C)** Consecutive amplification curves depending on the number of starting molecules (3000-300-30-3 molecules) of the PCR reaction.

Band diagnosis and DNA purification. Once the RT-PCR was completed, an aliquot of each reaction was run in a 1% agarose gel to determine the size (bp) of the allele amplified per sample (**Figure 3.11**). Same-allele PCR products from the same sample were pooled together and purified with 1x AMPure beads. After purification, the beads were eluted with 32 µl of MilliQ water, gently mixed by pipetting, and incubated at RT for 5 min.
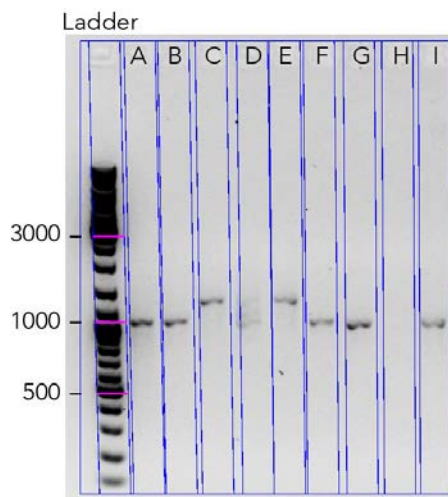
Figure 3.11: Example of amplification products from single molecule PCRs. 1% agarose gel showing different amplification product sizes (in bp) from single molecule PCR, corresponding to different alleles from heterozygous samples. A, B, F, G and I show 10-zinc finger alleles amplification while C and E show 12-zinc finger amplification.

Samples were separated on a magnet from the beads and transferred to a fresh tube and quantified in Qubit and run in a 10% PAA gel to check their quality. Once samples pass all quality checked, they were sent to Sanger-sequencing at Eurofins Scientific (https://www.eurofins.com/).

3.2.13.4 Sequence analysis

Sequences of *Prdm9* ZnF repeats were determined from bidirectional Sanger sequencing with the same primers used for amplification (see **Table 3.2**). Chromatograms were analyzed with BioEdit (version 7.2.5) (Hall 1999) to ensure reliable sequences, which were then translated using ExPasy: SIB bioinformatics resource portal (Artimo et al. 2012).

Different alleles were classified by assessing the number of ZnF repeats (from #8 to #16) and the variation in their amino acidic sequence in the highly variable positions -1, +3 and +6 of each ZnF repeat, following the nomenclature previously described (Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014). Each repeat was classified with a number (from #3 to #34), based on its amino acidic sequence and the different combination of repeats were classified as different alleles.

3.3 Data analysis

3.3.1 SNPs array

Mice from the BRbS were genotyped using the Mouse Universal Genotyping Array (MegaMUGA) in collaboration with Prof. J.B. Searle (Cornell University, USA). The MegaMUGA consisted in 77,808 evenly distributed SNP markers built on the Illumina Infinium platform (Morgan et al. 2016a). SNPs were filtered using PLINK version 1.9 (Purcell et al. 2007) with missing values above 5% threshold, yielding a final dataset of 63,344 informative SNPs evenly distributed throughout the genome with the exception of chromosomes 8 and Y. Then, the informative SNPs were used to estimate individual ancestry and admixture proportions assuming K populations based on a maximum likelihood method using the ADMIXTURE software (Alexander et al. 2009). Clusters from K=1 to K=10 were evaluated, and after applying Evanno's Δ K (Evanno et al. 2005), K=2, K=3, K=4 and K=5 resulted in the lowest likelihood values. Admixture analyses were plotted using Rstudio (R Core Team 2018) and Principal Component Analyses (PCA) were also included using a module from PLINK 1.9 (Purcell et al. 2007).

3.3.2 Genome-wide screening of genetic divergence and diversity

Genome-wide screening of genetic divergence was performed by computing Weir and Cockerham's F_{ST} mean values using VCFtools 0.1.15 (Danecek et al. 2011). First, genomic divergence between populations considering chromosomes involved and not involved in fusions. Estimated F_{ST} values were adjusted to minimize type I errors (Bonferroni correction). The number of alleles (Na) and allelic richness (Ar) were estimated using the R package hierfstat v.0.04.22 (Goudet 2005); the observed heterozygosity (Ho) and expected heterozygosity (Hs) were calculated with PLINK 1.9 (Purcell et al. 2007) and the inbreeding coefficient (Fis) and nucleotide diversity (pi) with VCFtools 0.1.15 (Danecek et al. 2011).

3.3.3 Estimates of recombination rates

The program LDhelmet (Chan et al. 2012) was applied for the estimation of variable recombination rates between pairs of SNPs at population scale (p = 4Ner). For recombination rate analysis, SNP data was phased with the SHAPEIT software (Delaneau et al. 2012) and likelihood tables generated with LDpop (Alexander and Machiela 2020). Recombination rates were established in windows of 50 SNPs across the mouse genome considering St and Rb mice as two groups.

Subsequent statistical tests to the divergence and recombination rate analyses were performed with JMP package version 5.1.2; SAS Institute Inc (SAS Institute Inc 2019) and with Rstudio (R Core Team 2018).

3.3.4 Hi-C data

The bioinformatic analysis of Hi-C data for the study of the structural organization of the mouse genome during spermatogenesis was performed by Dr. A. Paytuví-Gallart and L. Álvarez González, members of our research group. Briefly, in Hi-C processed samples, quality check and trimming were performed using BBDuk (version 10/2015) (Bushnell 2014). Then, reads were processed with TADbit (version 0.2.0.23) (Serra et al. 2017), mapping reads from 15 bp onward, using a step size of 5 bp. To remove possible artifacts, the following filters were applied: "self-circle", "dangling-end", "error", "extra dangling-end", "too short", "too large", "duplicated" and "random breaks". An inhouse script (Paytuví-Gallart 2019) was used for binning data into a raw matrix, which was normalized using HiCExplorer (version 1.8.1) (Ramírez et al. 2018).

Once the normalized matrices were obtained, averaged contact probability decay by genomic distance [P(s)] was established using HiCRep (version 1.4) (Yang et al. 2017). Then, the averaged contact probability [P(s)], the inter-/intra-chromosome interaction ratio and the inter-subtelomeric interaction quantification analyses were conducted with HiCExplorer. Following, compartments and TADs were determined with TADbit. A/B compartments were established by obtaining and analyzing the corresponding eigenvector values and TAD insulation scores were established with HiCExplorer. Differential matrices between different samples were also obtained and together with analyses of compartment switching, they were used to detect main structural differences between samples.

3.3.5 ChIP-seq data

Quality check and trimming of ChIP-seq reads were performed using BBDuk (Bushnell 2014). Reads were mapped to the mouse reference genome (mm10) using the Galaxy server (Afgan et al. 2018) and filtered using SAMtools (Li et al. 2009). Following, bigWig files were generated with DeepTools (Ramírez et al. 2016), to assess the genome coverage of the ChIP data. Peak calling was performed using MACS2 (Feng et al. 2012).

To assess the stability of peaks and their reliability, different peak callings were performed, applying q-values of 0.05, 0.01 and 0.001 together with fold change filters of 3, 4 and 4. Peakds were visualized with IGVtools (Thorvaldsdóttir et al. 2013) together with coverage and reads, allowing for the discrimination between clearly aligned peaks and noise. After assessment of peaks according to coverage and reads, it was determined that reliable peaks were those from q-value=0.01 and fold change=3.

BED files for reliable peaks were generated and annotated to the mouse reference genome (mm10) using the R/Bioconductor package ChIPseeker (Yu et al. 2015), in order to establish their relative position to genomic features such as promoters, enhancers or TSS. Peak overlap analyses an additional ChiP-seq data representation were conducted with the R/bioconductor package ChIPpeakAnno (Zhu et al. 2010).

3.3.6 RNA-seq data

RNA-seq data analysis was performed by Dr. Andreu Paytuví using the AIR platform (https://transcriptomics.sequentiabiotech.com/).

3.3.7 Figure representation with R

For data representation, different packages were used: ggplot2 (Wickham 2016), karyoploteR (Gel and Serra 2017), ChIPseeker (Yu et al. 2015), ChIPpeakAnno (Zhu et al. 2010), Gviz (Hahne and Ivanek 2016) and rtracklayer (Lawrence et al. 2009).

RESULTS

Results | 4

4.1 **Development of a reproducible flow cytometry method to isolate enriched germ cell populations from mouse testis**

4.1.1 Introduction

Mammalian spermatogenesis is an intricate process resulting in the formation of male haploid gametes (**section 1.2.1**). In order to attain a better understanding of the complexity of the process, the development of a reproducible cytometry method to isolate germ cells is much in need. The isolation of different germ cell populations from adult mice testis poses the difficulty of obtain specific cell fractions that are simultaneously present in the testis. In this context, the canonical work developed by Romrell and colleagues (Romrell et al. 1976) set the grounds for the development testis dissociation techniques. Their recommendations for the obtention of an optimal somatic-free germ cell suspension from which populations could be isolated have been extensively followed in the field (Fallahi et al. 2010; Liu et al. 2015; Rodríguez-Casuriaga et al. 2014), becoming the basis of studies based on germ cell isolation (Bryant et al. 2013; Fallahi et al. 2010; Gaysinskaya et al. 2014). In fact, different techniques have been developed since then to purify, to varying degrees, various meiotic fractions from both adult and immature animals. These included three different methodological approaches: (i) sedimentation (Barchi et al. 2009; Bryant et al. 2013), (ii) magnetic-activated cell sorting (MACS) (Gassei et al. 2009) and (iii) fluorescence-activated cell sorting (FACS) (Bastos et al. 2005; Fallahi et al. 2010; Gaysinskaya et al. 2014; Getun et al. 2011; Lima et al. 2016; Rodríguez-Casuriaga et al. 2014; Struijk et al. 2019).

Separation by sedimentation discriminates cells based on their size and density and can be performed either by sedimentation by unit gravity (STA-PUT) or by elutriation. STA-PUT separates germ cells by velocity sedimentation at unity gravity (Bryant et al. 2013; Dunleavy et al. 2019; Lam et al. 1970). This approach has been applied to different species such as mouse (Bellvé et al. 1977; Lam and Bruce 1971), rat (Go et al. 1971; Platz et al. 1975), trout (Louie and Dixon 1972), ram (Loir and Lanneau 1974) and cricket (Kaye and McMaster-Kaye 1974). Although STA-PUT normally yields high number of cells when compared with other techniques ($1 \cdot 10^8$ cells/spermatogenic cell type), it is inefficient in recovering different cell types (Meistrich et al. 1973; Romrell et al. 1976), being specially relevant the absence of a pure spermatogonia fraction (Bryant et al. 2013; Lam et al. 1970; Meistrich 1972; Romrell et al. 1976). On the other hand, elutriation separates cells also by velocity sedimentation, but it requires less time for cell recovery and yields a higher number of cells ($3\text{-}5 \cdot 10^8$ cells/spermatogenic cell type) (Barchi et al. 2009). Germ cells from rodents such as mice (Bryant et al. 2013; Grabske et al. 1975), rat (Sanborn et al. 1975) and hamster (Grabske et al. 1975) have been obtained using this method. However, the main limitation of the elutriation approach is the need for specific equipment to apply this technique (the elutriator) and, as in STA-PUT, it fails to obtain homogeneous and enriched populations of the different spermatogenesis cell types, as different populations present similar sedimentation velocities (Barchi et al. 2009).

In the case of MACS, cells are obtained based on their cell surface molecules (Miltenyi et al. 1990). Using this technique, Gassei and colleagues successfully isolated spermatogonia from rat (Gassei et al. 2009), human and monkey (Gassei et al. 2010). The use of spermatogonial markers such as ID4 or BMI [(de Rooij 2017) and references therein] allowed for the development of more refined studies on spermatogonial differentiation (Aloisio et al. 2014; Helsel et al. 2017; Komai et al. 2014). Nevertheless, the lack of knowledge of specific markers for the remaining germline populations limits the applicability of a MACS-only based methodology for cell discrimination.

As for FACS, it combines the premise of discriminating germ cells based on their DNA content in addition to their morphological characteristics (as STA-PUT and elutriation) with the use of specific markers (as in MACS). FACS is based on the staining of germ cells with a vital dye in order to discriminate the different cell types by their ploidy (DNA content) and the efficiency of DNA staining, which varies depending on the degree of chromatin condensation (cell complexity). Most of the studies have used Hoechst 33342 (Ho) as the vital dye to sort germ cells (Bastos et al. 2005; Fallahi et al. 2010; Gaysinskaya et al. 2014; Getun et al. 2011), given its proven efficiency in DNA labelling and its characteristic fluorescence emission pattern (Arndt-Jovin and Jovin 1977; Belloc et al. 1994; Latt and Stetten 1976). While FACS sorting of germ cells results in low number of cells recovered when compared to sedimentation techniques ($0.5\text{-}2.0 \cdot 10^6$ cells/spermatogenic cell after

FACS), it permits the recovery of more populations than sedimentation, including spermatogonia, primary spermatocytes, round spermatids and sperm (Getun et al. 2011). In addition, the potential use of specific markers to refine cell discrimination increases the potential of FACS for isolation of germ cells. However, the lack of reproducibility in cell discrimination (Gaysinskaya and Bortvin 2015; Getun et al. 2011) highlighted the need for a reliable, efficient and reproducible method. This is especially relevant when conducting high-throughput genomic downstream analysis (i.e., ChIP-seq, RNA-seq), where cell purity is essential for obtaining reliable results.

Given this background, the main aim of this study was to develop a reproducible cytometry protocol to isolate germ cells from mouse testis. We paid special attention to the optimization of what are considered the most critical steps: (i) testis disaggregation, which included mechanically and enzymatically disaggregation of testis and cell staining; and (ii) germ cell discrimination by FACS.

4.1.2 Results

4.1.2.1 Optimization of testis disaggregation

Testis dissociation.   C57BL/6J adult male mice (10-18 weeks) were used for the optimization of testis dissociation using different conditions (**Table 4.1**). The main goal of this critical step was to dissociate cells from the seminiferous tubules in an optimal state of preservation. To that aim, both collagenase type II and type IV were initially tested. While collagenase type II contains higher levels of protease activity, particularly clostripain (Gilles et al. 1979), collagenase type IV has low tryptic activity, thus better preserving membrane proteins and receptors. In the end, two incubation steps with collagenase II (0.5 mg/ml) (fifteen minutes each incubation) resulted to be more effective, yielding higher cell numbers (20-30 M cells per testis) than with collagenase IV (10-20 M cells per testis) (**Table 4.1**).

Table 4.1: Summary of the chemical disaggregation optimization.

Disaggregation volume	Collagenase type (0.5 mg/ml)	Tested trypsin concentrations	Average nº cells recovered/animal
Both testes in 3 ml	IV	0.8 mg/ml	3.75 M
Both testes in 10 ml	IV	0.25 to 0.5 mg/ml	10-13 M
Each testis in 10 ml	IV	0.125 mg/ml	14 M
Each testis in 10 ml	II	0.25 to 0.375 mg/ml	20-30 M

Additionally, different concentrations of bovine trypsin were tested during the optimization process (from 0.125 mg/ml to 0.8 mg/ml) (**Table 4.1**). As high concentrations of trypsin (0.8 mg/ml) resulted in a high cell mortality, concentrations were subsequently reduced. Thus, two incubations with trypsin were finally included in combination with collagenase II (0.5 mg/ml): a first one at 0.375 mg/ml (15 minutes, 33ºC), and a second one, doubling the trypsin concentration in the same volume from the first incubation, to enhance tissue disaggregation (15 minutes more at the same temperature). Hence, the combinatorial action of collagenase and trypsin was the main factor that influenced the total number of cells recovered from testis.

Cell staining.   Once a homogeneous suspension of germ cell was successfully obtained, the next step included cell staining with a fluorescent dye detectable by the laser of the flow sorter. Cells were dyed with Hoechst 33342 (Ho) in combination with propidium iodide (PI), a vital dye (Ciancio et al. 1988; Riccardi and Nicoletti 2006).

Following previous studies (Gaysinskaya and Bortvin 2015; Getun et al. 2011), our first approach was to sort living cells. This included two staining steps: the first one using 0.13 mg/ml Ho together with trypsin (0.5 mg/ml) and 50 µg/ml PI, incubated for 15 min at 33ºC, and a second incubation

with only Ho (5µg per two million cells) for 30-45 min at 33º (**Table 4.2**). This approach dramatically reduced cell viability from 70-90% after disaggregation to 14% living cells after staining, resulting in non-reproducible FACS profiles (**Figure 4.1**).

Consequently, the staining process was reduced to just one step using 5 µg Ho per million cells with 50 µg/ml PI in a 45 min incubation at 33ºC in an attempt to decrease cell mortality during staining (**Table 4.2**). Although we found an apparent discrimination of germ cell populations, FACS profiles were not reproducible between experiments (**Figure 4.1 B-D**). Therefore, a pre-FACS cell fixation approach was tested (**Table 4.2; Figure 4.1**), as fixation by crosslink is a step required in many high-throughput genomic studies (e.g., in nuclei Hi-C or ChIP-seq). After including a cell fixation step (1% Formaldehyde, 10 min) before the cell staining, FACS profiles were more consistent between experiments (**Figure 4.2 D**).

Table 4.2: Optimization of cell preparation for FACS.

Test	Fixation before FACS	Hoechst dye
#1	No	(i) Incubation 0.13 mg/ml Hoechst, 0.5 mg/ml trypsin and 50 µg/ml PI; 15 min, 33ºC. (ii) Hoechst dying: 5µg/ml Hoechst per 2 million cells; 30-45 min, 33ºC.
#2	No	5 µg Hoechst per million cells with 50 µg/ml PI. 45 min incubation at 33ºC.
#3	1% Formaldehyde, 10 min	5 µg per million cells Hoechst 30 min incubation at 33ºC.
#4	1% Formaldehyde, 10 min	5 µg/ml Hoechst 30 min incubation at 33ºC.
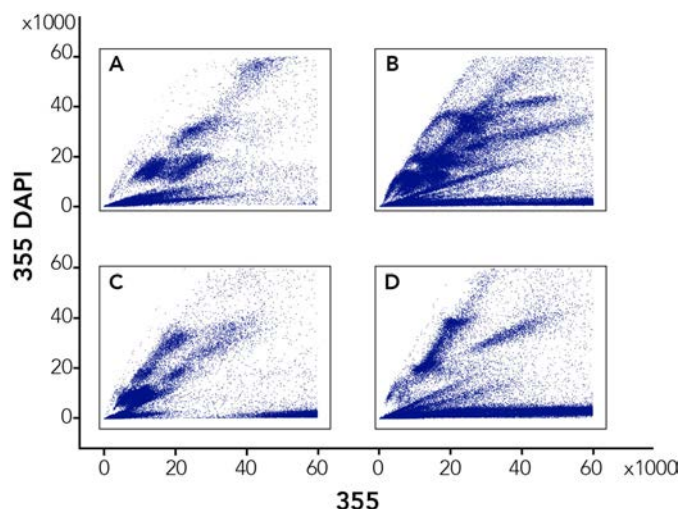
Figure 4.1: FACS profiles resulting from live cell staining tests. FACS profiles resulting from live cell staining tests. (A) FACS profile resulting from the #1 cell staining test. **(B-D)** FACS profiles resulting from #2 cell staining test. See table 4-2 for details. Ho red emission is represented in the X axis (355), discriminating by cell complexity while Ho blue emission is represented in the y axis (355-DAPI), discriminating by DNA content.

4.1.2.2 FACS setup

Once testis dissociation was optimal and provided reproducible results, FACS setup and subsequent purification of germ cells followed. Cells were sorted using a BD Influx™ cell sorter (BD Biosciences) at the CRG Flow Cytometry Unit, at Barcelona Biomedical Research Park (**section 3.2.2**).

Briefly, FACS setup consisted of establishing well-differentiated sorting gates to isolate distinct cell populations based on DNA content and cell complexity. Firstly, cells were discriminated from random particles by plotting their SSC vs. FSC profiles (**Figure 4.2 A**). A general population of cells (P1 fraction) was discriminated from cell debris. Then, the selected P1 fraction was plotted according to PE-TR-PI (fluorochrome laser for PI) and SSC. The resulting plot provided additional insights into cell morphology by ensuring the selection of whole cells with intact membrane and not cell

fragments (P2) (**Figure 4.2 B**). Subsequently, the P2 fraction was analyzed by 355 DAPI width (cell width) vs. 355-DAPI (Hoechst blue) (**Figure 4.2 D**), selecting single cells (P3 fraction) and excluding cell aggregates. Finally, the P3 fraction was analyzed by their Ho staining, distinguishing germ cells by their DNA content (355-DAPI) and cell complexity (355) (**section 3.2.2**). By this way, six main germ cells populations were detectable: P4, P5, P6, P7, P8 and P9 (**Figure 4.2 D**).
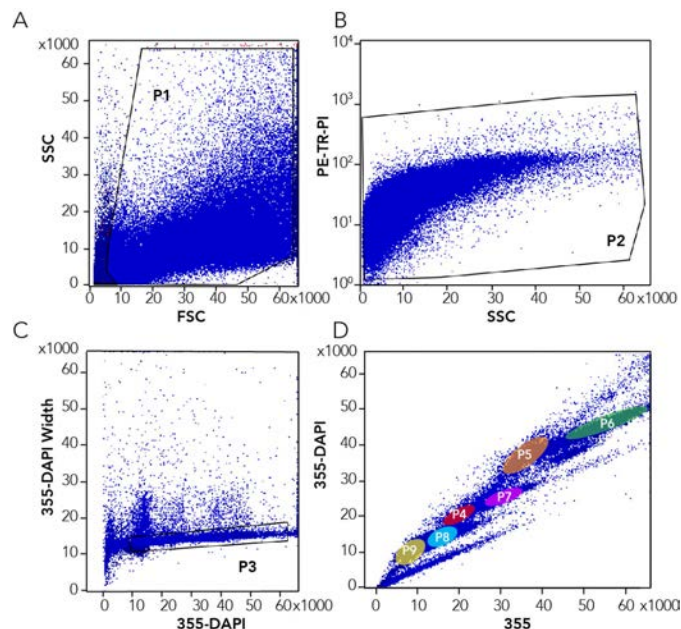
Figure 4.2: Representative spermatogenesis FACS profiles. (A) Representative scatter plot (FSC vs. SSC). Gate P1 selects cells, distinguishing them from debris. **(B)** PE-TR-PI (fluorochrome laser for PI) vs. SSC plots. Gate P2 contains cells with an intact membrane. **(C)** Analysis of cell aggregates. P3 gate corresponds to single cells, excluding cell aggregates. **(D)** Analysis of germ cells by their DNA content (355-DAPI) and cell complexity (355). Six different germ cell populations were distinguished: P4, P5, P6, P7, P8 and P9.

Enrichment analysis.   Once the FACS set up was established, all germ cell populations detectable (P4, P5, P6, P7, P8 and P9) were sorted (4,000-7,000 events per second). Cell enrichment analysis was conducted by immunofluorescence against specific proteins, such as γH2AX and SYCP3, analyzing DAPI morphology (**section 3.2.3**).

The analysis of DAPI morphology revealed that cells from the P4 fraction (**Figure 4.2 D**) contained spermatogonium with an average enrichment of 91%$\pm$5 (**Figure 4.3**). This was confirmed by immunofluorescence using an antibody against CD90.2, which labels Thy-1.2, a molecule present in the cell surface of adult germline stem cells (Hammoud et al. 2014) (**Figure 4.3 A**). Moreover, the position of the P4 fraction in the 355-DAPI/355 plot indicated that these were cells with 2c content. Since spermatogonium represent a heterogeneous population (de Rooij 2017; Hammoud et al. 2014; Helsel et al. 2017), we detected that intermediate spermatogonia were the most frequently represented sub-population (46.70%$\pm$11.79) followed by type B spermatogonia (26.58%$\pm$16.57) and type A spermatogonia (17.53%$\pm$10.73) (**Figure 4.3 B**). The remaining cells were mostly round spermatids (RS, around 4%), followed by elongated spermatids (ES, around 2%) and secondary spermatocytes (1%). Sperm and primary spermatocytes were observed is low proportions (0-1%). A mean number of 216,000 spermatogonia were recovered per adult mouse.
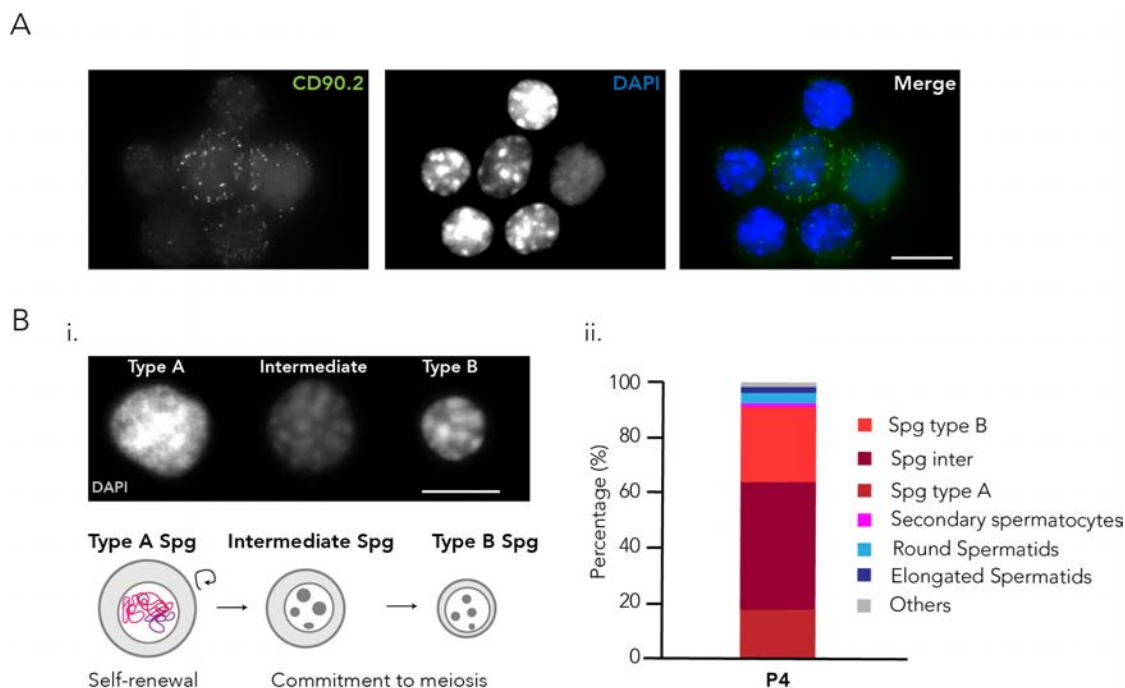
Figure 4.3: Enrichment analysis of P4 fraction. (A) Representative image of an immunofluorescence against CD90.2 (green). **(B)** Characterization of spermatogonium population. (i) Representative DAPI staining of type A, intermediate and type B spermatogonium (upper panel) and their schematic representation (lower panel). DAPI staining allows for cell type discrimination by their morphology. (ii) Percentage of each cell type found in the P4 fraction. "Others" include cells such as sperm or primary spermatocytes. Scale bars= 10μm.

In the case of both P5 and P6 fractions, their position in the 355-DAPI/355 plot corresponded to diploid cells (2n) with 4c content (**Figure 4.2 D**). In the case of the P5 fraction, the immunodetection of SYCP3 and γH2AX revealed a heterogeneous population, with 54.7%±10.7 being early prophase spermatocytes (leptonema/zygonema, L/Z) and 30.1%±6.01 late prophase spermatocytes (pachynema/diplonema, P/D) (**Figure 4.4**). The remaining cells were mainly spermatogonia (11%), followed by secondary spermatocytes (2.5%), RS (1.7%) and to a lesser extent, sperm (0.4%).
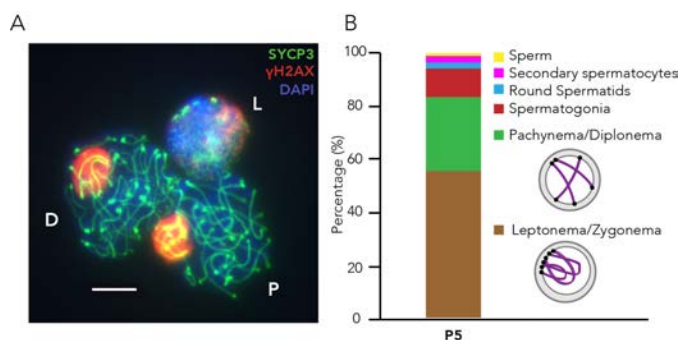
Figure 4.4: Enrichment analysis of P5 fraction. (A) Representative image of cells recovered from gate P5 showing the synaptonemal complex labeled with SYCP3 (green), asynapsis labeled with γH2AX (red) and chromatin labeled with DAPI (blue). Cells at leptonema (L), pachynema (P) and diplonema (D) are presented. Scale bar= 10μm. **(B)** Plot depicting the enrichment of the cell types found in P5 (left) with a schematic representation of each cell fraction (right).

Cells recovered from the P6 fraction (**Figure 4.2 D**) were primary spermatocytes at P/D with an average enrichment of 81.36%±9.40, followed by RS (15.74%±10.03). The 2.9% remaining cells of P6 were a mix of spermatogonium, metaphase I and sperm (**Figure 4.5**). The same immuno-analysis described for the P5 faction was performed, with 46.74%±13.85 of cells being at pachynema while 35.23%±12.70 at diplonema. A mean number of 650,000 P/D were recovered per adult mouse.
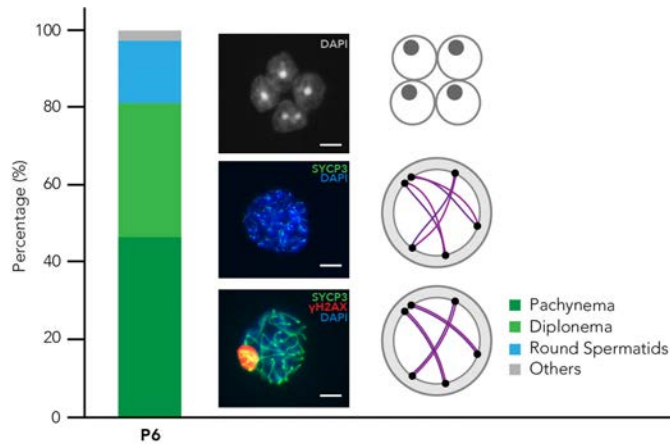
Figure 4.5: Enrichment analysis of P6 fraction. Plot depicting the percentage of cell types sorted from the P6 gate (left), with enrichment of primary spermatocytes at pachynema/diplonema (P/D) above 80% and a small (16%) fraction of round spermatids (RS). For each cell type, the corresponding immunofluorescence images is represented, accompanied by their schematic representation. The three cell fractions are DAPI stained (grey in RS and blue in P/D). Additionally, the synaptonemal complex (SYCP3) is labeled in green in P/D, and γH2AX (red) also labeled in pachynema. "Others" included sperm spermatogonia and metaphases II. Scale bars= 10 μm. Spg-Spermatogonia; Met I-metaphase I.

The P7 fraction corresponded to a heterogeneous population of 2c content (**Figure 4.2 D**). To further characterize this cell population, immunofluorescence with α-tubulin was performed to visualize the cytoskeleton (**Figure 4.6 A**). This analysis revealed that the cell fraction from the P7 sorting gate included secondary spermatocytes (13.9%$\pm$5.2), cells in anaphase/telophase transition (18.2%$\pm$13) and cells in diakinesis II (64.2%$\pm$15.2) (**Figure 4.6 B**). The remaining 3.7% of cells where a mix of RS, sperm, primary spermatocytes, spermatogonium and ES. Overall, a mean number of 430,000 cells from P7 was recovered per adult mouse.

Finally, cells from P8 and P9 fractions (**Figure 4.2 D**) corresponded to haploid cells with 'c' DNA content according to their position in the 355-DAPI/355 plot. These were round spermatids (P8 fraction) and sperm (P9 fraction). While RS represented 85% of cells from P8 fraction, sperm were more challenging to isolate from P9 fraction. As the sperm morphology resulted in a diffused FACS profile, the probability of placing the gate incorrectly and getting contamination from the neighbor population (RS) increases (**Figure 4.2 D**). Thus, to obtain an enriched fraction of sperm it was necessary to perform an alternative re-gating approach (see below).
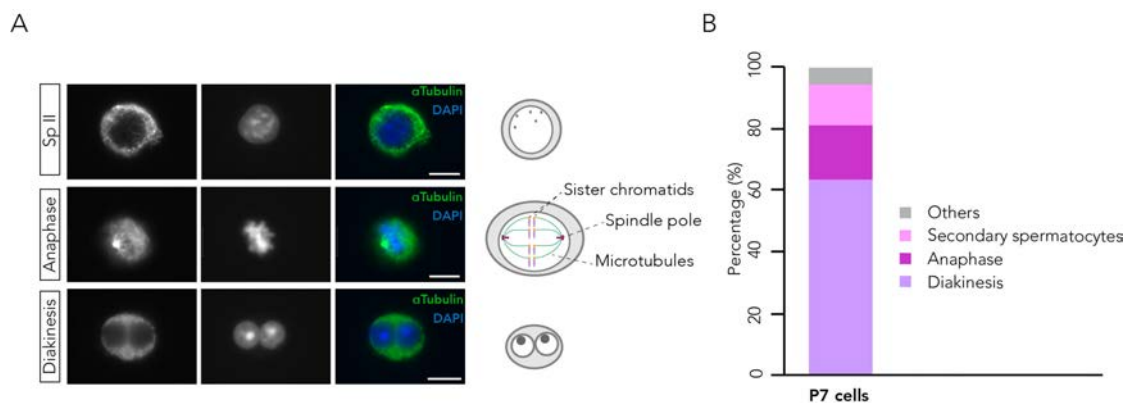
Figure 4.6: Enrichment analysis of P7 fraction. (A) Immunofluorescence with α-tubulin (green) and DAPI staining unveiled that the P7 fraction contained secondary spermatocytes (SpII), cells at anaphase and cells in diakinesis accompanied by its corresponding schematic representation (right). Scale bar= 10 μm. **(B)** Enrichment of the P7 fraction showing the percentage of each cell type. "Others" included single RS, sperm, SpI, Spg and ES. Sp I-primary spermatocytes; Spg-spermatogonia; RS-round spermatids; ES-elongated spermatids.

Purification of early primary spermatocytes. Due to the poor enrichment values obtained for primary spermatocytes from P5 fraction (54.7%$\pm$10.7 of L/Z and 30.1%$\pm$6.01 of P/D) we set up a specific isolation protocol based on immunolabeling of phase-specific proteins previous to FACS. To that aim, we immunolabeled simultaneously the proteins DMC1 and SYCP3. DMC1 is involved in the DSB repair pathway (Yoshida et al. 1998), thus being specifically expressed during L/Z, while absent in later prophase stages. By this way, cells discriminated positively for DMC1 and SYCP3 (labeling the synaptonemal complex) correspond to early prophase primary spermatocytes

at leptonema/zygonema, while cells only positive for SYCP3 and negative for DMC1 correspond to late prophase primary spermatocytes at P/D.

To ensure protein labeling, antibody titration was an essential preceding step to determine the lowest antibody dilution that resulted in a positive fluorescence signal. For that, 'in-solution' immunofluorescence was performed in a suspension of germ cells obtained from testis disaggregation (**section 3.2.7.2**), using an anti-mouse SYCP3 primary antibody (in combination with an anti-mouse Cy5 secondary antibody) and an anti-rabbit DMC1 primary antibody (in combination with an anti-rabbit FITC secondary antibody) (see **appendix table A.1.1**). For the titration experiment, serial dilutions (from $1:10^6$ to 1:100) of primary antibodies were performed, maintaining a constant number of cells (around 3 million cells) per each dilution tested. This was combined with serial dilutions for secondary antibodies. Control dilutions included:

a) Negative control: sample without antibodies, unstained.
b) Primary antibody negative control: sample with just secondary antibodies (anti-mouse Cy5 and anti-rabbit FITC)
c) Secondary antibody negative control: sample with just primary antibodies (anti-mouse SYCP3 and anti-rabbit DMC1).

Samples were then analyzed with the flow cytometer, comparing controls (**Figure 4.7 A**) with the serial dilutions performed for each antibody combination (**Figure 4.7 B-K**). For all samples, cells were firstly discriminated by Ho staining. The whole 4c population (including P5 and P6 gates from **Figure 4.2 D**) was gated and examined with the lasers 640-APC (640 nm emission from the Cy5 fluorochrome) and FITC (530 nm emission from the FITC fluorochrome) (**Figure 4.7**). The staining shift from negative to positive was evident in the combination anti-SYCP3 at $1:10^4$ with anti-mouse Cy5 at $1:10^6$ (**Figure 4.7 H**) and the optimal dilution determined was both SYCP3 and Cy5 at 1:103 (**Figure 4.7 I**). As for the primary anti-DMC1, a similar titration process was performed, expecting two populations in this case: DMC1 positive for L/Z and DMC1 negative for P/D. So, the optimal dilution was 1:100 for DMC1 combined with FITC diluted $1:10^3$ (**Figure 4.7 J**).

Once the optimal concentration was determined for each primary and secondary antibody, double immunofluorescence with both SYCP3-Cy5 and DMC1-FITC was performed at the respective optimal concentrations. The FACS profile resulted in the clear distinction of two main populations: P10 fraction (cells negative for DMC1 and positive for SYCP3) and P11 fraction (cells positive for both DMC1 and SYCP3) (**Figure 4.7 K**).
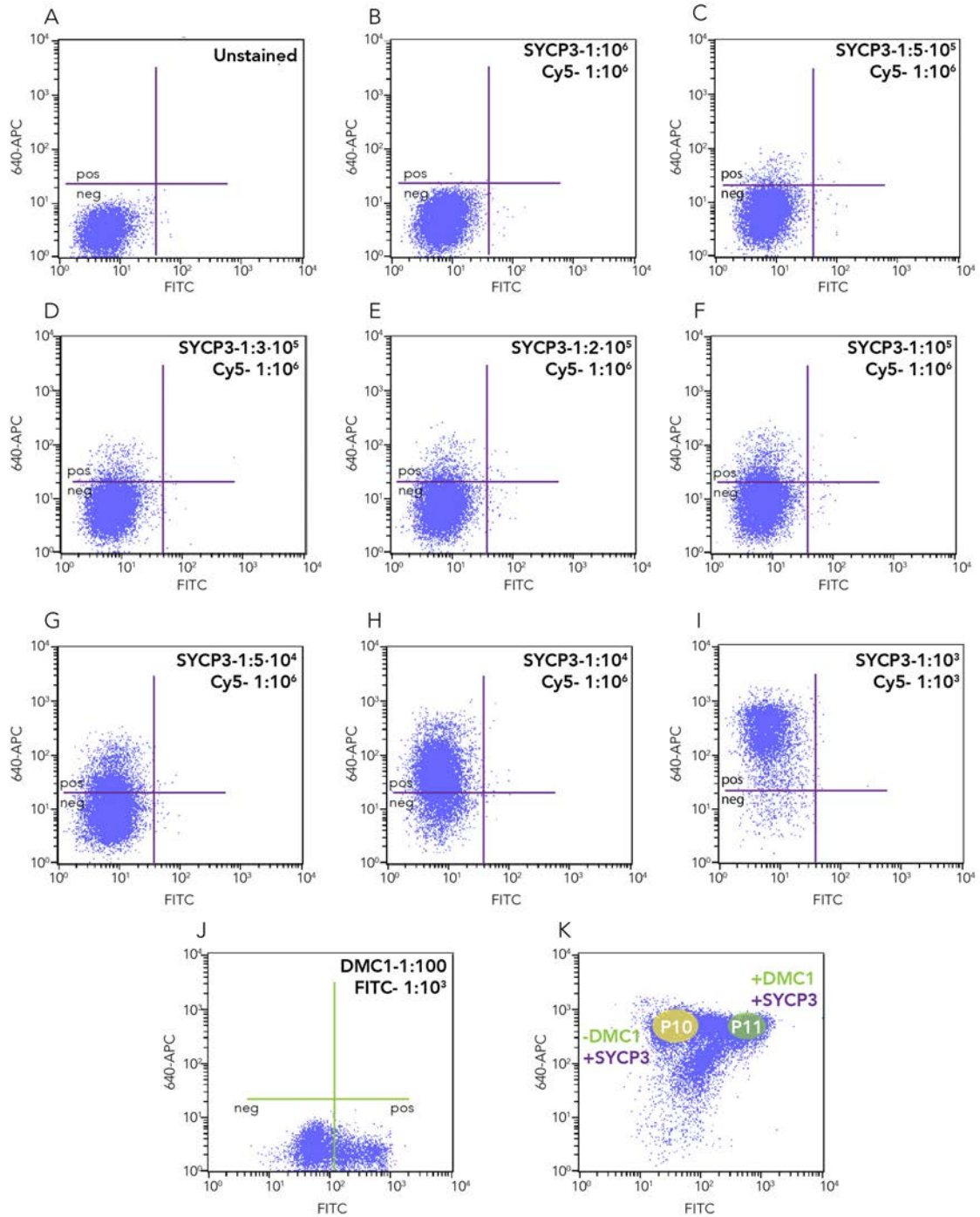
Figure 4.7: FACS profiles for antibody titration. (A) FACS profile of the negative control without antibodies (unstained). **(B-H)** FACS profiles of increasing anti-SYCP3 concentrations and stable anti-mouse Cy5 concentration (1:10[6]). The shift to positive staining is detectable from panel C, being clear from panel G. **(I)** FACS profile from the optimal concentration found for anti-SYCP3 and anti-mouse Cy5 (1:10[3] for both antibodies), where the shift to positive signal is clear and at high resolution. **(J)** FACS profile of anti-DMC1 (1:100) in combination with anti-rabbit FITC (1:10[3]), showing two evident cell populations: negative cells for DMC1 on the left and positive cells for DMC1 on the right. **(K)** FACS profile from 4c germ cells, resulting from the double immunostaining with DMC1/FITC and SYCP3/Cy5, distinguishing two main populations: P10 (negative for DMC1 and positive for SYCP3) and P11 (positive for both DMC1 and SYCP3).

Subsequently, P10 and P11 fractions were subjected to an enrichment analysis. The immunodetection of γH2AX in addition to DMC1 and SYCP3 permitted to accurately stage prophase I in the isolated fractions, confirming that the P10 fraction corresponded to primary spermatocytes at P/D (98.22%±2.05) and P11 to primary spermatocytes at L/Z (88.68%±4.04) (**Figure 4.8**). A mean number of 240,000 L/Z and 665,000 P/D cells per recovered per adult mouse.
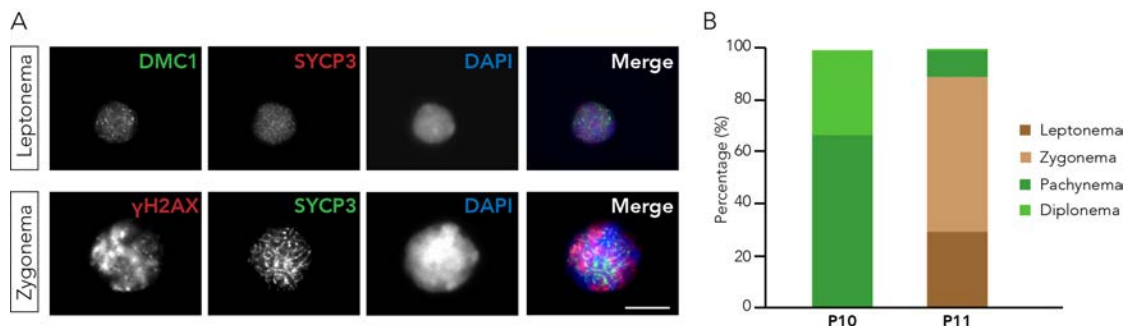
Figure 4.8: Enrichment analysis of P10 and P11 fractions. (A) Representative immunofluorescence of the predominant cell types found. For leptonema, a positive signal for DMC1 is shown in green, together with the SYCP3 signal in red that allows for spermatocyte staging. For zygonema, the γH2AX signal spread in the nucleus together with a higher level of synapsis detectable by SYCP3 confirms the stage of this cell. Scale bar= 10 μm. **(B)** Enrichment analysis of P10 and P11 fractions.

Purification of round spermatids and sperm   The fact that both round spermatids and sperm are haploid cells hampered their discrimination by FACS based solely on DNA content (**Figure 4.9 A**). Moreover, the specific morphology of the mouse sperm head resulted in an unclear and diffused population in the Ho emission plot [(**Figure 4.9 A**), gate P9]. However, the specific morphological features of each cell type served to apply an alternative re-gating approach (**Figure 4.9 D**).

Briefly, the whole c population [including P8 and P9 gates (**Figure 4.9 A**)] was gated and plotted by SSC (cell complexity) vs. FSC (volume), allowing for refined isolation of each population (**Figure 4.9 B**). This allowed to establish new fractions according to their morphological differences (the P12 fraction for RS and the P13 fraction for sperm). In the case of the P12 fraction, the premise was to obtain the most homogeneous group of cells from P8, according to their light diffraction patterns (SSc vs. FSC). As for the P13 fraction, additionally to the morphological parameters, we selected according to size (i.e., the smallest cells from P9 were the ones selected). Sperm was also recovered from the caudal epididymis, which was disaggregated and FACS sorted (P14) (**Figure 4.9 C**), discarding somatic cells that could contaminate the cell fraction.

Overall, the average enrichment obtained for RS was 91.66%±3.9 for the P9 fraction, and 93.23%±4.75 from the corresponding P12 fraction. As for sperm, the average enrichment was 87.54%±6.71 from the re-gating approach (P13 fraction), and 98.68%±5 from the fraction recovered from epididymis (P14). A mean number of 3.75 million RS, 1.6 million testicular sperm and 2 million epidydimal sperm were recovered per adult mouse.
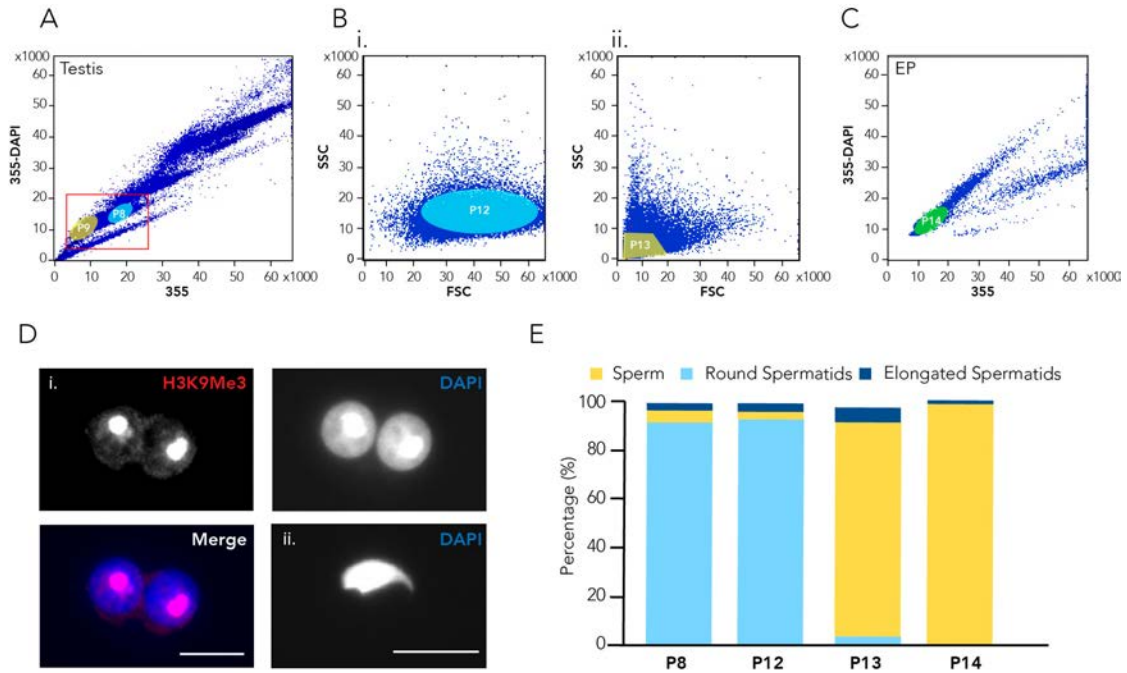
Figure 4.9: Re-gating approach to isolate round spermatids and sperm. (A) FACS profile of fixed germ cells. The red square frames cells (n,c) and primary gates are established (P8 and P9). **(B)** Refined morphology analysis of (n,c) populations by FSC vs. SSC profiles. (i) Cells from gate P8 are plotted and a gate established (P10) aiming to isolate highly enriched round spermatids (RS). (ii). Cells from gate P9 are plotted, and a gate established (P11) aiming to isolate highly enriched sperm. **(C)** FACS profile of sperm recovered from the epididymis (P14). **(D)** Microscopy images of RS (i) and sperm (ii), recovered from FACS. Scale bar=10 μm. **(E)** Enrichment analysis of re-gated populations P12 (RS) and P13 (Sperm), P8 fraction from the 355-DAPI vs. 355 DAPI plot (panel A, RS) and cells recovered from the epididymis (P14, Sperm). ES-elongated spermatids.

4.1.3 Discussion

Previous attempts to isolate germ cells using FACS faced numerous issues regarding cell viability, reproducibility and efficiency (Bastos et al. 2005; Gaysinskaya and Bortvin 2015; Getun et al. 2011; Mays-Hoopes et al. 1995). Our approach represents an improvement in cell disassociation, attaining an optimal germ cell suspension by the combinatorial action of collagenase type II and trypsin, which yielded high cell viability. In addition, we ensured high reproducibility in flow sorting by including cell fixation previous to FACS. The obtention of enriched fractions of spermatogonia, primary spermatocytes at L/Z and P/D, RS and sperm resulted from a thorough optimization of the FACS setup, including substantial improvements in the disaggregation protocol based on the analysis of FACS profiles and subsequent sorted fractions. So, we recovered high cell numbers per fraction with a high reproducibility when compared to peviously described FACS-based methods (Gaysinskaya and Bortvin 2015; Getun et al. 2011). Also, our gating protocol proved to be more efficient than similar FACS-based discriminating approaches (Lima et al. 2016). This has allowed us to apply high-throughput techniques to the enriched fractions, including Hi-C, ChIP-seq and Hi-C, that permitted the refined study on genome organization and its link with cohesin occupancy and transcription presented in this thesis (**sections 4.2 and 4.3**).

During the process of cell recovery, we faced some difficulties as we recovered low numbers of spermatogonia compared to the high numbers attained for RS (an average of 3.75 million RS/animal). As per the spermatogenic process, this RS recovery should represent four times the spermatogonia recovery but in our approach the recovered spermatogonia represented 1/16 of the RS fraction. The low spermatogonia recovery could be explained by the presence of the blood-testis barrier (BTB) (Cheng and Mruk 2002; Phillips et al. 2010). As spermatogonia are found in the basal compartment, mostly adjacent to the vasculature, the BTB would prevent an efficient enzymatic release of the spermatogonia from the seminiferous epithelium. A possible improvement for spermatogonial

recovery would be to perfuse the animal preceding the testes obtention, as classically described (Romrell et al. 1976), in this manner the BTB would loosen by the loss of blood cells, facilitating spermatogonial release during testis disassociation.

Regarding fraction purity, we detected groups of 4 RS in the 4c fraction. These groups of 4-RS could be the result of incomplete collagenase-based disaggregation which has been described to promote the formation of symplasts (multinucleated germ cells) by the widening of intercellular bridges between synchronously differentiating germ cell clusters (Bellvé 1993; Meistrich 1972; Romrell et al. 1976). Even though these symplasts were successfully reduced from the 4c fraction with trypsin concentration adjustments (**Table 4.1**), what appeared to be 2-RS symplasts were still present in the 2c fraction (**Figure 4.6**). Despite the possibility of these 2-RS cells being symplasts (thus artifactual), the reduction of symplasts in the 4c fraction after disaggregation optimization together with the fact that this 2-RS fraction was recovered jointly with immediate phases in the spermatogenic process (secondary spermatocytes and anaphase) (**Figure 4.6**), favored the classification of this fraction as cells in diakinesis.

The method presented here can have further applicability, given the description of new spermatogonial markers [(de Rooij 2017) and references therein] together with the identification of cell-specific novel markers through single-cell RNA-seq (Lukassen et al. 2018; Sohni et al. 2019). Recent studies apply similar FACS-based discriminating approaches to isolate enriched cell fractions from spermiogenesis (Struijk et al. 2019). Thus, our method represents a reliable base from which new experimental strategies can be developed for the isolation of specific germ cell populations.

Overall, we developed a reproducible cytometry method to isolate highly enriched populations of germ cells from mouse testis, which included spermatogonia, primary spermatocytes (L/Z and P/D), RS and sperm. We foresee that this protocol can be applied not only to mammal species, as shown by contemporary studies (Lima et al. 2016), but also to evolutionary distinct taxa. Thus, we anticipate that our methodological approach will provide impetus for further exploration of the functional and structural basis of genomes in a broad context, which will reinforce the link between developmental genetics and genome evolution.

4.2 Study of the chromatin 3D folding during spermatogenesis and its relation to insulator proteins and gene expression

Cell Reports

Three-Dimensional Genomic Structure and Cohesin Occupancy Correlate with Transcriptional Activity during Spermatogenesis

Graphical Abstract
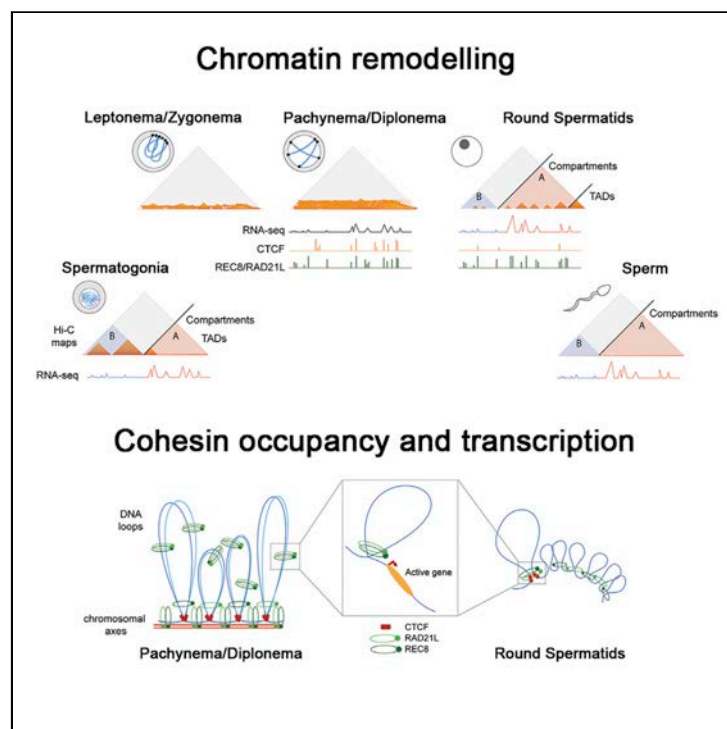
Authors

Covadonga Vara, Andreu Paytuví-Gallart, Yasmina Cuartero, ..., Paul D. Waters, Marc A. Marti-Renom, Aurora Ruiz-Herrera

Correspondence

martirenom@cnag.crg.eu (M.A.M.-R.), aurora.ruizherrera@uab.cat (A.R.-H.)

In Brief

The formation of mammalian germ cells involves dramatic chromosomal movements and chromatin remodeling, whose regulatory pathways are far from understood. Vara et al. show how the dynamics of insulator proteins' occupancy and transcriptional activity are coupled during the 3D genome re-organization that takes place during mouse spermatogenesis.

Highlights

- Meiotic and post-meiotic cells differ in their hierarchal genome organization

- Meiotic cohesin loading correlates with gene expression and local insulation

- Sperm cells contain compartments with active transcription but no clear TADs

- Genes that escape X chromosome silencing associate with promoter cohesin occupancy

CellPress

Cell Reports
Article

Three-Dimensional Genomic Structure and Cohesin Occupancy Correlate with Transcriptional Activity during Spermatogenesis

Covadonga Vara,[1,2,12] Andreu Paytuví-Gallart,[1,2,3,12] Yasmina Cuartero,[4,5,12] François Le Dily,[4,5] Francisca Garcia,[6] Judit Salvà-Castro,[1,2] Laura Gómez-H,[7] Eva Julià,[4,8] Catia Moutinho,[5] Riccardo Aiese Cigliano,[3] Walter Sanseverino,[3] Oscar Fornas,[4,9] Alberto M. Pendás,[7] Holger Heyn,[5,9] Paul D. Waters,[10] Marc A. Marti-Renom,[4,5,9,11,*] and Aurora Ruiz-Herrera[1,2,13,*]

[1]Departament de Biologia Cel.lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès 08193, Spain
[2]Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès 08193, Spain
[3]Sequentia Biotech, Carrer Comte D'Urgell 240, Barcelona 08036, Spain
[4]Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology (BIST), Carrer del Doctor Aiguader 88, Barcelona 08003, Spain
[5]CNAG-CRG, Centre for Genomic Regulation, The Barcelona Institute of Science and Technology (BIST), Baldiri Reixac 4, Barcelona 08028, Spain
[6]Unitat de Cultius Cel.lulars, Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès 08193, Spain
[7]Molecular Mechanisms Program, Centro de Investigación del Cáncer and Instituto de Biología Molecular y Celular del Cáncer (CSIC-Universidad de Salamanca), Salamanca 37007, Spain
[8]Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Carrer del Doctor Aiguader 88, PRBB Building, Barcelona 08003, Spain
[9]Pompeu Fabra University (UPF), Doctor Aiguader 88, Barcelona 08003, Spain
[10]School of Biotechnology and Biomolecular Sciences, Faculty of Science, UNSW Sydney, NSW 2052, Australia
[11]ICREA, Pg. Lluís Companys 23, Barcelona 08010, Spain
[12]These authors contributed equally
[13]Lead Contact
*Correspondence: martirenom@cnag.crg.eu (M.A.M.-R.), aurora.ruizherrera@uab.cat (A.R.-H.)
https://doi.org/10.1016/j.celrep.2019.06.037

SUMMARY

Mammalian gametogenesis involves dramatic and tightly regulated chromatin remodeling, whose regulatory pathways remain largely unexplored. Here, we generate a comprehensive high-resolution structural and functional atlas of mouse spermatogenesis by combining *in situ* chromosome conformation capture sequencing (Hi-C), RNA sequencing (RNA-seq), and chromatin immunoprecipitation sequencing (ChIP-seq) of CCCTC-binding factor (CTCF) and meiotic cohesins, coupled with confocal and super-resolution microscopy. Spermatogonia presents well-defined compartment patterns and topological domains. However, chromosome occupancy and compartmentalization are highly re-arranged during prophase I, with cohesins bound to active promoters in DNA loops out of the chromosomal axes. Compartment patterns re-emerge in round spermatids, where cohesin occupancy correlates with transcriptional activity of key developmental genes. The compact sperm genome contains compartments with actively transcribed genes but no fine-scale topological domains, concomitant with the presence of protamines. Overall, we demonstrate how genome-wide cohesin occupancy and transcriptional activity is associated with three-dimensional (3D) remodeling during spermatogenesis, ultimately reprogramming the genome for the next generation.

INTRODUCTION

Mammalian genomes are packaged into a tailored chromatin structure, the regulation of which depends on several superimposed layers of organization, including epigenetic modifications (of both the DNA and nucleosomes) and the higher-order organization of chromatin compartments inside the nucleus. This organization is achieved by chromatins folding into loops, topologically associating domains (TADs), and compartments (A and B), which can ultimately influence transcriptional activity (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Rao et al., 2014). How these different levels of chromatin organization interact during the cell cycle has just begun to be elucidated (Dekker et al., 2013). In somatic cells, the highly compartmentalized folding of the genome in interphase is lost during mitosis, when chromosomes are linearly organized in consecutive chromatin loops (Gibcus et al., 2018; Naumova et al., 2013). Recent studies in mice have suggested remarkable chromatin architecture reprogramming during the formation of germ cells (Alavattam et al., 2019; Patel et al., 2019; Wang et al., 2019) and early development (Du et al., 2017; Flyamer et al., 2017). However, how the higher-order chromatin organization is configured during all stages of spermatogenesis, and how insulator proteins and

cohesins determine this organization to regulate transcription activity, remains largely unexplored.

Germ cells represent a unique cell model, where unipotent diploid cells (gonia) undergo extensive cellular differentiation (gametogenesis) to form highly differentiated cells (oocytes and sperm) that ultimately form a totipotent embryo. In the case of mammalian males, this complex process is divided into three stages: (1) proliferation and differentiation of spermatogonia (Spg); (2) meiosis, a reductional division that produces haploid cells through two consecutive cell divisions (meiosis I and meiosis II); and (3) spermiogenesis, where round spermatids (RSs) are transformed into densely compacted spermatozoa. These sequential developmental stages involve dramatic and tightly regulated chromosomal re-organization and chromatin remodeling. It is during the first meiotic prophase (prophase I) that homologous chromosomes align, pair, synapse, and recombine. All these processes are interconnected through four sequential stages: leptonema, zygonema, pachynema, and diplonema (Handel and Schimenti, 2010).

At leptonema, chromosomes cluster by their telomeres to the nuclear envelope in the bouquet (Reig-Viader et al., 2016). This structure promotes the pairing of homologous chromosomes by the formation of proteinaceous structures along chromosomes formed by cohesins (i.e., REC8 and RAD21L; Gutiérrez-Caballero et al., 2011; Llano et al., 2012) and proteins of the synaptonemal complex (SC). Meiotic recombination is triggered by the formation of double strand breaks (DSBs), caused by the endonuclease protein SPO11 (Keeney et al., 1997). DSBs are then initiated at zygonema, leading to synapsis between homologous chromosomes. Subsequently, at pachynema, SCs are completely established, creating bivalent structures with resolved recombination producing crossover events. At diplonema, homologous chromosomes start to segregate by the disassembly of SCs to produce spermatocytes II, which undergo a second meiotic division resulting in RSs. Finally, spermatids become sperm ready for fecundation through spermiogenesis, a differentiation stage that includes changes in cell morphology and DNA packaging via the replacement of histones by protamines (testis-specific histone variants).

Despite recent analysis of genome conformational changes in male germ cells (Alavattam et al., 2019; Patel et al., 2019; Wang et al., 2019), how the higher hierarchal level of genome organization is related to gene expression and insulator proteins during spermatogenesis remains unknown. It has been generally accepted that there are two waves of active transcription: one before entering meiosis and a second in the transition from RSs to sperm (Sassone-Corsi, 2002). However, recent transcriptome analysis in germ cells suggests that these two transcriptional waves might take place earlier (da Cruz et al., 2016), which evidences a finely tuned regulation of chromatin remodeling and active transcription. Here, we implement a reproducible flow cytometry protocol to isolate enriched male mouse germ cell populations representing all stages of spermatogenesis: premeiotic (Spg), meiotic (leptonema, zygonema, pachynema, and diplonema), and post-meiotic cells (RSs and sperm). On these sorted germ cell populations, we performed genome-wide chromosome conformation capture analysis (*in situ* Hi-C [chromosome conformation capture sequencing]), coupled with RNA sequencing (RNA-seq) and chromatin immunoprecipitation sequencing (ChIP-seq) of CCCTC-binding factor (CTCF) and meiotic cohesins (Figure 1A). These data have permitted the comprehensive study of the close interplay between chromatin higher-order organization dynamics and function during mouse spermatogenesis.

RESULTS

Dynamic Overall Chromatin Structure Reorganization during Spermatogenesis

To unveil changes in chromosome conformation during spermatogenesis, we developed a reproducible fluorescence-activated cell sorting (FACS) protocol to obtain, based on DNA content and chromatin complexity, highly enriched (90.4% average enrichment) cell fractions for Spg, primary spermatocytes at the leptonema-zygonema (L/Z) and pachynema-diplonema (P/D) stages, RSs, and sperm (Figures 1B, 1C, and S1) (STAR Methods). For each germ cell fraction, as well as for a mouse primary fibroblast cell line as a somatic profile, we performed *in situ* Hi-C (Rao et al., 2014) (Figures 1D–1F and S1). After filtering the raw Hi-C interactions, an average of 237.86 million valid interactions were obtained per cell type (Tables S1 and S2). The comparison between biological replicates resulted in highly reproducible Hi-C maps (Figure S2).

Genome organization changed during spermatogenesis (Figures 1F and 2A–2F), as reflected by the analysis of distance-dependent interaction frequencies (Figure 2C) and inter- and intra-chromosomal interaction ratios (Figure 2D). Fibroblasts showed high inter- and intra-chromosomal interaction ratios (>0.6) inversely correlated with chromosomal size (p < 0.001), suggesting distinct chromosomal compartmentalization within nuclei (Figure 2D). In contrast, inter- and intra-chromosomal interaction ratios decreased 2-fold to about 0.3 for all chromosomes in Spg, suggesting that a commitment to enter meiosis is accompanied by a drastic remodeling of chromosomal territories within the nucleus. Importantly, this decrease in inter- and intra-chromosomal interactions was concomitant with the dynamic changes of the so-called A-B compartments (Lieberman-Aiden et al., 2009). Likewise, fibroblasts and Spg, both in interphase, shared similar contact probability patterns at short distances (from 0.5 to 7 Mbp) and had the lowest intermediate interactions (between 1 and 10 Mb) (Figure 2C). However, at larger genomic distances (>10 Mb), fibroblasts showed a slight change in the slope at 10 Mbp with fewer interactions, whereas Spg maintained the same trend up to 100 Mbp.

As meiosis progressed, compartments were mostly lost in primary spermatocytes (L/Z and P/D), coinciding with prophase I, when homologous chromosomes condensate, align, pair, synapse, and recombine (Figures 2B, 3A, and 3B). Consistent with this absence of compartments during prophase I, eigenvector values were close to 0 (Figures 2B, 3A, and 3B), and inter- and intra-chromosomal interaction ratios reached a minimum for all chromosomes (Figure 2D). An exception was the sex chromosomes in P/D, which had detectable variations in the inter- and intra-chromosomal interactions ratio. This unusual pattern most likely reflects meiotic sex chromosome inactivation (MSCI), the process by which the X and Y chromosomes
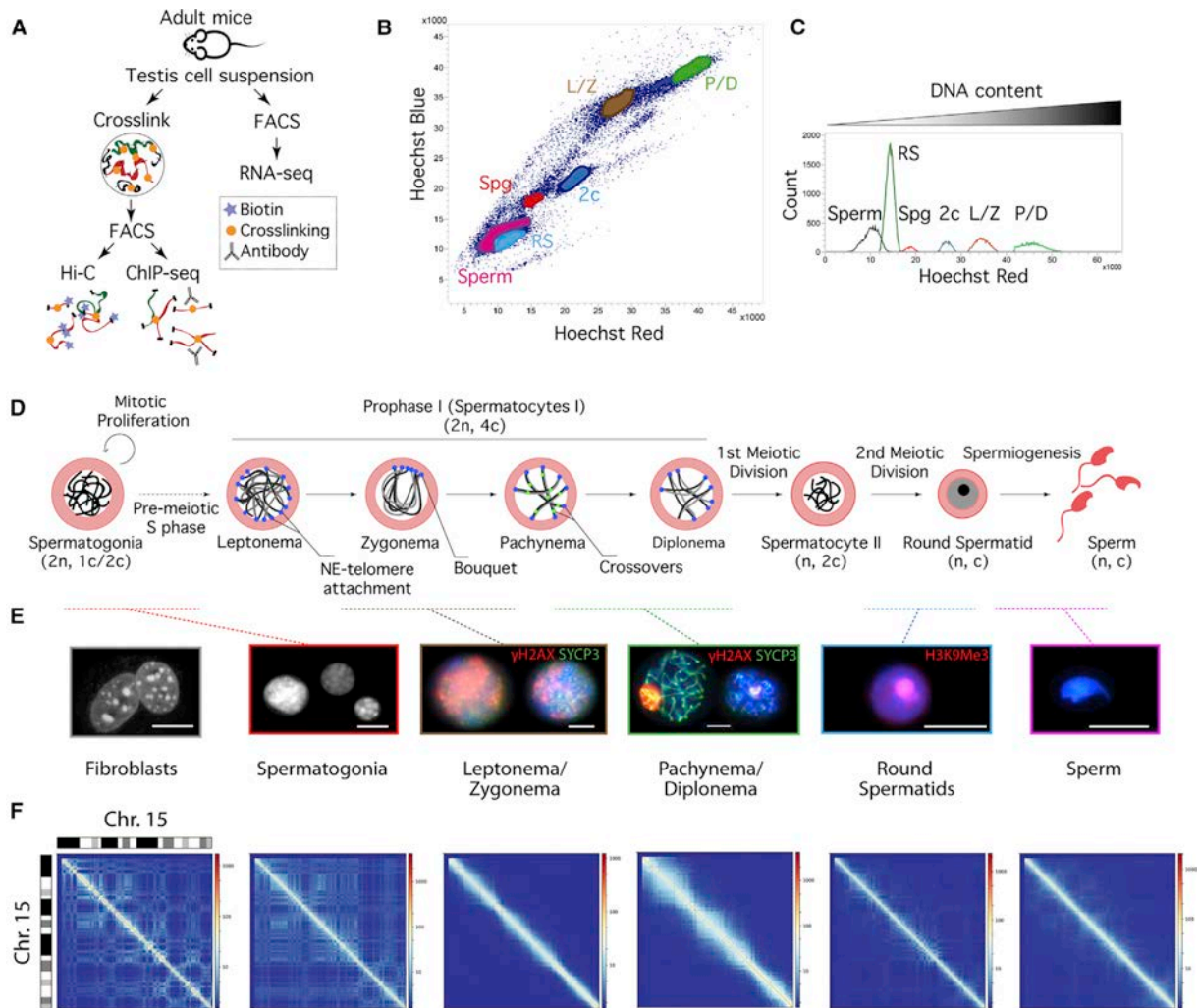
Figure 1. Higher-Order Chromatin Structure during Spermatogenesis

(A) Experimental workflow: Hi-C, ChIP-seq, and RNA-seq on fluorescence-activated cell sorting (FACS)-enriched mouse germ cells.

(B) Flow cytometry Hoechst Blue (UV355-460/50) and Hoechst Red (UV355-670/30) plot showing spermatogonia (Spg), leptonema-zygonema (L/Z), pachynema-diplonema (P/D), secondary spermatocytes (2c), round spermatids (RSs), and sperm. Recovered germ cell fractions presented the following average enrichment: 91% for Spg, 88.7% for L/Z, 89.2% for P/D, 92.9% for RS, and 90% for sperm.

(C) Histogram of the differential DNA content showing cell events for each FACS-isolated germ cell population (Hoechst Red, UV355-670/30).

(D) Overview of the spermatogenesis process (adapted from Reig-Viader et al., 2016). Numbers between parentheses indicate the diploid (2n) haploid (n) number for each cell type and the number of chromatids per chromosome (4c, 2c, or c).

(E) Representative immunofluorescence images showing DAPI-stained DNA (gray/blue) and specific meiotic proteins for the different cell populations included in the present study. Fibroblasts and Spg have DAPI-stained DNA in gray. For L/Z and P/D, DAPI is shown in blue, SYCP3 in green, and γH2AX in red. The image represents a mosaic of two individual captured cells. In RSs, DAPI is blue, and H3K9me3 (marker for the constitutive heterochromatin at centromeres- chromocenters) is red. Scale bars, 10 um.

(F) Iterative correction and eigenvector decomposition (ICE)-corrected Hi-C matrices for chromosome 15 at a 50-kbp resolution for the cell types analyzed. Deep blue lines indicate non-mapped bins.

See also Figure S1.

are transcriptionally inactivated in primary spermatocytes, forming the sex body at the periphery of the nucleus (Turner, 2007). Another structural feature characteristic of early meiosis that was revealed in our analyses was the bouquet structure. We detected high inter-chromosomal contact between telomeres in

primary spermatocytes, being more prominent in L/Z (Figures 2E and 2F). Both L/Z and P/D had an enrichment of counts at the most proximal sub-telomeric bins, but cells at L/Z had significantly higher interactions (Wilcoxon test, $p < 0.05$) than any other cell type (Figures 2E and 2F). The analysis of the
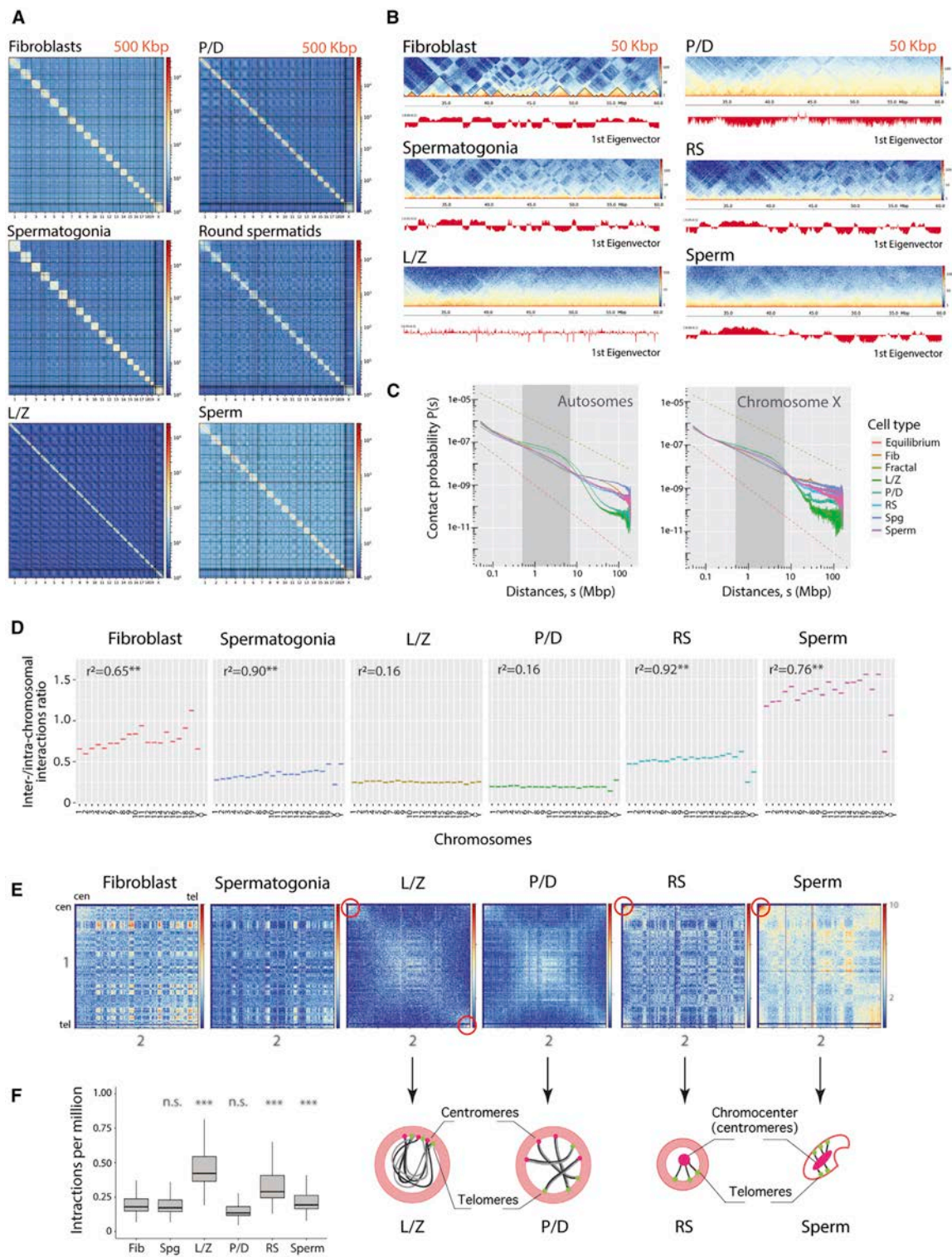
(legend on next page)

distance-dependent interaction frequencies revealed a distinct chromosome organization for primary spermatocytes, with two abrupt changes in slope: the first at 2.5−4.5 Mbp and the second at 40 Mbp (Figure 2C).

After meiosis, haploid cells (RSs and sperm) had a distinctive higher-order chromatin structure. Although A-B compartments re-appeared after being lost in previous stages, they presented as a blurry plaid pattern of larger mean size (0.86 Mbp in RSs and 0.93 Mbp in sperm) compared to Spg (Figures 2F, 3B, S3A, and S3B; Table S3). Interestingly, the proportion of genomic bins with the same compartment status (A or B) was higher between haploid cells (r^2 = 0.80) than between haploid cells and Spg (RSs versus Spg, r^2 = 0.58; sperm vs Spg, r^2 = 0.48). This pattern of differential compartmentalization in RSs was also validated by three-dimensional fluorescence *in situ* hybridization (3D-FISH), where physical distances between pairs of loci increased compared to Spg and fibroblasts (Figure S4; Table S4). Moreover, the inter- and intra-chromosomal interaction ratio values were higher in RSs for all chromosomes, with the exception of sex chromosomes (Figure 2D), which are known to form post-meiotic sex chromatin (PMSC) attached to the chromocenters (Namekawa et al., 2006). Thus, the formation of PMSC correlates with chromatin remodeling that results in low inter- and intra-chromosomal interactions. Remarkably, in sperm, inter-chromosomal interactions were greater than intra-chromosomal interactions, compared to previous stages, and inversely correlated with chromosomal size (p < 0.001) (Figure 2D). Since ratio values were higher than in fibroblasts, the higher-order chromatin structure is likely densely packed in sperm but remains in chromosome territories. The sex chromosomes did not follow the autosomal pattern; the interactions ratio decreased by 2-fold (Figure 2D). In both haploid cell types, we detected higher inter-chromosomal contact between centromeres, suggestive of the presence of the chromocenters (Figures 2E and 2F). In fact, it is known that in mouse sperm, centromeres are located at the center of the nucleus, while telomeres attach to the nuclear envelope (Haaf and Ward, 1995). This particular compartmentalization in both cell types was also reflected by a decrease in interaction frequencies as genomic distance increased (Figure 2C).

These observations indicate that the genome suffers a major structural re-organization during spermatogenesis with dynamic and dramatic changes in chromosome occupancy and compartments.

Two Rounds of TAD Reorganization in Primary Spermatocytes and Sperm

To further investigate the dynamics of the higher-order chromatin structure at the sub-megabase scale, we identified TADs and examined the robustness of their boundaries using TADbit (Serra et al., 2017) at 50-Kbp resolution. Similar to the A-B compartment patterns, TADs were well defined in both fibroblasts and Spg (Figure 3C). In primary spermatocytes, however, there was a substantial reduction in the variance of TAD insulation score as well as an increase in TAD size, especially in L/Z (Figures 3C−3E). TAD insulation scores were partially recovered in RSs but, in contrast to previous observations (Jung et al., 2017; Wang et al., 2019), were drastically reduced in sperm cells (close to 0) (Figures 3D, 3E, and S5).

A total of 2002 TADs, with an average length of 1.3 Mbp, were identified in fibroblasts, which was more than in Spg (834 TADs, mean size of 3.26 Mbp) (Figure 3C; Table S3). Although fewer TADs were identified in primary spermatocytes (305 TADs in L/Z and 294 TADs in P/D), their boundaries had high strength scores (74.25% of TADs in L/Z, and 79.59% in P/D, had scores >9; Table S3). This pattern contrasted with RSs, which had a large number of small TADs (n = 4,649) with low border strength scores (Figure 3C; Table S3). Meta-border plots confirmed this dynamic (Figure 3F). These results demonstrate that TADs also underwent a large reorganization in genome structure during spermatogenesis, concomitant to larger changes in chromosome territories and compartments.

In summary, our data suggest at least four distinct patterns of chromatin interactions during spermatogenesis progression: (1) interphase-like organization (e.g., Spg); (2) a condensed pattern in prophase I (i.e., L/Z and P/D) where A-B compartments and TADs are largely lost; (3) RSs with a blurry compartment plaid pattern; and (4) sperm where TADs are reduced, but A-B compartments are observed.

Functional Compartment Switching during Spermatogenesis

To assess whether A-B compartmentalization changes during spermatogenesis correlated with differential expression of resident genes, we investigated changes of compartment type between the stages where compartments could be clearly observed. The proportion of the genome organized in the

---

Figure 2. Chromosomal Organization in Interphase, Pre-meiotic, Meiotic, and Post-meiotic Cells

(A) Genome-wide ICE-corrected heatmaps at 500 kbp for the cell types analyzed.

(B) Chromosome 18 region-specific ICE-corrected heatmaps at 50 kbp (from 20 Mbp to 60 Mbp), depicting compartment signal (first eigenvector) for all cell types.

(C) Contact probability P(s) as a function of genomic distance in all cell types for autosomes (left panel) and the X chromosome (right panel). Discontinuous straight lines correspond to the fractal (green) and equilibrium (red) models (Lieberman-Aiden et al., 2009). Gray-shadowed area expands the genomic region from 0.5 to 7 Mbp.

(D) Inter- and intra-chromosomal interaction ratios for each chromosome and cell type. Correlation values (**p < 0.001) between the inter- and intra-chromosomal ratio and chromosomal size (autosomes only) are shown for each cell type.

(E) Heatmaps showing normalized inter-chromosomal interactions between chromosomes 1 and 2 in all cell types. Red circles represent high-contact regions.

(F) Left panel: boxplots depicting inter-chromosomal interactions per million at sub-centromeric regions (from the centromere up to 3.5 Mbp) for all cell types (Wilcoxon test, **** < 0.0001; n.s., not statistically significant when compared to fibroblasts). Right panel: schematic representation of chromosomes and centromeres and telomeres in L/Z, P/D, RSs, and sperm. Dots represent centromeres (pink) and telomeres (green). In RSs and sperm, all centromeres associate in the center of the cell forming the chromocenter.

Fib, fibroblast; Spg, spermatogonia; L/Z, leptonema-zygonema; P/D, pachynema-diplonema; RS, round spermatids; cen, centromeres of acrocentric chromosomes; tel, telomeres of non-centromeric ends. See also Figures S2–S5.
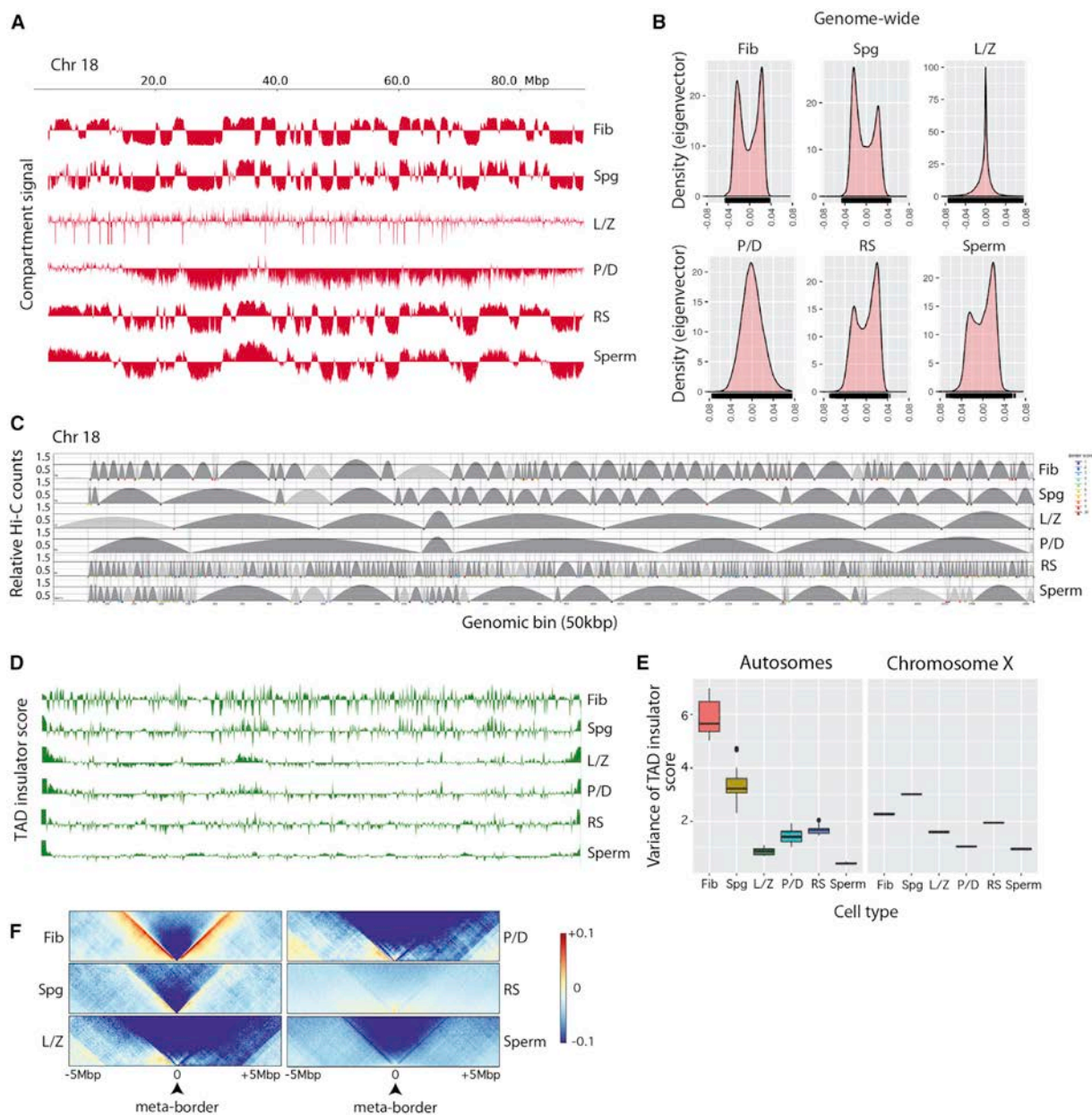
Figure 3. Chromosome-specific A-B Compartment Profiles and TAD Signals

(A) Compartment signal (first eigenvector) across chromosome 18.

(B) Density plots of eigenvector values considering autosomes.

(C) TAD border alignments along chromosome 18. Dark gray arches represent TADs with higher intra-TAD interactions than expected. TAD border robustness (from 1 to 10) is represented by a color gradient.

(D) Representation of TAD insulator score in mouse chromosome 18.

(E) Variance of the TAD insulation scores for autosomes (left panel) and the X chromosome (right panel).

(F) Meta-plots for all TAD boundaries detected in fibroblasts (n = 2002), Spg (n = 834), L/Z (n = 305), P/D (n = 294), RSs (n = 4649), and sperm (n = 1042).

Fib, fibroblast; Spg, spermatogonia; L/Z, leptonema-zygonema; P/D, pachynema-diplonema; SpII, spermatocytes II; RS, round spermatids. See also Figures S3–S5.
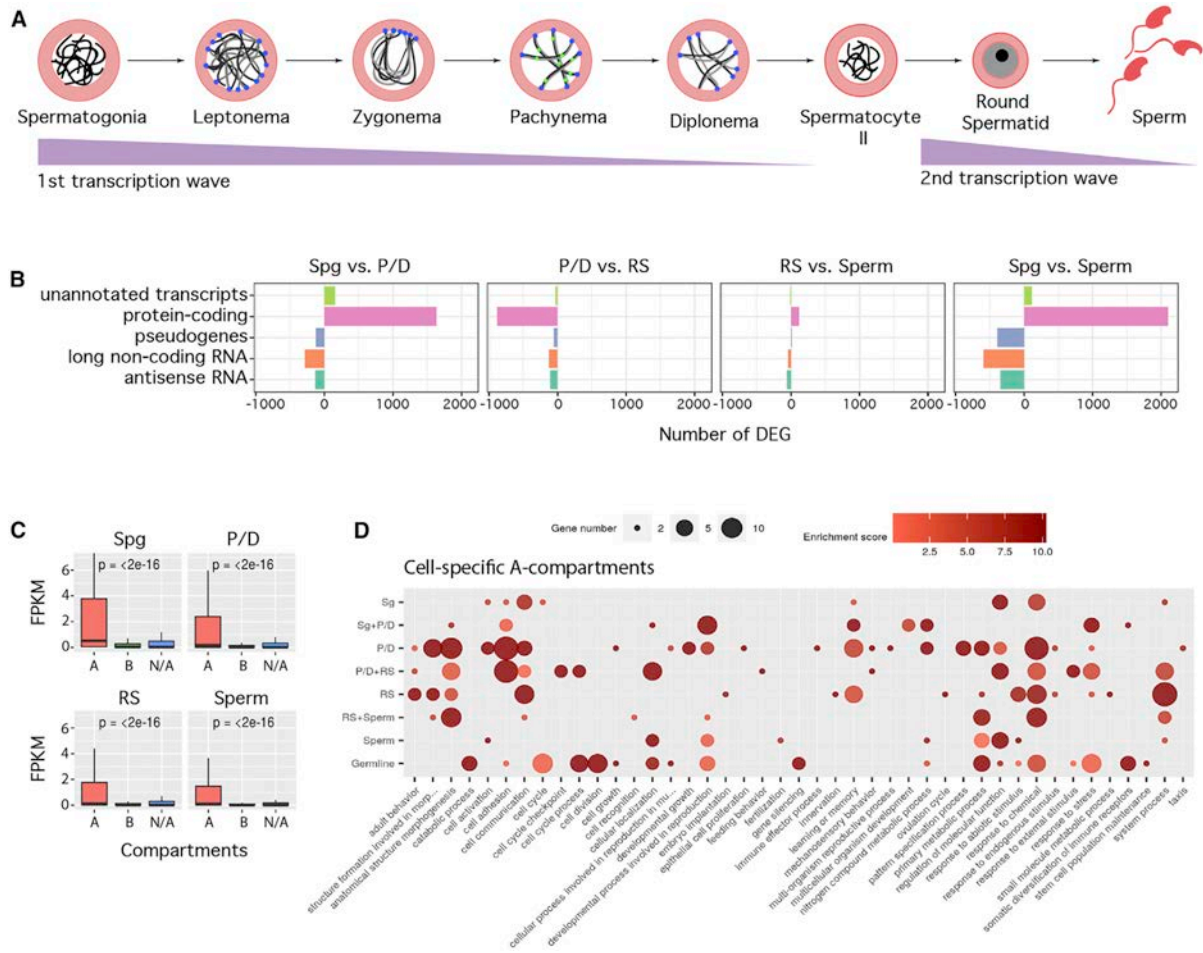
Figure 4. Dynamics of Gene Expression during Gametogenesis

(A) Two extensive waves of transcription take place during spermatogenesis.

(B) Differentially expressed genes (DEGs) for each pairwise comparison, including known nuclear protein-coding RNA (RefSeq), long noncoding RNA, antisense RNA, processed transcripts, and mitochondrial nuclear protein-coding RNA.

(C) Boxplots representing genome-wide expression (as fragments per kilobase of transcript per million mapped reads [FPKM] values) according to A-B compartment assignment (N/A, not assigned compartments).

(D) Gene Ontology (GO) analysis of expressed genes in cell-specific A compartments. Bubble size represents the number of genes assigned to each GO. Only GO terms with two or more genes are represented.

Spg, spermatogonia; P/D, pachynema-diplonema; RS, round spermatids.

A compartment was 45.7% in fibroblast and reduced to 39.4% in Spg, before a rise in RSs (46.9%) and sperm (48.6%). Since A compartments correlate with open chromatin and active genes (Lieberman-Aiden et al., 2009), the switching of compartment types can provide insight into genome function during spermatogenesis. We therefore performed low-input RNA-seq (four biological replicates) on a selected group of the enriched germ cell populations: Spg, primary spermatocytes (P/D), RSs, and sperm. After filtering, an average of 35.37 million paired-end reads was obtained per cell type (Table S5).

The total number of expressed genes decreased as spermatogenesis progressed, with 19,145 expressed in Spg, 15,480 in P/D, 14,706 in RSs, and 13,646 in sperm. Pairwise differential gene expression analysis between cell types generated lists of differentially expressed genes (DEGs), which were classified as (1) protein-coding RNA; (2) long non-coding RNA (lncRNA); (3) antisense RNA (asRNA); (4) processed and unprocessed pseudogenes; or (5) unannotated and unconfirmed transcripts (Figure 4; Table S6). The majority (76.55%) of DEGs were protein-coding, with their abundance reducing as spermatogenesis progressed. This was coupled with an increased expression of lncRNAs and non-coding asRNAs and pseudogenes (Figure 4; Table S6). Consistent with the global shutdown in gene expression that occurs during meiosis (Sassone-Corsi, 2002), the net balance between pairwise comparisons (Spg versus P/D, RSs versus sperm, and Spg versus sperm) was negative, which

suggests higher transcriptional activity in RSs than in primary spermatocytes (Figures 4A and 4B).

Consistent with a correlation between chromatin remodeling and active transcription, genes in the A compartments had significantly higher expression than those in the B compartments (Figures 4C and 4D). As expected, genes related to spermatogenesis (e.g., morphogenesis and cell differentiation) were enriched in the A compartments (Figure 4D). In RSs, during the second wave of transcription (Figure 4A), the most representative Gene Ontology (GO) term was "system process," which included 27 genes involved in sensory perception, including olfactory receptors. Likewise, genes with important roles in fertilization were also expressed in RSs, including the acrosome reaction (e.g., *Plcz1* and *Smcp*). Therefore, the transformation of RSs into spermatozoa was accompanied by the transcription of genes related to spermiogenesis and sperm function located in newly created A compartments.

Differential CTCF and Cohesin Loading Correlates with Gene Expression and Chromatin Remodeling

Western blots confirmed the presence of CTCF and meiotic cohesins (REC8 and RAD21L) in whole-testis protein extracts. Immunofluorescence (IF) then revealed a previously unreported pattern for CTCF, with signal along all chromosome axes in primary spermatocytes that were more intense on autosomals than on the X (Figure S6). There was a weak CTCF signal in RSs and a clear, cloud-like signal for cohesins (Figure S6). ChIP-seq analysis confirmed these differences in IF signal intensity and density (Figure S6). In fact, we detected 19,347 CTCF ChIP-seq peaks in primary spermatocytes (P/D), with the vast majority (97.1%) being lost in RSs (Figures 5A and 5B).

We then performed ChIP-seq in P/D and RSs on the meiotic cohesins REC8 and RAD21L. In primary spermatocytes (P/D), we detected 11,618 REC8 and 9639 RAD21L peaks distributed across the genome, with substantial overlap (55.8% for RAD21L and 46.3% for REC8) (Figures 5A and 5B). Of the total peaks, 3999 were common to all CTCF (20.6%), RAD21L (41.5%), and REC8 (34.4%). In RSs, we found 11,559 REC8 peaks and 12,507 RAD21L peaks, of which 6383 (51% of RAD21L and 55.2% of REC8) overlapped (Figures 5A and 5B). The vast majority (90.5%) of REC8 peaks observed in primary spermatocytes were maintained in RSs, which contrasted RAD21L, where 40% of the peaks were specific to RSs.

Despite the overlap of RAD21L and REC8 peaks detected genome-wide in P/D, the distribution of REC8 and RAD21L immunolabeling along the axes was not continuous or significantly correlated, as measured by super-resolution microscopy (r^2 = 0.15; Figure S6). The close proximity of meiotic cohesins genome-wide (on average, peaks are scattered every 264.3 Kbp for RAD21L and 219.2 Kbp for REC8) was much lower than the estimated DNA loop size at the axes for mouse pachynema (1.5−2 Mbp; Patel et al., 2019). This pattern suggests that the peaks correspond not only to cohesins loaded at the chromosomal axes (expected every 1.5−2 Mbp), but also to DNA loops out of the axes. Given the close interplay among the organization of chromosomal axes, DNA loops, and the formation of DSBs during early prophase I (Kleckner et al., 2003; Ruiz-Herrera et al., 2017; Wang et al., 2015), we examined the correlation

between cohesins and DSBs (SPO11-oligos hotspots; Lange et al., 2016) and H3K4me3 (Brick et al., 2012). Interestingly, both SPO11 hotspots and H3K4me3 marks correlated with cohesin occupancy (STAR Methods; p < 0.001), concomitant with open chromatin states. Since DSBs are known to occur genome-wide before being recruited at the chromosomal axes to be repaired (Lange et al., 2016), the observed correlation between SPO11 hotspots and cohesion peaks also suggests the presence of cohesins in DNA loops out of the axes.

Remarkably, we detected a correlation of cohesin genomic distribution with gene expression and local insulation in both primary spermatocytes and RSs (Figures 5 and 6). In P/D, most cohesin peaks (80.7% of RAD21L and 83.3% of REC8) were located in promoter regions (less than 2 kbp from the transcriptional start site [TSS]) of genes with significantly (p < 0.01) higher expression than genes without promoter-associated cohesin peaks (Figures 5C−5E). In RSs, cohesin peaks associated with TSSs were reduced to 77.7% for REC8 and 45.5% for RAD21L, suggesting an unequal re-distribution of meiotic cohesins later in spermatogenesis (Figures 5C and 5D). Regardless of RAD21L re-distribution, the overlapping of genes with RAD21L in their promoters between P/D and RSs was statistically significant (p = 0e+00, Fischer's exact test). Despite this reduction of peaks in RSs, genes with cohesin peaks in their TSSs still had significantly higher expression than genes without peaks (Figures 5E and 5F). In contrast, only 25.9% of the CTCF peaks were within promoter regions (i.e., 2 kbp upstream of TSSs) in primary spermatocytes, with the majority located in intergenic regions (Figure 5C). In all cases, CTCF and cohesin peaks were preferentially located (nearly 80% of peaks) in A compartments (permutation test, STAR Methods; p < 0.001), consistent with a correlation between chromatin remodeling and active transcription. Although the TAD insulation score was reduced in primary spermatocytes (Figures 3D and 3E), some TADs still remained, with boundaries enriched for CTCF and cohesins (STAR Methods; p < 0.01) (i.e., genomic regions with lower TAD insulation scores; Figure 6B) and associated with gene expression (Figure 6C). Meta-border plots of CTCF and cohesin peaks confirmed this trend (Figure 6D). The same pattern was observed in RSs (Figures 6B−6D), suggesting that despite global chromatin remodeling, local insulation is maintained by insulator proteins and can affect gene expression.

Expressed genes with CTCF, RAD21L, and REC8 in their promoters in P/D had GO term enrichments related to protein regulation, modification, and polymerization (e.g., *Usp42*), as well as DNA repair and cellular response to DNA damage stimulus (e.g., *Herc2*), suggesting a regulatory role in spermatogenesis progression. Interestingly, genes with CTCF at their promoters were involved in the transcriptional machinery (e.g., *Nsa2*), whereas genes with both cohesins but not CTCF were essential for posterior neural development of the embryo (e.g., *Cdh2*). In RSs, CTCF was almost absent, but remarkably, when it co-localized with promoters, it was associated with genes implicated in key pathways of embryo development (e.g., *Nanog*). Genes with both cohesins in their promoters were related to the regulation of cell growth and the Wnt signaling pathway (e.g., *Amer3*) and general nervous system development (e.g., *Ccd88a*; Figure 5F). Genes with RS-specific RAD21L peaks had functions involved
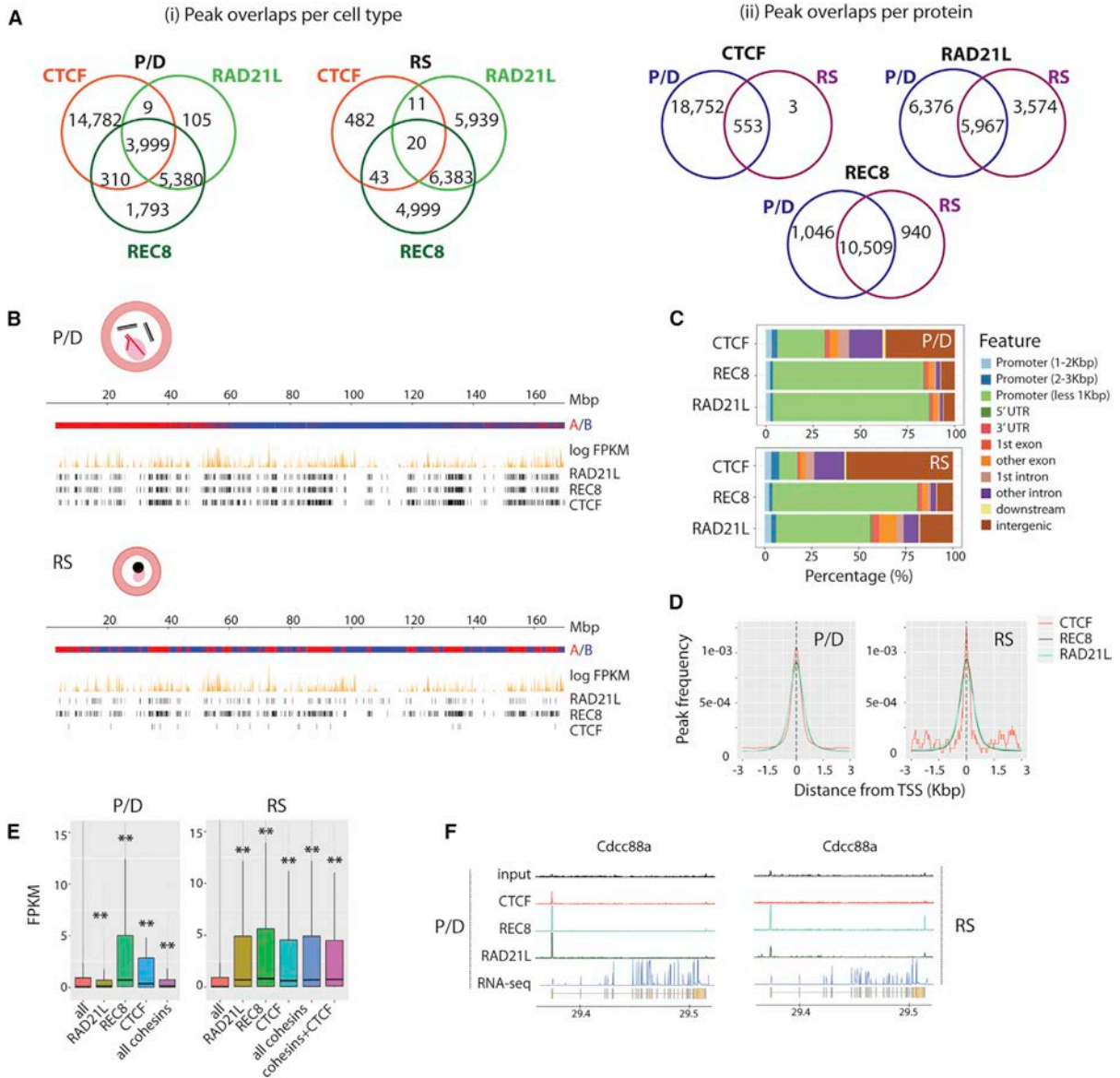
Figure 5. CTCF and Cohesin Profiles in Primary Spermatocytes and RSs

(A) Venn diagrams for CTCF and cohesins in P/D and RSs considering peak overlaps per cell type and peak overlaps per protein.

(B) Representative examples of CTCF and cohesins' genomic distribution along chromosome 1 in P/D and RSs. For each cell type, A-B compartments, gene expression (represented as log FPKM), CTCF peaks, and cohesin peaks are displayed.

(C) Genome-wide distribution of CTCF and cohesin-occupied sites in relation to TSSs and other genomic features in P/D and RSs.

(D) Insulator peak frequencies relative to TSSs of genes in P/D and RSs.

(E) Boxplots representing expression (FPKM values) of genes with CTCF and cohesin peaks located at the TSS. Asterisks represent statistically significant differential gene expression when compared with all genes in the mouse genome ($p < 0.01$).

(F) Examples of CTCF and cohesion-occupied sites in P/D and RSs for the expressed gene Cdc88a.

See also Figure S6.

in oligondendrocyte differentiation and cardiac ventricle development (e.g., *Notch1*), whereas RS-specific REC8 peaks were found to be involved in nervous system development (e.g., *Kif3c*).

Sex Chromosome Silencing Is Coupled with Higher-Order Chromatin Restructuring

The eutherian mammal X chromosome is composed of evolutionary strata that were isolated from recombination with the
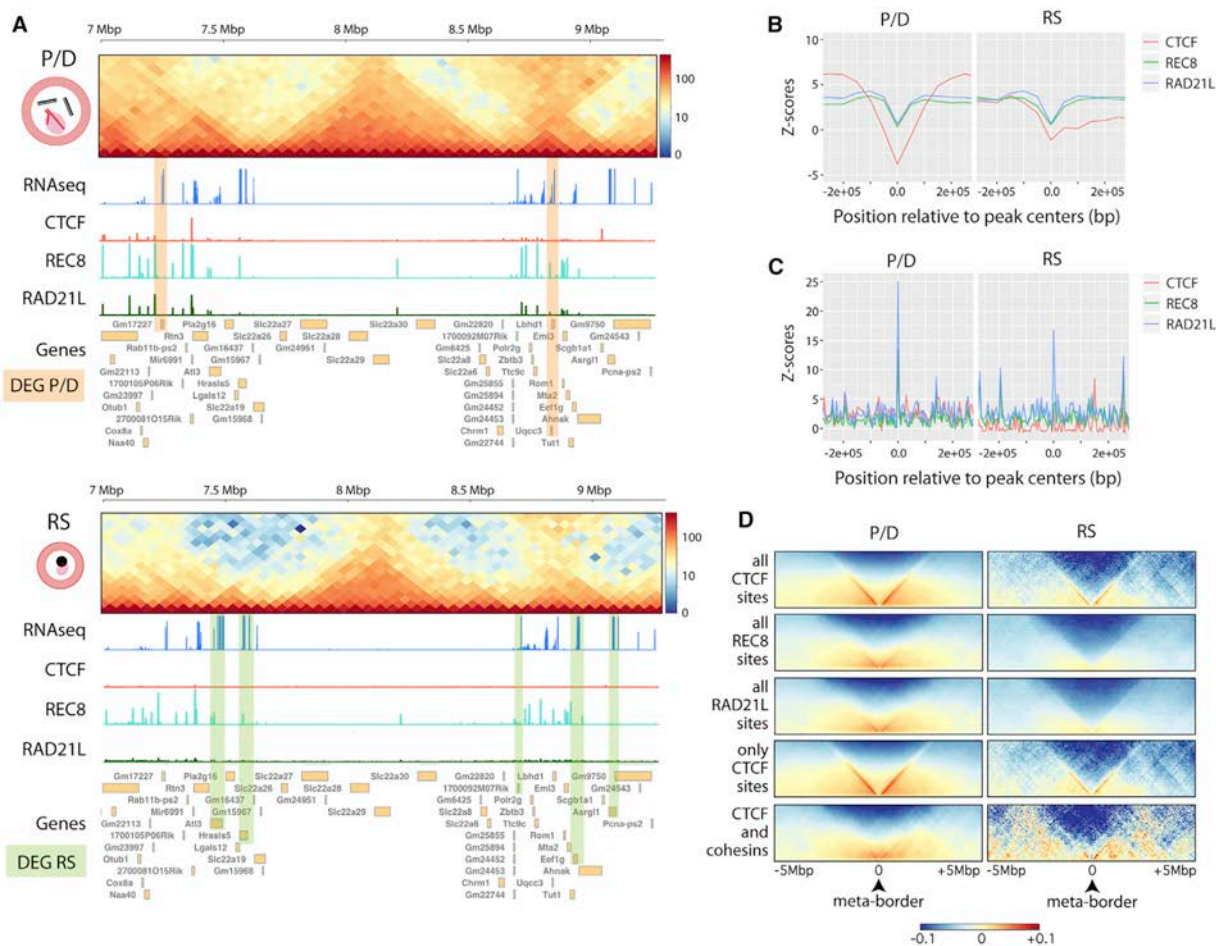
Figure 6. Local Insulation, Cohesin Occupancy, and Gene Expression

(A) Representative examples of CTCF and cohesins' genomic distributions across a specific region of chromosome 19 (from 7 to 9.3 Mbp) in P/D and RSs. For each cell type, Hi-C interaction maps (50-Kbp bins), gene expression (represented as log FPKM), CTCF peaks, cohesin peaks, and genes from NCBI Ref Seq annotation are displayed. Green and orange highlights indicate differentially expressed genes (DEGs) in each cell type: Gm17227 and Uqcc3 for P/D and Atl3; Hrasls5, 1700092M07Rik, Eef1g, and Asrgl1 for RSs.

(B) Distribution of CTCF and cohesin peaks at TAD borders. The y axes represent the TAD insulation Z-score relative to random genomic regions (based on 10,000 permutation tests with randomization, $p < 0.01$).

(C) Distribution of gene expression for CTCF and cohesin peaks located at TAD borders. The y axes represent the TAD insulation Z-score relative to random genomic regions (based on 10,000 permutation tests with randomization, $p < 0.01$).

(D) Meta-plots for all peaks detected in P/D and RSs.

P/D, pachynema-diplonema; RS, round spermatids. See also Figure S6.

Y chromosome at different evolutionary times (Lahn and Page, 1999). In mice, the X chromosome evolutionary strata are highly rearranged, and there are 22 ampliconic regions known to escape MSCI (Mueller et al., 2008) (Figure S7; Table S7). The X chromosome lost its A-B compartment pattern once meiosis was initiated, from L/Z onward (Figures 7A and 7B), consistent with MSCI chromatin remodeling. This translated into an overall reduction in gene expression in P/D compared to Spg, although ampliconic regions were still expressed in P/D, RSs, and sperm (Figure 7C). Importantly, this organization of the X chromosome was maintained in post-meiotic cells (RSs and sperm) (Figures 7A and 7B). In fact, distance-dependent interaction frequencies revealed that genomic interactions at medium genomic distances (1−10 Mbp) were higher in the X chromosome in post-meiotic cells than in autosomes of the same cell type (Figure 2C).

For higher-order chromatin structures at the sub-megabase scale, all chromosomes except X had equivalent TAD dynamics during spermatogenesis. In the X of meiotic and post-meiotic cells, there was a trend for fewer and larger TADs compared to those observed in Spg (Figure 7D; Table S8). The genomic positions of TADs were not equally distributed along the X; in RSs, there were more short TADs detected in the distal regions of
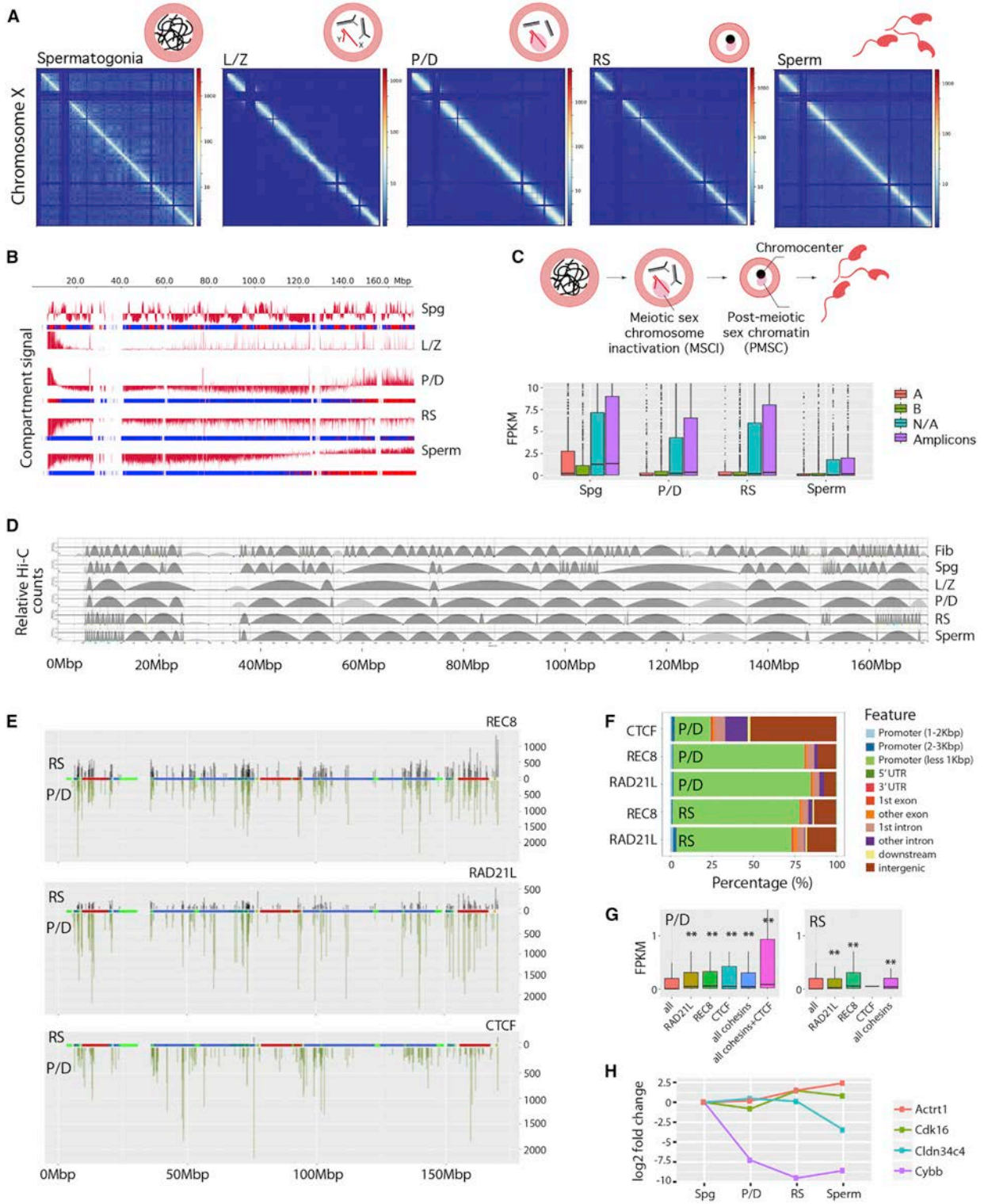
(legend on next page)

the X (Figure 7D), consistent with the increase of autosomal variance of TAD insulation scores in RSs compared to L/Z and P/D (Figure 3E). However, the remainder of the X in RSs maintained the fewer and larger TADs observed in L/Z and P/D. In sperm, the high density of short TADs was only observed in proximal X chromosome long arm (Xq).

Differential CTCF and Cohesin Loading in Autosomes and the X Chromosome

The proportion of cohesin peaks was lower for the X chromosome than for any autosome in both P/D and RSs (Figure 7E). The X represents ~4.9% of the mouse genome, but only accounts for 1.82% of the total RAD21L peaks in P/D and 1.54% of peaks in RSs. Likewise, REC8 peaks on the X chromosome were underrepresented, with 2.12% of the total peaks in P/D and 2.03% of peaks in RSs. The underrepresentation of REC8 ChIP-seq peaks was confirmed by super-resolution microscopy in pachynema, with decreased labeling of REC8 (44.7% ± 10.7%) on the X chromosome axis relative to those of autosomes (Figure S6). In contrast, RAD21L labeling actually increased (25.6% ± 10.6%) on the X (Figure S6).

Overall, 75% of cohesin peaks co-localized with promoter regions (less than 2 kbp upstream of TSSs) in both P/D and RSs in the X chromosome (Figure 7F). These genes had significantly higher expression than genes without promoter-associated cohesin peaks, despite MSCI and PMSC (Figures 7G and 7H), suggesting that genes that escape MSCI correlate with promoter cohesin occupancy. As with the cohesins, CTCF was depleted on the X chromosome relative to the autosomes (1.26% of total peaks were on the X in P/D). However, CTCF was almost completely depleted from the X chromosome in RSs (just two peaks were detected). This was not observed for individual autosomes, on which the proportion of CTCF peaks were generally stable between P/D and RSs. Collectively, these observations point to a novel role for meiotic cohesins in genome organization and function during meiotic prophase I and spermiogenesis.

DISCUSSION

Here, we provide a high-resolution structural and functional atlas of mouse spermatogenesis. Our data reveal the compartmental-ization of meiotic chromosomes in both early and late stages of spermatogenesis, which was reflected at different levels: (1) inter- and intra-chromosomal interaction ratios; (2) distance-dependent interaction frequencies; (3) genomic compartments; (4) topological domains; and (5) occupancy of insulator proteins. We provide evidence of a delicate fine-tuning among chromatin remodeling, architectural proteins, and cell-specific gene expression.

Our analyses complement and extend recently published works (Alavattam et al., 2019; Patel et al., 2019; Wang et al., 2019) with a comprehensive view of the sequential developmental stages during mouse spermatogenesis. Pre-meiotic, meiotic, and post-meiotic cells all presented differences in inter- and intra-chromosomal interaction ratios that were distinct from somatic cells (i.e., fibroblasts). Although Spg maintained the equivalent proportion of A-B compartments as fibroblasts, they showed a drastic remodeling of chromosomal compartmentalization. As the first meiotic prophase begins, additional chromatin remodeling appears. The inter- and intra-chromosomal interaction ratio reaches a minimum in primary spermatocytes, with the A-B compartments almost lost in both L/Z and P/D, in contrast to previous observations (Patel et al., 2019). Thus, chromosomal occupancy and compartmentalization inside nuclei were re-arranged (i.e., higher-chromatin structure relaxation) during prophase I, permitting DNA-scaffold assembly and the formation and repair of DSBs, with no distinction between autosomes.

The distance-dependent interaction we detect suggests differences in previously reported mitotic and meiotic chromosome folding (Gibcus et al., 2018; Naumova et al., 2013; Wang et al., 2019). Prophase I cells display two changes in contact probability: the first between 2.5 and 4.5 Mbp and the second at 40 Mbp. This organization can result from two features of chromosome assembly during prophase I. First, the chromatin is anchored as long DNA loops in a protein scaffold composed of specific meiotic cohesins (e.g., REC8 and RAD21L) (Gutiérrez-Caballero et al., 2011; Llano et al., 2012) and proteins of the SC (e.g., SYCP3) (Henderson and Keeney, 2005), preventing interactions below 40 Mbp. Second, the association of recombination hotspots and cohesins in primary spermatocytes suggests that cohesion-mediated transcription in genomic regions out of the

---

Figure 7. Higher-Order Chromatin Structure and Gene Expression in the X Chromosome

(A) ICE-corrected Hi-C matrices for the X chromosome in mouse germ cells, at a 50-Kbp resolution. Plaid blue regions correspond to non-mapped bins.

(B) Representation of compartment signal (first eigenvector) along the mouse's X chromosome.

(C) Upper panel: Overview of spermatogenesis with two pairs of autosomes (black and gray lines) and the sex chromosomes (X and Y as red lines). Meiotic sex chromosome inactivation (MSCI) characterizes prophase I (shown as a pink cloud; also known as the sex body). In RSs, all centromeres associate in the center of the cell forming the chromocenter, with the X adjacent, forming the post-meiotic sex chromatin (PMSC). Lower panel: Boxplots of expression (FPKM values) of X genes binned according to A-B compartment or campliconic. N/A are not assigned to a compartment or amplicon.

(D) X chromosome TAD alignments. Arches represent TADs with higher (darker gray) and lower (lighter gray) than expected intra-TAD interactions.

(E) CTCF and cohesin distribution on the X in P/D (green) and RSs (black). Campliconic regions are green, and evolutionary strata are displayed, with blue representing older strata and red representing newer strata (see Figure S7 for further details).

(F) X chromosome distribution of CTCF and cohesins' occupancy relative to TSSs and other genomic features in P/D and RS.

(G) Expression (FPKM values) of genes with CTCF and cohesin peaks located at promoter regions (−2 kbp from TSS). Asterisks represent statistically significant differential gene expression when compared with all genes in the mouse genome (p < 0.01). Boxes represent first and third quartiles, whereas black bars in boxes represent the median values.

(H) Expression changes (versus Spg) of representative X genes that reduce expression in P/D and maintain low levels during spermatogenesis (e.g., Cybb), genes that reduce expression in P/D and increase in RSs and sperm (i.e., Cdk16), genes that increase expression in RSs (e.g., Actr1), and genes that increase expression in P/D and then reduce (i.e., Cldn34c4).

Fib, fibroblast; Spg, spermatogonia; L/Z, leptonema-zygonema; P/D, pachynema-diplonema; RS, round spermatids. See also Figure S7.

axes can provide an environment conducive to both gene expression and the formation of DSBs. This is reflected by the interactions observed at shorter distances (2.5–4.5 Mbp).

Compartments re-emerged in post-meiotic cells with a remarkable high-level organizational difference between RSs and sperm. Chromosomal territories were re-established in both cell types but with chromatin less densely packed in RSs, as revealed by the interaction ratio and 3D-FISH. In RSs, intra-chromosomal interactions were greater than inter-chromosomal interactions, the opposite to what was observed in sperm. In both cell types, A compartments (which correlate with gene expression) were larger than in pre-meiotic cells and were over-represented when compared to fibroblasts and Spg. Moreover, the transformation of RSs into spermatozoa was accompanied by the transcription of genes related to spermiogenesis and sperm function located in newly created A compartments. This agrees with a second burst of gene expression in RSs (da Cruz et al., 2016) and supports the more recent idea that sperm are not inactive cells (Jodar et al., 2016; Jung et al., 2017). In fact, it was shown that a large number of sperm promoters are in an active epigenetic state, suggesting that this genetic information can influence embryo development upon fertilization (Jung et al., 2017).

There was a remarkable reduction of TADs in sperm. One strength of our approach that differs from others (Jung et al., 2017; Wang et al., 2019) was the cell-sorting strategy, which permitted the isolation of sperm from the rest of the germ cell populations. Coupled with the Hi-C simulations, this permitted us to attribute higher-order chromatin structures specifically to sperm. Thus, the reduction of the TAD insulation score in sperm can be linked to histone replacement by protamines. Protamines replace most histones in sperm, with the help of transition proteins and H2A histone variants, folding DNA into toroidal subunits at the kbp level (Barral et al., 2017). Because of the structural constraints of protamines, it is tempting to speculate that the compact chromatin organization in sperm is associated with the presence of A compartments at the Mbp scale, which likely correlates with active histone marks in sperm (Jung et al., 2017), but not with the formation of fine-scale TAD structures (i.e., kbp).

The genomic atlas for CTCF and meiotic-specific cohesins presented here provides an unprecedented view of the connection among chromatin remodeling, architectural proteins, and gene expression. Strikingly, the vast majority of cohesin peaks detected in primary spermatocytes were localized within promoter regions of genes, with significantly higher expression than non-cohesin-associated genes. The conflicting patterns of RAD21L on the X chromosome, with reduced ChIP peak numbers but increased super-resolution microscopy labeling intensity, are counterintuitive. The increased labeling intensity at the axis suggests that RAD21L has an important X chromosome scaffolding function, but the reduced number of peaks (compared to the autosomes) indicates that the detected peaks correspond to cohesins that do not interact directly with DNA at the axis. In fact, REC8 and RAD21L bind to both head domains of a structural maintenance of chromosome (SMC) heterodimer, forming a ring-like protein structure topologically encircling sister chromatids to the SC, a core axis from which chromatin

loops emerge (Haering and Jessberger, 2012). This structure could block REC8 and RAD21L access to chromatin, preventing detection at the axis by ChIP-seq. Our data suggest that cohesins associate with active promoters most probably located in DNA loops out of the axes, hinting at a functional role that adds to the well-known structural role of these cohesins in the formation of meiotic chromosomal axes (Llano et al., 2012). In fact, mice with REC8 and RAD21L deficiencies are infertile; meiosis is arrested in early prophase I (Bannister et al., 2004; Herrán et al., 2011), precluding analysis of potential regulatory roles for cohesins during spermiogenesis.

Our results also point to the role of meiotic cohesins in regulating gene expression after meiosis, which was characterized by the occupancy of RAD21L at the TSSs of expressed genes in RSs. We suggest that CTCF and cohesins could have a synergistic role in establishing transcriptional hubs for early embryonic development in meiotic prophase I (where the three proteins co-localize), while also fine-tuning subsequent spermatogenesis progression. Moreover, cohesins in RSs appear to correlate with the expression of genes implicated in early embryonic development (e.g., *Nanog*), providing immediate access to the major molecular pathways involved in organogenesis upon fertilization. These results suggest that some cellular functional activity is predetermined in early spermiogenesis, before RSs are differentiated into sperm, which raises the intriguing possibility that cohesins play a role in such functional predetermination.

Here, we extend initial observations (Alavattam et al., 2019; Patel et al., 2019; Wang et al., 2019) of X chromosome remodeling in prophase I to pre-meiotic and post-meiotic cells. Our functional and structural analyses show that silencing of the X chromosomes (MSCI) is accompanied by distinct changes in the higher-order chromatin structure at different levels: (1) more intra- than inter-chromosomal interactions, a pattern already present in Spg; (2) strong compartmentalization; and (3) a reduction in both the TAD number and signal once prophase I initiates. Remarkably, architectural proteins were remodeled, which included a reduction of cohesin peaks (compared to autosomes) and an almost complete depletion of CTCF. These data demonstrate that a reduction of the cohesin load correlates with an absence of chromosome X compartmentalization (A/B). Despite MSCI, gene expression still takes place, which correlates with meiotic cohesin occupancy but not with CTCF. Altogether, these observations support a role for cohesin in regulating gene transcription.

In summary, we have implemented a robust workflow to provide an integrated structural and functional framework of the 3D organization of the mouse genome in germ cells. We detected a fine-tuned balance among chromatin remodeling, architectural proteins, and gene expression during spermatogenesis. Overall, our results provide insights into how the structural organization of the genome influences cellular differentiation, especially in the context of the dramatic chromatin changes that take place during the formation and differentiation of the mammalian male germline. Future functional perturbation analyses will help us understand the mechanism by which 3D genome folding changes shape transcriptional activity during spermatogenesis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
  - Cell lines
- METHOD DETAILS
  - Fluorescence Activated Cell Sorting (FACS) of mouse male germ cells
  - Fibroblast cell culture
  - In nuclei Hi-C
  - Hi-C data processing, binning and normalization
  - Correlation coefficient analysis
  - Averaged contact probability P(s)
  - Inter-/intra-chromosome interaction ratio
  - Inter-subtelomeric interaction quantification
  - Sperm simulations
  - A-B compartments and TAD calling
  - Compartment switching
  - 3D-FISH and confocal microscopy
  - Western blot analysis of CTCF and cohesins in germ cells
  - Spermatocyte spreads and immunofluorescence analysis of CTCF and cohesins
  - Stimulated emission depletion (STED) microscopy
  - ChIP-sequencing
  - ChIP-seq peak calling and annotation
  - RNA-sequencing
  - RNA profiling
  - Analysis of correspondence between compartments and gene expression
  - Local insulation, cohesin occupancy and gene expression
  - X chromosome evolutionary strata and ampliconic regions
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.06.037.

AUTHOR CONTRIBUTIONS

A.R.-H. conceived and devised the study. C.V., A.P.-G., Y.C., F.L.D., H.H., M.A.M.-R., and A.R.-H. designed experiments. C.V., Y.C., F.L.D., F.G., J.S.-C., L.G.-H., E.J., C.M., O.F., and A.R.-H. performed experiments. C.V., A.P.-G., Y.C., F.L.D., L.G.-H., A.M.P., P.D.W., M.A.M.-R., and A.R.-H. analyzed the data. R.A.C., W.S., A.M.P., H.H., M.A.M.-R., and A.R.-H. contributed to reagents and data collection. C.V., A.P.-G., P.D.W., and A.R.-H. wrote the first draft of the manuscript with input from all authors. All authors read and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPORTING CITATIONS

The following reference appears in the Supplemental Information: Cortez et al. (2014).

REFERENCES

Alavattam, K.G., Maezawa, S., Sakashita, A., Khoury, H., Barski, A., Kaplan, N., and Namekawa, S.H. (2019). Attenuated chromatin compartmentalization in meiosis and its maturation in sperm development. Nat. Struct. Mol. Biol. *26*, 175–184.

Bannister, L.A., Reinholdt, L.G., Munroe, R.J., and Schimenti, J.C. (2004). Positional cloning and characterization of mouse mei8, a disrupted allelle of the meiotic cohesin Rec8. Genesis *40*, 184–194.

Barral, S., Morozumi, Y., Tanaka, H., Montellier, E., Govin, J., de Dieuleveult, M., Charbonnier, G., Couté, Y., Puthier, D., Buchou, T., et al. (2017). Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. Mol. Cell *66*, 89–101.e8.

Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R.D., and Petukhova, G.V. (2012). Genetic recombination is directed away from functional genomic elements in mice. Nature *485*, 642–645.

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner. (Department of Energy), Joint Genome Institute.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S.; the AmiGO Hub; the Web Presence Working Group (2008). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288–289.

Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P.D., Grützner, F., and Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. Nature *508*, 488–493.

da Cruz, I., Rodríguez-Casuriaga, R., Santiñaque, F.F., Farías, J., Curti, G., Capoano, C.A., Folle, G.A., Benavente, R., Sotelo-Silveira, J.R., and Geisinger, A. (2016). Transcriptome analysis of highly purified mouse spermatogenic cell

populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. BMC Genomics 17, 294–313.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat. Rev. Genet. 14, 390–403.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., He, J., Xiang, Y., Wang, Q., Li, Y., et al. (2017). Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature 547, 232–235.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. 7, 1728–1740.

Flyamer, I.M., Gassler, J., Imakaev, M., Brandão, H.B., Ulianov, S.V., Abdennur, N., Razin, S.V., Mirny, L.A., and Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature 544, 110–114.

Gel, B., and Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics 33, 3088–3090.

Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics 32, 289–291.

Gibcus, J.H., Samejima, K., Goloborodko, A., Samejima, I., Naumova, N., Nuebler, J., Kanemaki, M.T., Xie, L., Paulson, J.R., Earnshaw, W.C., et al. (2018). A pathway for mitotic chromosome formation. Science 359, 652.

Gutiérrez-Caballero, C., Herrán, Y., Sánchez-Martín, M., Suja, J.Á., Barbero, J.L., Llano, E., and Pendás, A.M. (2011). Identification and molecular characterization of the mammalian α-kleisin RAD21L. Cell Cycle 10, 1477–1487.

Haaf, T., and Ward, D.C. (1995). Higher order nuclear structure in mammalian sperm revealed by in situ hybridization and extended chromatin fibers. Exp. Cell Res. 219, 604–611.

Haering, C.H., and Jessberger, R. (2012). Cohesin in determining chromosome architecture. Exp. Cell Res. 318, 1386–1393.

Handel, M.A., and Schimenti, J.C. (2010). Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. Nat. Rev. Genet. 11, 124–136.

Henderson, K.A., and Keeney, S. (2005). Synaptonemal complex formation: where does it start? BioEssays 27, 995–998.

Herrán, Y., Gutiérrez-Caballero, C., Sánchez-Martín, M., Hernández, T., Viera, A., Barbero, J.L., de Álava, E., de Rooij, D.G., Suja, J.Á., Llano, E., and Pendás, A.M. (2011). The cohesin subunit RAD21L functions in meiotic synapsis and exhibits sexual dimorphism in fertility. EMBO J. 30, 3091–3105.

Jodar, M., Sendler, E., and Krawetz, S.A. (2016). The protein and transcript profiles of human semen. Cell Tissue Res. 363, 85–96.

Jung, Y.H., Sauria, M.E.G., Lyu, X., Cheema, M.S., Ausio, J., Taylor, J., and Corces, V.G. (2017). Chromatin states in mouse sperm correlate with embryonic and adult regulatory landscapes. Cell Rep. 18, 1366–1382.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell 88, 375–384.

Kleckner, N., Storlazzi, A., and Zickler, D. (2003). Coordinate variation in meiotic pachytene SC length and total crossover/chiasma frequency under conditions of constant DNA length. Trends Genet. 19, 623–628.

Lahn, B.T., and Page, D.C. (1999). Four evolutionary strata on the human X chromosome. Science 286, 964–967.

Lange, J., Yamada, S., Tischfield, S.E., Pan, J., Kim, S., Zhu, X., Socci, N.D., Jasin, M., and Keeney, S. (2016). The landscape of mouse meiotic double-strand break formation, processing, and repair. Cell 167, 695–708.e16.

Leng, N., Dawson, J.A., Thomson, J.A., Ruotti, V., Rissman, A.I., Smits, B.M.G., Haag, J.D., Gould, M.N., Stewart, R.M., and Kendziorski, C. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29, 1035–1043.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293.

Llano, E., Herrán, Y., García-Tuñón, I., Gutiérrez-Caballero, C., de Álava, E., Barbero, J.L., Schimenti, J., de Rooij, D.G., Sánchez-Martín, M., and Pendás, A.M. (2012). Meiotic cohesin complexes are essential for the formation of the axial element in mice. J. Cell Biol. 197, 877–885.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550–571.

Marco-Sola, S., Sammeth, M., Guigó, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. Nat. Methods 9, 1185–1188.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45, D183–D189.

Mueller, J.L., Mahadevaiah, S.K., Park, P.J., Warburton, P.E., Page, D.C., and Turner, J.M.A. (2008). The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. Nat. Genet. 40, 794–799.

Namekawa, S.H., Park, P.J., Zhang, L.-F., Shima, J.E., McCarrey, J.R., Griswold, M.D., and Lee, J.T. (2006). Postmeiotic sex chromatin in the male germline of mice. Curr. Biol. 16, 660–667.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science 342, 948–953.

Patel, L., Kang, R., Rosenberg, S.C., Qiu, Y., Raviram, R., Chee, S., Hu, R., Ren, B., Cole, F., and Corbett, K.D. (2019). Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. Nat. Struct. Mol. Biol. 26, 164–174.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44 (W1), W160–W165.

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature Communications 9, 189.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680.

Rau, A., Gallopin, M., Celeux, G., and Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. Bioinformatics 29, 2146–2152.

Reig-Viader, R., Garcia-Caldés, M., and Ruiz-Herrera, A. (2016). Telomere homeostasis in mammalian germ cells: a review. Chromosoma *125*, 337–351.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Ruiz-Herrera, A., Vozdova, M., Fernández, J., Sebestova, H., Capilla, L., Frohlich, J., Vara, C., Hernández-Marsal, A., Sipek, J., Robinson, T.J., and Rubes, J. (2017). Recombination correlates with synaptonemal complex length and chromatin loop size in bovids-insights into mammalian meiotic chromosomal organization. Chromosoma *126*, 615–631.

Sánchez-Guillén, R.A., Capilla, L., Reig-Viader, R., Martínez-Plana, M., Pardo-Camacho, C., Andrés-Nieto, M., Ventura, J., and Ruiz-Herrera, A. (2015). On the origin of Robertsonian fusions in nature: evidence of telomere shortening in wild house mice. J. Evol. Biol. *28*, 241–249.

Sassone-Corsi, P. (2002). Unique chromatin remodeling and transcriptional regulation in spermatogenesis. Science *296*, 2176–2178.

Serra, F., Baù, D., Goodstadt, M., Castillo, D., Filion, G.J., and Marti-Renom, M.A. (2017). Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput. Biol. *13*, e1005665.

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. Genome Res. *21*, 2213–2223.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. *45*, W122–W129.

Turner, J.M.A. (2007). Meiotic sex chromosome inactivation. Development *134*, 1823–1831.

Wang, S., Zickler, D., Kleckner, N., and Zhang, L. (2015). Meiotic crossover patterns: obligatory crossover, interference and homeostasis in a single process. Cell Cycle *14*, 305–314.

Wang, Y., Wang, H., Zhang, Y., Du, Z., Si, W., Fan, S., Qin, D., Wang, M., Duan, Y., Li, L., et al. (2019). Reprogramming of meiotic chromatin architecture during spermatogenesis. Mol. Cell *73*, 547–561.e6.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

Yang, T., Zhang, F., Yardımcı, G.G., Song, F., Hardison, R.C., Noble, W.S., Yue, F., and Li, Q. (2017). HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. Genome Res. *27*, 1939–1949.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383.

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237–247.

STAR★METHODS

KEY RESOURCES TABLE

Reagent or Resource	Source	Identifier
Antibodies		
Anti-mouse SYCP3	Abcam	Cat#ab97672; RRID:AB_10678841
Anti-Rabbit DMC1	Santa Cruz Biotechnologies	Cat#sc-22768; RRID:AB_2277191
Anti-mouse Cy5	Jackson Immunoresearch	Cat#115-175-146; RRID:AB_2338713
Anti-rabbit FITC	Jackson Immunoresearch	Cat#111-095-003; RRID:AB_2337972
Anti-CTCF	Millipore	Cat#07-729; RRID:AB_441965
Anti-REC8	Courtesy of A.M. Pendás	N/A
Anti-RAD21L	Courtesy of A.M. Pendás	N/A
Anti-HRP-PO	Bio-Rad	Cat#1706515; RRID:AB_11125142
Chemicals, Peptides and Recombinant Proteins		
Collagenase Type II	Life Technologies	#17101015
DNase I	Sigma-Aldrich	#DN25-10MG
Trypsin from bovine pancreas	Sigma-Aldrich	#T9935-100MG
Hoechst 33342	Life Technologies	#H3570-10ml
Complete Protease Inhibitor	Roche	#1187358001
MboI	New England Biolabs	#R0147M
Biotin-14-dATP	Life Technologies	#19524-016
DNA Polymerase I, large (Klenow) Fragment	New England Biolabs	#M0210M
T4 DNA Ligase	New England Biolabs	#M0202M
Dynabeads MyOne Streptavidin T1	Life Technologies	#65001
T4 Polunucleotides Kinase	New England Biolabs	#M0201L
T4 DNA Polymerase	New England Biolabs	#M0212M
Klenow Fragment 3′- > 5′ exo-	New England Biolabs	#M0212M
AMPure XP Beads	Beckman-Coulter	#A63880
2x Laemmli Buffer	Bio-Rad Laboratories	#1610737
Trans-Blot® Turbo Transfer Packs	Bio-Rad Laboratories	#1704158
Ponceau S solution	Sigma-Aldrich	#P7170-1L
Clarity™ Western ECL Substrate, 500 ml	Bio-Rad Laboratories	#1705061
OneDay ChIP kit	Diagenode	C01010080
Unblocked protein A	Diagenode	C03020002
NEBNext® End Repair Module	New England Biolabs	E6050S
NEBNext® End Repair Reaction Buffer	New England Biolabs	B6052S
NEBNEXT® dA-tailing Reaction Buffer	New England Biolabs	B6059S
Deposited Data		
Hi-C data	This paper	GEO:GSE132054
ChIP-seq data	This paper	GEO:GSE132054
RNA-seq data	This paper	GEO:GSE132054
SPO11-oligos hotspots	Lange et al., 2016	GEO:GSE84689
H3K4me3 data	Brick et al., 2012	GEO:GSE35498
Hi-C data from sperm	Jung et al., 2017	SRR3225862 and SRR3225863
Experimental Models: Organisms/Strains		
C57BL/6J	Charles River Laboratories	N/A

Continued

Reagent or Resource	Source	Identifier
Software and algorithms		
BBDuk (version 10/2015)	Bushnell, 2014	https://sourceforge.net/projects/bbmap/
TADbit (version 0.2.0.23)	Serra et al. (2017	https://github.com/3DGenomes/TADbit
GEM (version 1.7.1)	Marco-Sola et al., 2012	https://sourceforge.net/projects/gemlibrary/
HiCExplorer (version 1.8.1)	Ramírez et al., 2018	https://github.com/deeptools/HiCExplorer
HiCRep (version 1.4)	Yang et al., 2017	https://github.com/MonkeyLB/hicrep
BEDtools (version 2.17)	Quinlan and Hall, 2010	https://github.com/arq5x/bedtools2
SAM tools	Li et al., 2009	http://samtools.sourceforge.net/
Deep tools	Ramírez et al., 2016	https://github.com/deeptools/deepTools
MACS2	Feng et al., 2012	https://github.com/taoliu/MACS
IGV tools	Thorvaldsdóttir et al., 2013	https://igv.org/
ChIPseeker	Yu et al., 2015	https://github.com/GuangchuangYu/ChIPseeker
KaryoploteR	Gel and Serra, 2017	http://bioconductor.org/packages/release/bioc/html/karyoploteR.html
RegioneR	Gel et al., 2016	https://bioconductor.org/packages/release/bioc/html/regioneR.html
Panther	Mi et al., 2017	http://www.pantherdb.org/
AmiGo	Carbon et al., 2008	http://amigo.geneontology.org
AIR	This paper	https://transcriptomics.cloud

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the Lead Contact: Aurora Ruiz-Herrera (aurora.ruizherrera@uab.cat).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals
C57BL/6J (B6) male mice at 8-17 weeks of age were purchased from the Charles River Laboratories. Animal maintenance and experimental procedures were carried out according to the Ethics Committee on Animal and Human Experimentation (CEEAH) guidelines from Universitat Autònoma de Barcelona (UAB).

Cell lines
A primary fibroblast cell line derived form a male mouse (C57BL/6J strain) previously established in our lab (Sánchez-Guillén et al., 2015), and cultured in DMEM medium supplemented with 10% fetal bovine serum and 1% PenStrep at 37°C and 5% CO_2.

METHOD DETAILS

Fluorescence Activated Cell Sorting (FACS) of mouse male germ cells
Male mice were dissected by an inguinal incision to remove whole testis and cauda epididymis. Sperm cells were recovered from the epididymis by pressing the deferent tube gently. Both decapsulated testis and epididymis were mechanically disaggregated together with the ejaculate, incubated in GBSS with collagenase type II (Life Technologies) (0.5 mg/ml) and DNase I (Sigma Aldrich) (1 ng/μl) at 33°C for 15min with agitation. Fetal bovine serum (ThermoFisher Scientific) was added to the mix at a proportion of 5% in order to inactivate trypsin. The cell suspension was filtered through a 70 μm diameter cell strainer and subsequently centrifuged for 3 min at 1,800 xg. Cells were then dyed with 5 μg/ml of Hoechst 33342 (Life Technologies) for 30 min at 33°C with agitation, then held at 4°C until sorting. Germ cells were sorted using a BD InfluxTM (BD Biosciences) coupled with an ultraviolet laser (355 nm). Subsequently, three sorting strategies were followed in order to obtain a total of six highly enriched germ cell populations: (i) spermatogonia (2n, 2c), (ii) spermatocytes at leptonema/zygonema (L/Z) stage (2n, 4c), (iii) spermatocytes at pachynema/diplonema (P/D) stage (2n, 4c), (iv) secondary spermatocytes (1n, 2c), (v) round spermatids and (vi) sperm.

Strategy #1: Four main germ cell populations (spermatogonia, P/D, secondary spermatocytes and round spermatids) were first isolated by plotting Hoechst Blue (UV355-460/50) versus Hoechst red (UV355-670/30) emissions to discriminate cells by both their DNA content and their complexity (Figure S1A). The spermatogonia fraction (2n, 2c) represented a heterogeneous population of undifferentiated and differentiating spermatogonia.

Strategy #2: In order to discriminate primary spermatocytes in L/Z and P/D stages, an "in solution" immunofluorescence using DMC1 (1:300, Santa Cruz Biotechnologies) and SYCP3 (1:1000, Abcam) antibodies was performed before sorting. Briefly, DMC1 (1:300, Santa Cruz Biotechnologies) and SYCP3 (1:1000, Abcam) antibodies were added to the cell suspension that resulted from testis disaggregation. Cells showing DMC1(+)/SYCP3(+) staining corresponded to primary spermatocytes in L/Z stage, whereas cells at P/D stage appeared with DMC1(-)/SYCP3(+) staining (Figure S1B).

Strategy #3: A third approach was followed to obtain highly enriched populations of sperm and round spermatids as both cell types have the same DNA content and can only be discriminated by morphology. We followed a re-gating approach since cell morphology can be discriminated in the flow sorter by how cells project a side shadow (side scatter, SSC) and a forward shadow (forward scatter, FSC), a plot for each initial gate was drawn (SSC versus FSC plot). With this approach, and applying a restrictive re-gating approach, sperm and round spermatids were isolated with high enrichment (Figure S1C). This allowed us to exclude elongating spermatids from these samples. Likewise, isolation of sperm from cauda epididymis was conducted by re-gating of the c fraction, resulting in isolated sperm Figure S1D).

Irrespective of the sorting strategy, all cell populations were collected after sorting in 1x PBS, centrifuged for 5 min at 1,800 xg. The supernatant was discarded, and cell pellets were flash-frozen at $-80°C$ until use. Sorts were typically 3-6 hours to collect between 0.2×10^6 and 3.2×10^6 cells, depending on the germ cell population.

Cell enrichment of each flow-sorted population was evaluated by immunofluorescence using specific meiotic proteins and DAPI morphology (Figure S1). Spermatogonia were distinguished using an anti-CD90 (Thy-1+) (1:10) antibody and DAPI morphology. For primary spermatocytes, prophase-I stages (leptonema, zygonema, pachynema and diplonema) were identified based on SYCP3 (1:400) and γH2AX patterns (1:300). For secondary spermatocytes, immunofluorescence using anti-H3K9me3 (1:500), anti-actin (1:400) and anti-tubulin (1:1000) was performed. Cell enrichment of round spermatids and sperm fractions was determined based on nucleus morphology and DAPI pattern. Cells were fixed on slides and then mounted with DAPI diluted in Vectashield (Vector Laboratories). In all cases, slides were analyzed using a fluorescence microscopy (Axiophot, Zeiss) coupled with a ProGres ® CS10plus, Jenoptik camera. Representative images were captured with ACO XY (A. Coloma, Open Microscopy). Between 50 and 100 cells were counted for each flow-sorted population. Only sorted populations with enrichment above 80% were considered for subsequent experiments.

Fibroblast cell culture

The mouse primary fibroblast cell line was cultured to 100% confluence. A cell fraction was kept for quality controls (see below) and the rest were fixated accordingly to the crosslink step of the Hi-C method (see *in nuclei* Hi-C section) for adherent cells.

Quality controls consisted of generating a modal karyotype and flow cytometry cell cycle analysis. Chromosome spreads were obtained using standard protocols. Briefly, cells were arrested in metaphase by adding 80 μL of Colcemid (10 μg/ml) (Sigma) to 10ml of medium for 2h and then trypsinised. Cells were spun down at 600 xg for 7 minutes and resuspended in 5ml of hypotonic solution (0.075M KCl) for 20 minutes at 37°C. Chromosomes were then fixed by addition of fixative solution (3:1 methanol/acetic acid) and metaphase spreads were obtained by dropping 15 μl of cell suspension onto a cleaned dry slide. Slides were kept at $-20°C$ until use. Metaphases were stained homogenously with Giemsa solution for analysis of the modal karyotype. An optical microscope (model Zeiss Axioskop) equipped with a charged coupled device camera (ProGresR CS10Plus, Jenoptik) was used for the microscope analysis. Good-quality metaphases were captured with the program Progress Capture 2.7.7 and analyzed for each specimen, obtaining the modal karyotype.

For the cell cycle analysis, fibroblasts were cultured until cells reached 100% confluence. Next, cells were fixated with 1% formaldehyde for 10' at RT and incubated with glycine 0.125M for 5' at RT and for 15' at 4°C to stop the crosslinking reaction. Subsequently, cells were scrapped off the flask, pelleted and resuspended in 1x PBS to finally be stained with Hoechst 33342. Cell cycle was analyzed using a BD InfluxTM (BD Biosciences) coupled with an ultraviolet laser (355 nm) to reveal that over 74% of the cells were in G1-phase (Figure S1).

In nuclei Hi-C

The generation of Hi-C libraries was optimized from the protocol developed by Rao et al. (2014). Different approaches were used depending on the cell type analyzed (somatic and germ cells, see below). Two replicates for each FACS-sorted population were obtained (with the exception of primary spermatocytes at L/Z stage where one single experiment was performed) from a total of 7.7×10^6 spermatogonia, 1.7×10^6 primary spermatocytes at L/Z stage, 20.9×10^6 primary spermatocytes at P/D stage, 15.7×10^6 secondary spermatocytes, 107.7×10^6 round spermatids and 113.8×10^6 sperm.

Somatic cells: Fibroblasts were grown until reaching confluence in supplemented DMEM medium. Cells were then washed with PBS and fixed with 1% formaldehyde for 10' at RT. Cells were then incubated with glycine 0.125M for 5' at RT and for 15' at 4°C to stop the crosslinking reaction. Following, 2 mL Trypsin (Fischer Scientific) were added, and cells were incubated for 8' at RT and then washed twice with PBS. Cells were scraped and collected in a tube and centrifuged for 5 minutes at a maximum speed of 1,800 xg.

Germ cells: Germ cells at a concentration of 1 million per 500 μl were incubated in formaldehyde (1%) for 10 min with agitation prior to FACS. Glycine (Sigma Aldrich) was added at a final concentration of 0.125 M and incubated with agitation at room temperature for

5 min and then at 4°C for 15 min. Cells were then centrifuged for 10 min at 290 xg at 4°C and resuspended in 3 mL 1x PBS in case of direct Hoechst staining, or in the according volume of block solution if immunofluorescence was the following step.

Each crosslinked cell aliquot was resuspended with lysis buffer and incubated on ice for 30' and then centrifuged for 5' at 1,800 xg. Pellets were washed with 1x NEB2 buffer (twice) and resuspended with NEB2 buffer with 10% SDS at RT and incubated for 10' at 65°C with agitation (300 rpm). NEB2 buffer with 10% Triton X-100 solution was added and cells were incubated for 30' at 37°C. Following, cells were centrifuged for 5' at 1,800 xg (4°C) and washed with 1x NEB2 buffer twice. An aliquot for ND (Non-Digested) control was taken from the sample to be processed and incubated at 37°C together with the digested sample. 400 U of MboI were added to the rest of the samples and chromatin was digested O/N at 37°C with agitation. An aliquot from the digested samples was taken for digestion controls. The full samples were kept at 37°C while the digestion control was performed. Proteinase K (10 mg/ml) (ThermoFisher) was added and the aliquoted samples were incubated for 45'-60' at 65°C followed by Phenol:Chloroform purification. The quality of the sample was checked by running on an 0.8% agarose gel.

After digestion, a small aliquot from the samples was kept as a Non-Ligated control and another aliquot was directly ligated without reparation (ligation control). Samples were centrifuged for 5' at 1,800 xg. Samples were then washed with 1x NEB2 buffer, twice. After the second wash, samples were directly resuspended with the reparation mix (1x NEB2 buffer, 0.05 mM dCTP, 0.05 mM dTTP, 0.05 mM, 50 mM biotin-dATP (ThermoFisher), 50U Kleenow (NEB, M02010M), H_2O). Samples were incubated for 45' at 37°C and for 10' at 65°C and centrifuged for 5' at 1,800 xg and then resuspended with ligation buffer [1x NEB T4 ligase buffer, 10% Triton x-100, 0.1 mg/ml BSA, 5 µl (10000 U cohesive) ligase (2000 U/ µl) (NEB, M0201M), 963 µl H_2O]. Samples were incubated at 16°C for at least 4h or O/N with mixing, then centrifuged for 5' at 1,800 xg and resuspended in 1x NEB2 buffer. Samples were incubated with RNaseA (10 mg/ml) (ThermoFisher) for 15 minutes at 37°C. The mix was incubated with Proteinase K (10 mg/ml) at 65°C O/N to reverse the cross-link. Samples were cooled to RT and purified with Phenol/Chloroform/Isoamyl alcohol. DNA content was measured on a Qubit. Samples were sonicated: 20 s time ON, 60 s time OFF, 8 cycles. Samples were then loaded in an electrophoresis gel of 1.2% agarose to check fragment size.

Samples were incubated for 30' with rotation at RT with Dynabeads MyOne Streptavidin T1 beads (Life Technologies) and 2x Binding Buffer (10 mM TrisHCl, 1mM EDTA, 2M NaCl). Beads were washed twice with Binding Buffer and resuspended in the end repair mix [1x NEB T4 DNA ligase buffer with 10 mM ATP, 25 mM dNTP mix, 10U/µl NEB T4 PNK (NEB M0201), 3U/µl T4 DNA polymerase I (NEB M0203), 5U/µl NEB DNA polymerase I (Klenow) (NEB 0210)]. Samples were incubated for 30' at RT. Beads were washed with 1x Binding Buffer, twice. Beads were subsequently resuspended in the dATP attachment master mix (1x NEBuffer 2, 0.5mM dATP, 5U/µl NEB Klenow exo minus (NEB, M0212), H_2O). Samples were incubated at 37°C for 30'. The beads were washed with 1x Binding Buffer, twice and resuspended with 1x NEB Quick ligation buffer (NEB, B6058).

NEB T4 DNA ligase and Illumina indexed adapters were added and thoroughly mixed. Samples were incubated for 15' at RT, beads were captured, and the supernatant was discarded. Beads were then washed with 1x Binding Buffer, twice. The beads were resuspended in Tris buffer (Elution buffer). PCR (5µl beads + 5.5 µl H_2O, 1.25 µl primer P5 25 mM, 1.25 µl primer P7 25 mM, 12,5 cocktail master mix NEB) was performed (98°C 30 s, 98°C 10 s, 60°C 30 s x 8 cycles, 72°C 30 s, 72°C 5'). After PCR, beads were captured on a magnet and the PCR was transferred to a new tube. The library was quantified using a Qubit fluorometer. 5µl from the PCR products were run in a 1.2% agarose electrophoresis gel to confirm range of sizes. A 1:1 amount of AmpureBeads (Beckman Coulter) was added to the samples and incubated for 10' at RT. The beads were washed twice with 70% ethanol without mixing. The beads were then eluded and incubated at RT, for 5'. AmpureBeads were separated on a magnet, and the solution was transferred to a fresh tube. 5µl of each sample was loaded to an electrophoresis gel of 1.2% agarose for a final size check. DNA quantity was measured with the Qubit fluorometer. Libraries were submitted for Illumina sequencing (paired-end 75bp each side on HiSeq 2500, v4).

Hi-C data processing, binning and normalization

Quality check and trimming were conducted using BBDuk (version 10/2015) (Bushnell, 2014). Setting a minimum read length of 35 bp and a minimum Phred quality score of 20, adapters and low-quality reads were removed while preserving their longest high-quality regions. Then, reads were processed with TADbit (version 0.2.0.23) (Serra et al., 2017), which makes use of the GEM (version 1.7.1) mapper (Marco-Sola et al., 2012) to iteratively map them against the mouse genome (version mm10). Reads were mapped from 15 bp toward using a step size of 5 bp. The filters used in order to remove possible artifacts were the following: "self-circle," "dangling-end," "error," "extra dangling-end," "too short," "too large," "duplicated," and "random breaks." The maximum molecule length parameter was set at 2 times the 99.9 percentile of the insert size distribution, returned by the "insert_size" from TADbit. The maximum distance of a read to a cleavage site was set to the 99.9 percentile of the insert size distribution.

An in-house script was used for binning and data normalization. This script imported the "HiC_data" module of TADbit, read the map files generated after the artifacts filtering step, binned the reads into a square matrix of 50 Kbp, and stored the matrix into a file in NPZ format (raw matrix). Afterward, HiCExplorer (version 1.8.1) (Ramírez et al., 2018) was used to normalize with the ICE (Iterative Correction and Eigenvector decomposition).

Correlation coefficient analysis

Pairwise comparisons between biological replicates were performed using HiCRep (version 1.4) (Yang et al., 2017), under a smoothing parameter of 5 and a considered distance over 10 Mbp. Since HiCRep only handles intra-chromosome raw matrices, each pairwise comparison yielded 20 correlation scores (19 autosome chromosomes + sex chromosome X). The Y chromosome was

excluded from the analysis due to the lower number of interactions detected in our analysis (less than 1% of the overall detected interactions) and the highly repetitive nature of this chromosome. The correlation between 2 replicates was defined as the mean of the 20 correlation scores.

Averaged contact probability P(s)
ICE-normalized matrices were scaled with a factor of 1/sum(matrix). The resulting matrices were then input to ''hicPlotDistVsCounts'' from the HiCExplorer package in order to obtain the contact probability P(s).

Inter-/intra-chromosome interaction ratio
ICE-normalized data stored in matrices were exported with HiCExplorer to the GInteractions format, which consists of 7 columns: chromosome, start and end from bin 1, chromosome, start and end from bin 2, and the amount of interaction. The GInteractions tables were imported in R for further quantification of intra-chromosome and intra-chromosome interactions and plotting.

Inter-subtelomeric interaction quantification
ICE-normalized matrices were scaled with a factor of 1,000,000/sum(matrix) and exported with HiCExplorer to GInteractions format. The GInteractions tables were imported in R for this inter-telomere interaction quantification. Since the telomeric and centromeric regions (annotated from the beginning of each chromosome to 2.9 Mbp according to the UCSC Table Browser) were masked due to the low-count filtering step prior to ICE normalization, we only considered inter-chromosome interactions between loci located within genomic positions 3 to 3.5 Mbp in each chromosome. Differences in the subtelomeric interaction frequencies between cell types were assessed with the Wilcoxon test.

Sperm simulations
In order to validate our enriched sperm population, we simulated six Hi-C sperm datasets of 100×10^6 reads with different proportions, from 0 to 100% by steps of 20%, of fibroblast reads. Both sperm and fibroblasts reads were derived from our generated libraries. Previously published data on sperm (SRR3225862 and SRR3225863 accessions from Jung et al. (2017) were also downloaded. These datasets underwent a quality check, Hi-C data processing, binning and normalization steps as described above. The resulting raw Hi-C matrices were used for correlation coefficient analysis while the ICE-normalized matrices were used to calculate the averaged contact probability P(s) (Figure S5).

A-B compartments and TAD calling
Raw matrices were used for the definition of A-B compartments. Columns with a low number of counts were filtered out using TADbit, setting the parameter min_count to 10. Since TADbit fits the column count distribution into a polynomial distribution, columns with a number of counts smaller than the first antimode of the distribution, which cannot be smaller than the min_count parameter, are filtered out. Then, the genome-wide matrices were normalized by the expected interactions at a given distance and by visibility by means of one-iteration of the ICE method. The correlation analysis was also performed with TADbit. In-house scripts computed A-B compartments from the first eigenvector, using 0 as threshold to differentiate both compartments and the gene density to label them.

TADs were identified using an in-house script that imported the ''Chromosome'' module of TADbit and added the raw and the ICE-normalized matrices of each chromosome separately. Filtered bins, due to low counts, were included in order to mask them when calling TADs. TAD insulation scores were obtained by first normalizing the different matrices for read depth in order for the scores to be comparable. Each matrix was then scaled to have 100M reads. Afterward, TAD insulation scores were obtained from the output given by the ''hicFindTADs'' program from HiCExplorer.

Compartment switching
BED files with a resolution of 50 Kbp were available from the compartments definition step. Each genomic bin of 50 Kbp had its corresponding compartment attributed. Pairwise comparisons between cell types -genome-wide and per-chromosome- were performed; the ratio of compartment switching was calculated as the number of genomic bins with a compartment change (A > B or B > A) divided by the total number of bins. From these files, a matrix file was created with 50 kbp-binned genomic coordinates as rows and cell types as columns, filled by the corresponding compartment labeling in each bin and cell type. Cell-specific A compartments were defined as those bins being compartment A in a cell type and compartment B in the remaining cell types.

3D-FISH and confocal microscopy
A total of 10 commercially available Bacterial Artificial Chromosomes (BACs) (Source BioScience) from mouse chromosomes 12 and 14 were selected according to their genomic location in A or B compartments (5 BACs for each compartment) (Table S4). The selected probes were separated by a range of genomic distances ranging from 66 kbp and 1.45 Mbp. Probes were selected based on their mappability and repetitive content in the mm10 genome assembly using the UCSC browser (https://genome.ucsc.edu/).

Probes were labeled with either dUTP-Dig (Sigma Aldrich) or dUTP-Cy3 (Enzo LifeSciences) by Nick Translation (Abbot Molecular). The 3D fluorescence *in situ* hybridization (3D-FISH) protocol was performed on mouse fibroblasts as following. Briefly, fibroblasts were cultured on slides overnight, whereas the germ cell suspension obtained after testis disaggregation was placed on slides

and incubated 1h at 37°C for cells to adhere on the surface. Slides were then fixed with 4% paraformaldehyde, washed in 1X PBS and then incubated in 60% Glycerol for 30-60 min, then finally snap-frozen in liquid nitrogen and kept in 50% formamide/2x SSC. As required, slides were thawed, treated with pepsin (0.01N HCl with 0.005% pepsin), washed with saline solutions (2x SSC, 50 mM MgCl$_2$, 1x PBS) and post-fixed with 1% paraformaldehyde. After 1h incubation in 50% formamide/2x SSC, the hybridization solution was placed on the slides. Both slides and probes were simultaneously denatured at 75°C for 2 min and then incubated for 48h at 37°C. Slides were washed in 2xSCC at 37°C and in 0.1x SSC at 60°C and dUTP-Dig was detected followed incubation with anti-Dig FITC (1:150) for 45 min at 37°C. Finally, slides were washed with 4x SSC with 0.2% Tween-20 and mounted with DAPI diluted in Vectashield. Germ cells from four males were subject to the protocol simultaneously.

Physical distances between hybridization signals were evaluated using a confocal microscopy (Leica SP5) using a 63x objective lens. Stacks of 0.34 µm-wide slices were captured for each cell nucleus, with a mean number of 20 stacks per nuclei. Images were reconstructed and analyzed using IMARIS (IMARIS Image Analysis Software), establishing nucleus volume (µm^3), physical distances between signals (µm) and relative physical distance between signals and nucleus surface (µm). For the data analysis, pairwise distances were computed for particular nuclei.

In total, we imaged 10 different loci (5 located in A compartments and 5 loci located in B compartment) across chromosomes 12 and 14. For each pair of probes, we measured the correlation between physical distances (in µm) and genomic distance (kbp) in pre-meiotic (fibroblasts and spermatogonia) and post-meiotic cells (round spermatids) (Figure S4; Table S4). A total of 527 measurements between seven pairs of probes were performed (Figure S6). We observed a positive correlation between physical and genomic distance (Spearman p value < 0.05) (Figure S4). But more importantly, we detected a differential pattern for both types of compartments (A and B) in round spermatids when compared to spermatogonia and fibroblasts (Figure S4). Lineal regressions were almost identical in spermatogonia and fibroblasts. In round spermatids, however, physical distances between pairs of probes were greater, confirming different chromatin organization in late spermatogenesis.

Western blot analysis of CTCF and cohesins in germ cells

Proteins from adult mice testis and fibroblast were extracted with RIPA buffer (NaCl 150 mM, Triton X-100 1%, Sodium deoxycholate 0.5%, SDS 0.1% and Tris-HCl 50 mM pH = 8.0). After protein quantification with the PierceTM BCA Protein Assay Kit (ThermoFisher Scientific), 6 replicates of both fibroblast and testis extract (three replicates of 30 µg and three of 40 µg protein) were denaturalized with 2x Laemmli and loaded into an 8% polyacrylamide gel until proteins reached the end of the gel [determined by the Precision Plus ProteinTM Dual Color Standards (Bio-Rad)]. Proteins were then transferred to a nitrocellulose membrane using the Trans-Blot® TurboTM Transfer System from Bio-Rad (10 minutes of transfer time). Next, membranes were stained with the Ponceau S solution (Sigma-Aldrich) to ensure that proteins had correctly transferred to the membrane. Then, the membranes were washed and incubated with blocking solution (TBS 1x with 0.1% Tween-20, 5% fat-free milk and 1x PBS) for at least an hour. Each set of testis and fibroblast protein extracts were incubated overnight at 4°C with anti-rabbit CTCF (1:2500), anti-rabbit RAD21L (1:2000) and anti-rabbit REC8 (1:2000) respectively, and all were simultaneously incubated with anti-rabbit ßTubulin (1:5000) as a control. Antibodies were detected the following day with 1-hour incubation with anti-rabbit HRP-PO (1:15000). Finally, membranes were detected with ClarityTM Western ECL Substrate and the results captured with the Molecular Imager VersadocTM (Bio-Rad).

Spermatocyte spreads and immunofluorescence analysis of CTCF and cohesins

Spermatocyte spreads were obtained from frozen mouse testis. Testes were mechanically disaggregated until obtaining a cell suspension in 1x PBS. The cell suspension was then distributed into different slides and incubated with 1% Lypsol for 16 minutes followed by a 20-minute incubation with 4% paraformaldehyde. Then slides were left to dry and washed with twice PhotoFlo 1% (Kodak) and then blocked with PBS-Tween-20 (0.005%). Slides were incubated overnight at 4°C with the following antibodies: anti-rabbit CTCF (1:50), anti-rabbit RAD21L (1:20) and anti-rabbit REC8 (1:20); one per slide and all of them combined with anti-mouse SYCP3 (1:400). Primary antibodies were detected with anti-rabbit Cy3 (1:200) combined with anti-mouse FITC (1:200). Slides were finally mounted with DAPI and analyzed with a fluorescence microscopy (Axiophot, Zeiss) coupled with a ProgRes ® CS10plus, Jenoptik camera. Representative images were captured with ACO XY (A. Coloma, Open Microscopy).

Stimulated emission depletion (STED) microscopy

Fab fragments were used for blocking and double labeling of REC8 and RAD21L antibodies (both raised in rabbit). STED microscopy (SP8, Leica) was used to generate the super-resolution images of REC8 and RAD21L foci along the chromosome axes. Secondary antibodies for STED imaging were conjugated to Alexa 555 and 488 (Invitrogen). Slides were mounted in Prolong Antifade Gold without DAPI. Fluorescence signals (red to green ratio) were measured along the 19 autosomal and XY axial elements of pachytene cells using the LAS X software from Leyca. Signal intensities were standardized and the overlay profiles of RAD21L and REC8 were plotted. Regression analysis was performed to determine the correlation between their profiles. The values of the coefficients of determination R^2 are shown in the scatterplots.

ChIP-sequencing

For chromatin immunoprecipitation, antibodies for CTCF (10 µl per sample) and cohesins RAD21L and REC8 (30 µl per sample) were used. Two biological replicates of ChIP-sequencing (ChIP-seq) were performed using FACS-sorted primary spermatocytes at P/D

stage (14 million of cells) and round spermatids (20 million of cells) from adult mice. In short, cells were incubated on ice in lysis buffer I (5 mM PIPES, 85 mM KCl, 0.5% NP-40, Protease Inhibitors) and lysis buffer II (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, Protease Inhibitors). Samples were then sonicated with the Biorruptor pico (30 s ON, 30 s OFF, 10 cycles) to obtain fragments around 200 bp. Sonicated lysates were centrifuged, and the supernatant was diluted in cold IP buffer (Diagenode). A small aliquot of each sample was kept as input and the remaining was divided in three aliquots, where antibodies for CTCF, RAD21L and REC8 were added (one in each aliquot) and then incubated overnight at 4°C. The antibody-chromatin pull-down was performed with Unblocked Protein A beads (Diagenode) and beads were then eluded with elution buffer (1% SDS, 0.1M NaHCO$_3$). After elution samples were centrifuged, and the supernatant was incubated with 200 mM NaCl at 65°C overnight to reverse the crosslink. Finally, proteins were digested and the DNA purified with phenol:chloroform:isoamyl alcohol. Libraries were prepared for each sample, repairing fragment ends by incubating the samples for 30 minutes with the NEBNext end repair mix (New England Biolabs). Then samples were purified with AMPure beads. Samples were then incubated with the A-tailing NEBNext mix (New England Biolabs) and purified again with AMPure beads. Then, adaptors were ligated to the sample, which was subsequently purified with AMPure beads (0.8X). Libraries were PCR-enriched using NEBNext indexed primers, 12 PCR cycles for RAD21L and CTCF samples and 14 cycles for REC8 samples. Libraries were finally purified with AMpure beads (0.7X).

ChIP-seq peak calling and annotation

Quality check and trimming were performed using BBDuk. Setting a minimum read length of 35 bp and a minimum Phred quality score of 20, adapters and low-quality reads were removed while preserving their longest high-quality regions. Single end reads were obtained for CTCF ChIP and paired end reads for the cohesin ChIP. Reads were trimmed and mapped to the mouse reference genome (mm10) using a Galaxy server. After mapping, non-unique mapped reads with a mapping quality less than 30 were filtered out with SAMtools (Li et al., 2009). Then, to assess read coverage distribution across the genome, bigWig files for each sample were generated with DeepTools (Ramírez et al., 2016). Following the server recommendations, filtered files were merged with Picard (Broad Institute) and split again with SAMtools before peak calling. Peak calling was performed using MACS2 (Feng et al., 2012). To find the optimal parameters for running the peak call function, a cross-correlation analysis between reads mapping to plus and minus strands was performed, and the d parameter was estimated for each sample. Peaks were called using as extension size the d estimated for each sample respectively. Coverage, reads and peaks were visualized with IGVtools (Thorvaldsdóttir et al., 2013).

BED peak files were imported to Rstudio and annotated to the reference mouse genome (mm10) using the R/bioconductor package ChIPseeker (Yu et al., 2015) together with the knownGene table from the UCSC resources (TxDb.Mmusculus.UCSC.mm10.knownGene). The number of peaks overlapping between CTCF and cohesins and between cell types was obtained by using the bioconductor package ChIPpeakAnno (Zhu et al., 2010). ChIP-seq data was represented using the packages ggplot2 (Wickham, 2016) and karyploteR (Gel and Serra, 2017) in combination with ChIPseeker and ChIPpeakAnno.

RNA-sequencing

Full-length single-cell RNA sequencing libraries were prepared using the Smart-seq2 protocol (Picelli et al., 2014) with minor modifications. Pools containing between 20,000 and 40,000 cells were obtained by FACS-sorting for four cell types (spermatogonia, primary spermatocytes at P/D, round spermatids and sperm) from adult mice. Four independent biological replicated were included in the analysis. Briefly, cells were sorted into 1.5ml eppendorfs containing lysis buffer. Reverse transcription was performed using SuperScrpit II (Invitrogen) in the presence of oligo-dT30VN, template-switching oligonucleotides and betaine. The cDNA was amplified using the KAPA Hifi Hotstart ReadyMix (Kappa Biosystems), ISPCR primer and 16 cycles of amplification. Following purification with Agencourt Ampure XP beads (Beckmann Coulter), product size distribution and quantity were assessed on a Bioanalyzer using a High Sensitvity DNA Kit (Agilent Technologies). Amplified cDNA was fragmented using Nextera® XT (Illumina) and amplified with indexed Nextera® PCR primers. Products were purified twice with Agencourt Ampure XP beads and quantified again using a Bioanalyzer High Sensitivity DNA Kit. Sequencing of Nextera® libraries was carried out on a HSeq2500 (Illumina) to obtain > 30 million pair ends reads per sample.

RNA profiling

The Artificial Intelligence RNA-seq Software as a Service (SaaS) platform (https://transcriptomics.cloud) was used to analyze RNA-seq data. AIR accepts raw next generation sequencing Illumina FastQ data as input. RNA-seq data was uploaded to the site and validated in order to automatically pair forward and reverse files (in case of paired-end samples) as well as to check its format and integrity. Quality was assessed using FastQC. A new analysis was defined from the "new analysis" screen, where the samples to be included were selected for analysis, along the reference genome (> 120,000 available genomes from NCBI, Ensembl and JGI). The analysis included quality trimming, Differential Gene Expression (DGE) followed by a Gene Ontology Enrichment Analysis (GOEA). Once the analysis was launched, bad quality reads were removed using BBDuk by setting a minimum length of 35 bp and a minimum Phred-quality score of 25. Afterward, high quality reads were mapped against the reference genome with STAR (Dobin et al., 2013) using the end-to-end alignment mode and gene expression quantification was performed with featureCounts (Liao et al., 2014).

The statistical analysis started by filtering lowly expressed genes using HTSFilter (Rau et al., 2013). Four statistical methods were then used for the identification of differentially expressed genes: DESeq2 (Love et al., 2014), edgeR (Robinson et al., 2010), EBSeq (Leng et al., 2013) and NOISeq (Tarazona et al., 2011). Data normalization was performed with the Trimmed Mean of M-values (TMM) method. Finally, GOEA was performed with in-house scripts based on hypergeometric tests (Tian et al., 2017). Multiple testing corrections controlling false positives from high-throughput experiments were also performed with the Benjamini-Hochberg method. The statistics section included: the Principal Component Analysis (PCA) clusterization of the samples, general plots for the interpretation of the experiment and several tabs in which the Differential Expressed Genes (DEG) and GOEA data is shown in different tables.

Specifically, we selected the mouse genome (GRCm38) from the Ensembl release 89. NOISeq was used to explore the DEG due to the variability of biological replicates. Raw expression and FPKM (Fragments Per Kilobase Million) values were downloaded from AIR. The Principal Component Analysis (PCA) was performed using the noise correction function from the NOISeq package. The raw expression values were converted to Counts Per Million (CPM) using edgeR and, afterward, expression values for each cell type were averaged for further analyses.

Analysis of correspondence between compartments and gene expression

Cell-specific A-B compartments were intersected with BEDTools (version 2.26) against a BED file with the TSS of genes derived from the GRCm38 gene annotation from Ensembl (release 89). Genes in each compartment were grepped (Bash command) with the table of FPKM values downloaded from AIR (see above), producing the expression profiles represented as boxplots for each cell type and compartment. Statistical significance among pairwise comparisons was tested using the Wilcoxon test.

For GOEA in cell-specific A-compartments we extracted gene Ontology (GO) terms from expressed genes (CPM > 1) located in A compartments. GOEA was performed following the approach used by AIR on the DEG. It consists of a series of hypergeometric tests carried out for each GO term thus identifying significant enriched GO terms relative to the expected genome background. P values were corrected with the Benjamini-Hochberg procedure to reduce false positives. We considered as significantly enriched the terms with corrected P values $\leq$ 0.01. Results were displayed using bubble plots.

Moreover, expressed genes with peaks on their promoters were classified according to cell type and protein (CTCF, RAD21L and REC8). Genes were then analyzed with PANTHER (Mi et al., 2017) using the online tool AmiGO (Carbon et al., 2008).

Local insulation, cohesin occupancy and gene expression

The GRCm38 gene annotation from Ensembl (release 89) was downloaded in GTF format. It was parsed with an in-house script to extract in BED format the promoter regions 2 kbp upstream the Transcriptional Start Site (TSS) for each gene. Afterward, ChIP-seq peaks were intersected with BEDTools to obtain gene lists with ChIP-seq peaks in their promoters. These gene lists were grepped (Bash command) with the downloaded table with the expression values from the AIR platform (see below).

Protein peaks were associated with changes in the corresponding TAD insulation score or gene expression using the function meanInRegions of the R package regioneR (version 1.10) (Gel et al., 2016). This package generates an expected distribution from 10.000 permutations from the values (e.g., TAD insulation score or FPKM values) observed in random genomic locations, thus calculating the Z-score of the TAD insulation score or the gene expression observed on ChIP peaks. Z-scores from flanking regions (+/− 250 Kbp) of peaks were calculated with the function localZScore.

Meta-border plots were created using ICE-normalized matrices normalsed by the number interactions expected at a given distance with "hicFindEnrichedContacts" program from the HiCExplorer package (parameters: "–method obs/exp," "–perchr"). Subsequently, sub-matrices of interaction counts 5 Mbp up and downstream of specific regions (e.g., TAD boundaries or protein peaks) were generated and averaged among them. The interaction counts underwent a log10 transformation and were plot with "hicPlotTADs" program from HiCExplorer.

X chromosome evolutionary strata and ampliconic regions

Mouse X strata were extrapolated from the human X chromosome using synteny information (https://www.ensembl.org/Mus_musculus/Location/Synteny?r=X). Likewise, the boundaries of each amplicon in the current assembly were defined by flanking paralogs (Table S7).

QUANTIFICATION AND STATISTICAL ANALYSIS

All sequencing data was checked and trimmed with BBDuk (Bushnell, 2014). Hi-C data was processed with TADbit (version 0.2.0.23) to obtain raw interaction matrices, compartments and TADs. HiCExplorer was used to: (i) normalize raw interaction matrices, (ii) create interaction heatmaps, (iii) predict averaged contact probabilities P(s) and (iv) obtain TAD insulation scores. Correlation values between biological replicates were performed using HiCRep (Yang et al., 2017). Normalized interaction matrices were exported with HiCExplorer to the GInteractions format and imported in R to calculate the inter-intra-chromosome interaction ratio. Normalized interaction matrices were scaled with a factor of 1,000,000/sum(matrix) with a custom Python script, exported with HiCExplorer to GInteractions format and imported in R to create boxplots. Differences between cell types were assessed with the Wilcoxon test (p value < 0.05).

ChIP-seq data was processed using the Galaxy server. BigWig files for each sample were generated with DeepTools (Ramírez et al., 2016) and peaks were called using MACS2 (Feng et al., 2012). BED peak files were imported to R and annotated using the R/Bioconductor package ChIPseeker (Yu et al., 2015) and UCSC resources (TxDb.Mmusculus.UCSC.mm10.knownGene). ChIPpeakAnno (Zhu et al., 2010) was used for peak intersection. Changes in the insulation score on peaks were statistically assessed using the function meanInRegions of the R package regioneR (version 1.10) (Gel et al., 2016).

RNA-seq data was processed with the Artificial Intelligence RNA-seq (AIR) Software as a Service (SaaS) platform (https://transcriptomics.cloud). In this case, raw expression values were converted to Counts Per Million (CPM) using edgeR, and averaged for each cell type. FPKM values were averaged for each cell type and BEDTools (version 2.26) was used to intersect genomic coordinates to relate gene expression with compartments, insulation scores and ChIP-seq marks. Differences in gene expression in the intersected regions between cell types were assessed by means of Wilcoxon tests (p value < 0.05).

DATA AND CODE AVAILABILITY

The accession number for the Hi-C, RNA-seq and ChIP-seq data reported in this paper is GEO: GSE132054.

Supplementary Information

Three-dimensional genomic structure and cohesin occupancy correlate with transcriptional activity during spermatogenesis

Supplementary Figure 1: FACS strategies for isolation of highly enriched germ cell populations, related to Figure 1. (A) Representative immunofluorescence images for the flow-sorted germ cells populations. (i) Pre-meiotic cells include spermatogonia positive for CD90.2 (green). (ii) Meiosis I includes spermatocytes I at leptonema/zygonema and pachynema/diplonema stages. Synaptonemal complexes (SYCP3) are labelled in green, ɣH2AX in red and DNA in blue (DAPI). (iii) Meiotic II cells represent a heterogeneous population including spermatocytes II in G_1, metaphase and diakinesis. Tubulin is labelled in green and DNA in blue (DAPI). (iv) The spermiogenesis stage includes round spermatids and sperm. Legend - n, chromosome number per cell; c, number of chromatids per chromosome. Scale bar: 10 μm. (B) Highly enriched sub-populations of primary spermatocytes. (i) Flow cytometry Hoechst Blue (UV355-460/50) and Hoechst Red (UV355-670/30) plot after "in solution" immunofluorescence with DMC1 and SYCP3. Germ cells are differentiated by distinct DNA content: 4c, 2c and c. (ii) Re-gating of the 4c population displaying three different cell populations: +DMC1/+SYCP3, -DMC1/+SYCP3 and -DMC1/-SYCP3. (iii) Enrichment values for +DMC1/+SYCP3 and -DMC1/+SYCP3 sorted populations. (C) Isolation of highly enriched round spermatids and sperm through re-gating. (i) Main FACS profile from whole testis, where primary gating is established for each germline population. (ii) Isolation of round spermatids and sperm, where cells are re-gated based on how cells project a side shadow (side scatter, SSC) and a forward shadow (forward scatter, FSC), attaining highly pure fractions of these populations. (iii) Representative images and enrichment values for round spermatids (RS) and sperm fractions after re-gating. Legend: RS – round spermatids, ES – elongated spermatids. (D) Isolation of sperm from caudate epididymis. (i) Representative cells obtained from the caudate epididymis stained with DAPI prior to FACS. (ii) Flow cytometry Hoechst Blue (UV355-460/50) and Hoechst Red (UV355-670/30) plot showing isolated sperm from caudate epididymis. Two populations of cells are evidenced (2c and c). Re-gating of the c fraction resulted in isolated sperm. (iii) Representative image and enrichment values after sorting. Legend: ES – elongated spermatids. (E) FACS profile of mouse primary fibroblast cell culture.
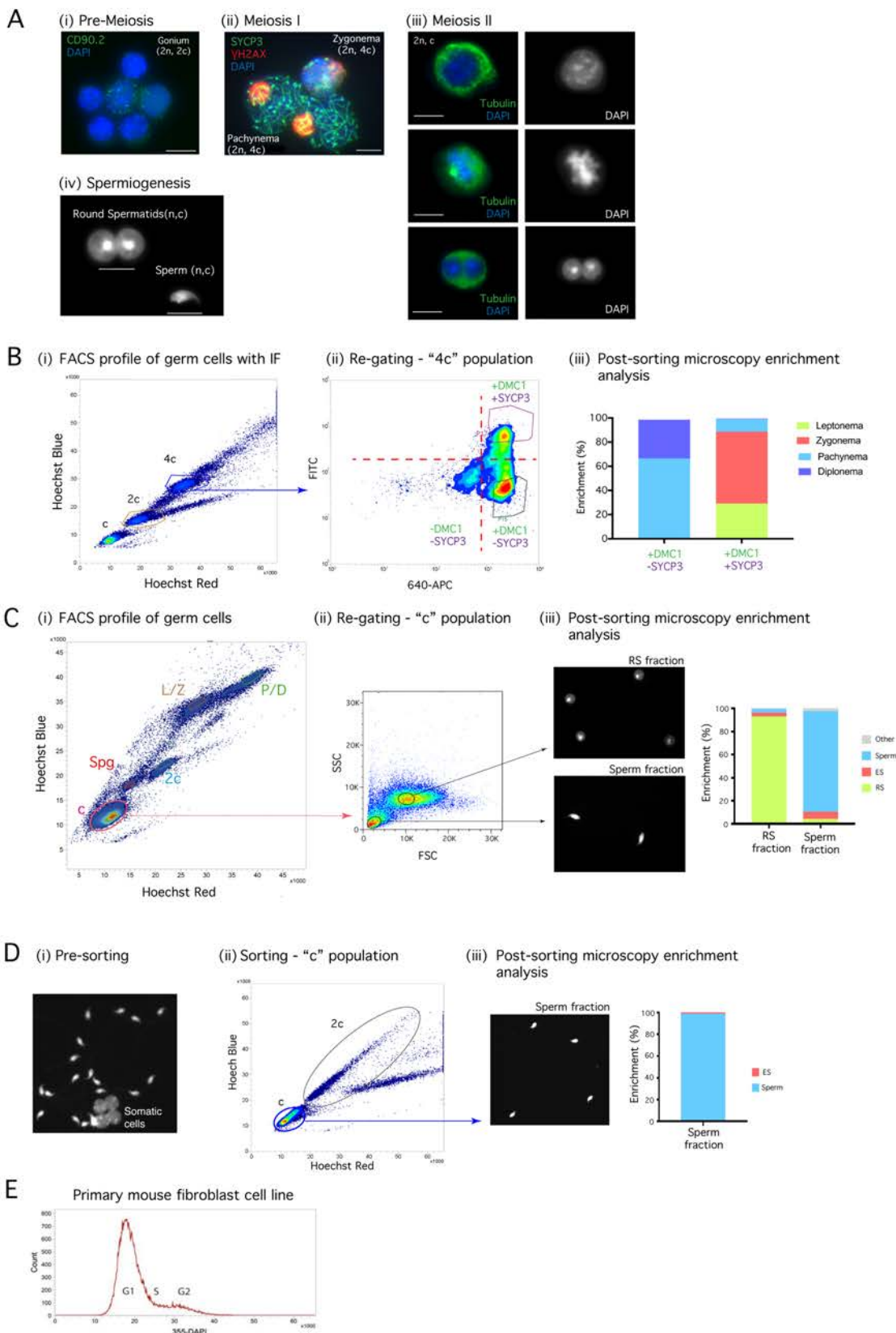
Supplementary Figure 2: Correlation among Hi-C replicates, related to Figure 2. Heatmap showing the correlation values among replicates based on the pairwise similarity score calculated using HiCRep. Legend – Fib: Fibroblast, Sg: Spermatogonia, L/Z: Leptonema/Zygonema, P/D: Pachynema/Diplonema, SpII: Spermatocytes II, RS: Round Spermatids, Rep#: replicates.
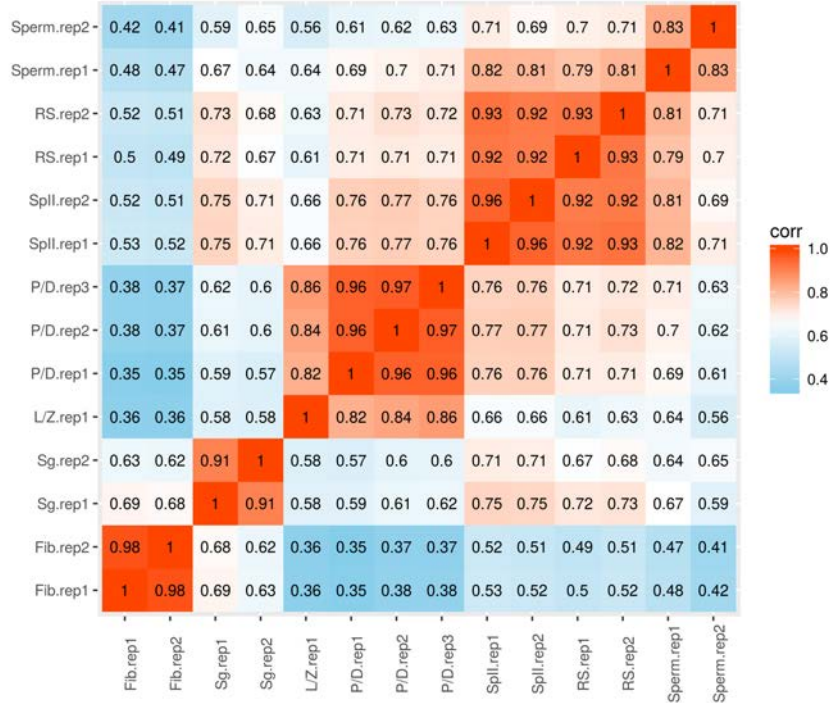
Supplementary Figure 3: Compartment switching during spermatogenesis, related to Figure 3. (A) Representation of the genomic regions corresponding to A (red) and B (blue) compartments in mouse chromosome 19 for all cell types analysed. The distribution of gene density is also displayed at the bottom of the figure (purple spikes). (B) Pair-wise representation of eigenvectors between cell types genome-wide. Each dot represents a 50Kbp bin in the genome. Bins representing A compartment conservation are depicted in red, whereas in blue are depicted bins with B compartment conservation. Bins with unclear signal or compartment switching are represented in grey. The purple line is a LOESS curve showing the tendency of the compartment switching. Correlation values are represented for each pair-wise comparison. Legend – Fib: Fibroblast, Spg: Spermatogonia, L/Z: Leptonema/Zygonema, P/D: Pachynema/Diplonema, SpII: Spermatocytes II, RS: Round Spermatids.
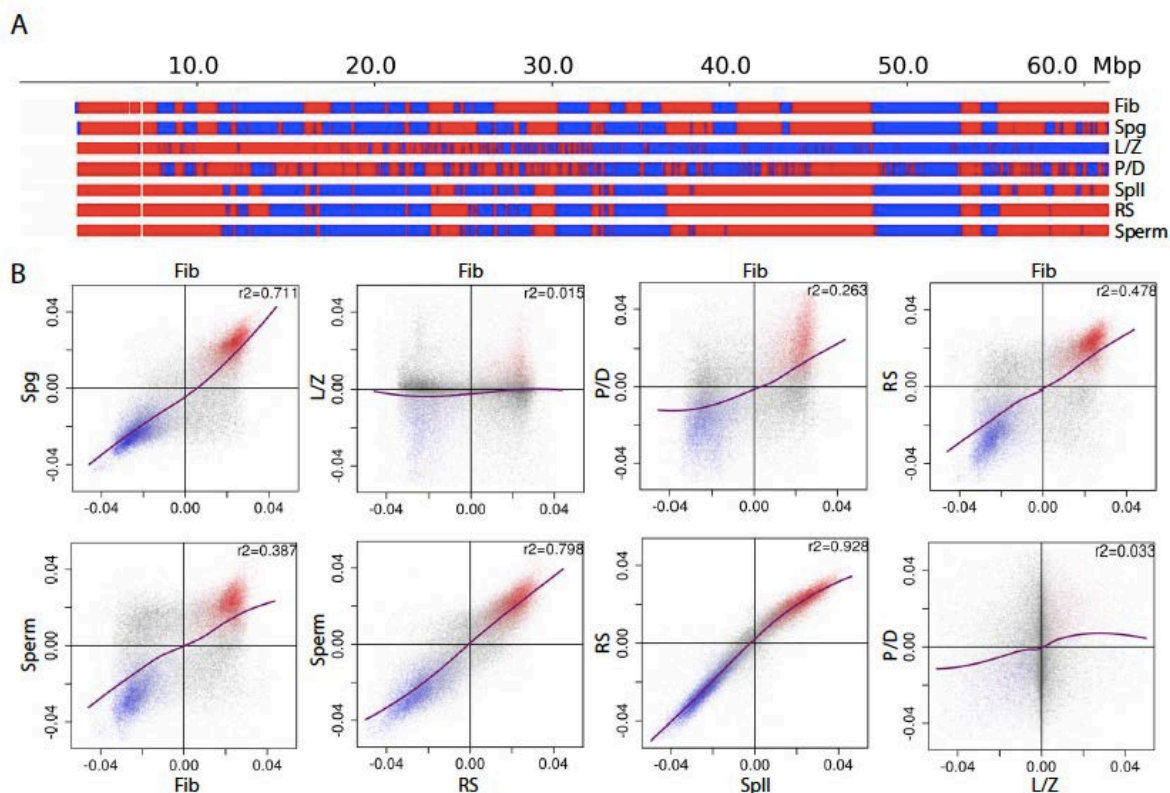
Supplementary Figure 4: Analysis of A/B compartments by 3D-FISH, related to Figure 3. (A) Heatmaps from the four chromosomal regions analysed by 3D-FISH in mouse chromosomes 12 and 14. Probes located at A-compartments are red, whereas probes located at B compartments are blue. (B) Two-colour 3D-FISH images of example pairs of probes within A-compartments in fibroblasts, spermatogonia and round spermatids (see supplementary methods for probe localisation). In all cases, left panels (i) represent the confocal x-y (upper section) and x-z sections (lower section), whereas right panels (ii) represent IMARIS three-dimensional reconstructions. (C) Lineal plot of the median genomic distances (in Kbp) between pairs of probes versus physical distances (in mm) in fibroblasts (n = 275), spermatogonia (n = 155) and round spermatids (n = 97) for both types of compartments (A in red and B in blue).
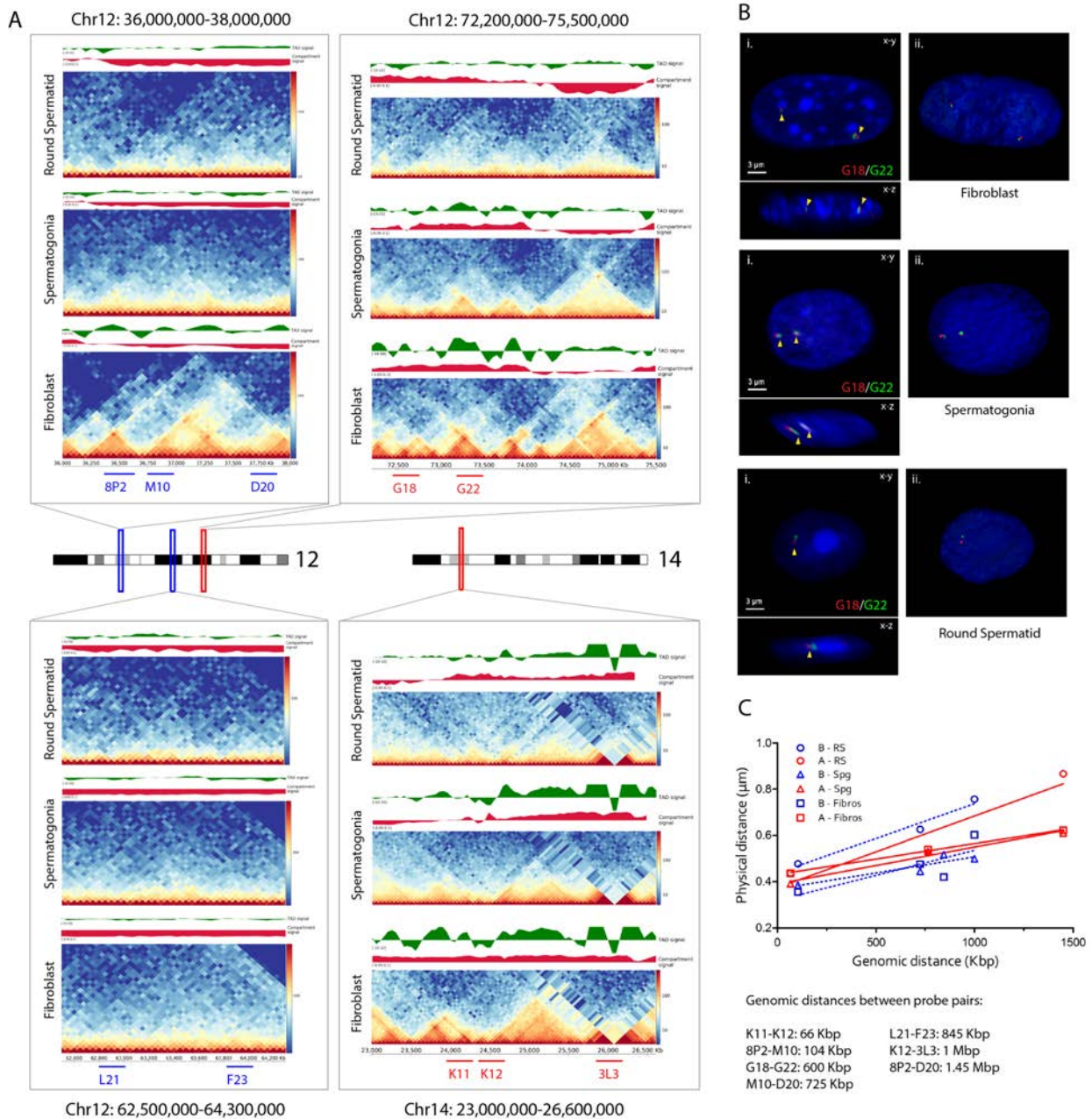
Supplementary Figure 5: Simulations of samples with different fibroblast and sperm content, related to Figure 3. (A) Heatmap showing correlation values, based on the pairwise similarity score calculated using HiCRep, among samples with different fibroblast and sperm content as well as the merged sperm replicates from SRP071784 (retrieved from Jung et al. 2017). (B) Contact probability as a function of genomic distance for samples with different fibroblast and sperm content as well as the merged sperm replicates from SRP071784. (C) Chromosome 8 region-specific ICE-corrected heatmaps at 50 kbp for samples with different fibroblast and sperm content, as well as from SRP071784.
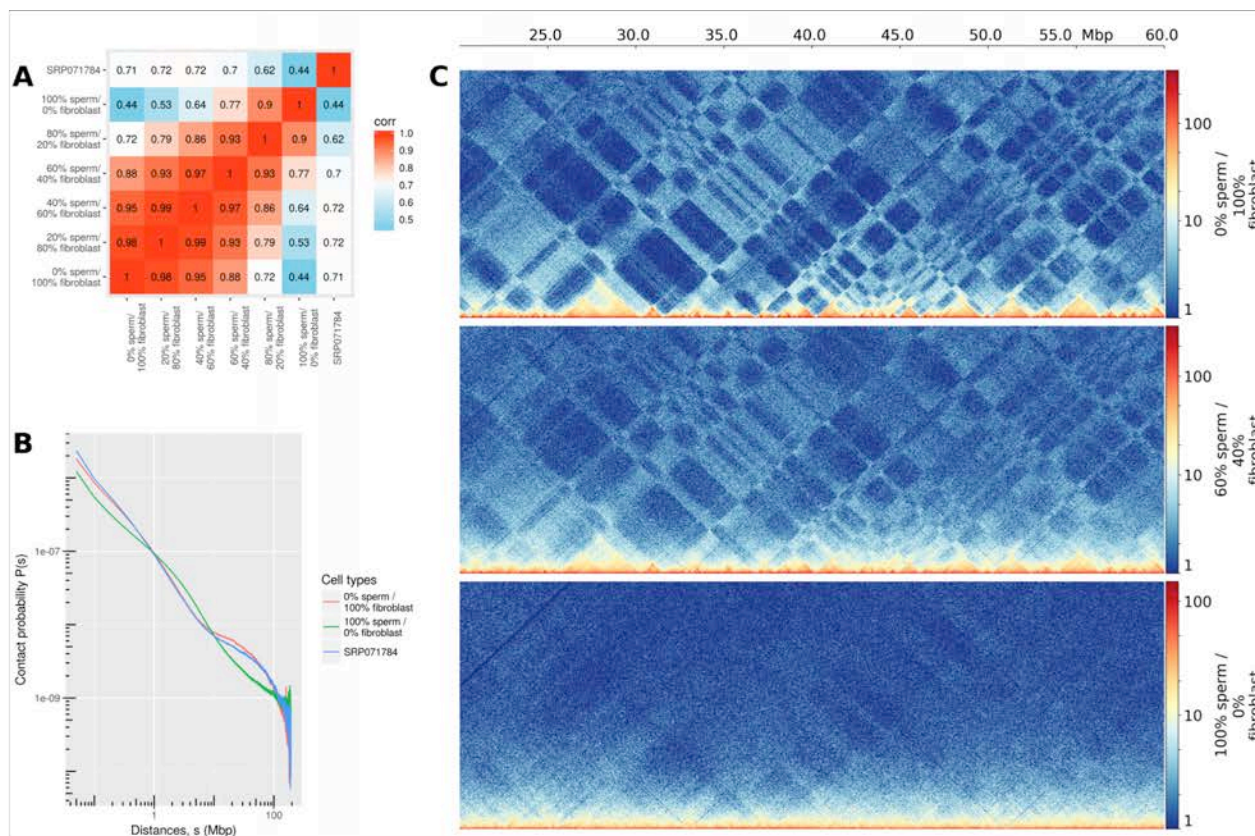
Supplementary Figure 6: Expression and intracellular localisation of CTCF and cohesins (REC8 and RAD21L) in mouse germ cells, related to Figures 5 and 6. (A) Western blotting analyses on whole nuclear extracts from mouse testis (two replicates) and fibroblasts (two replicates). Abbreviations: T: testis, F: fibroblast, R#: replicate. The band corresponding to CTCF was detected at 120 KDa, REC8 and RAD21L at 63 KDa, approximately. The β-tubulin loading control was detected at about 50 kDa. (B) Examples of immunofluorescence images showing the intracellular distribution of CTCF, RAD21L and REC8 in primary spermatocytes at pachytene stage. Double immunolabelling of CTCF (red)/ RAD21L (red)/ REC8 (red) with SYCP3 (green) are shown. Scale bar: 10 μm. (C) Examples of immunofluorescence images showing the intracellular distribution of CTCF, RAD21L and REC8 in round spermatids. Double immunolabelling of CTCF (red)/ RAD21L (red)/ REC8 (red) with SYCP3 (green) are shown. Scale bar: 10 μm. (D) Double immunolabelling of REC8 (green) and RAD21L (red) at pachytene stage showing only partial overlapping at the chromosome axes by STED microscopy. Plot of the mean correlation between REC8 and RAD21L at the autosomes (n=76 axial elements, AEs; mean ± s.d). (E) Example of an autosome axes immunolabelled for REC8 (green) and RAD21L (red) (upper panel). Plot represents normalised signal intensity profile of REC8 (green) with RAD21L (red) (middle panel). Lower plot shows regression analysis of the correlation between REC8 and RAD21L (lower panel). (F) Example of differential immunolabelling of (i) RAD21L (increased labelling) and (ii) REC8 (decreased labelling) at the XY bivalent in a spermatocyte at pachytene. (iii) Magnification of the XY bivalent is shown for each cohesin. (iv) Plot representation of the labelling gain/loss of RAD21L and REC8 fluorescence intensity (depicted in %) along the XY axes in comparison with autosomes (n=19 cells).

A

B   Primary Spermatocytes

C   Round Spermatids

D

E

F
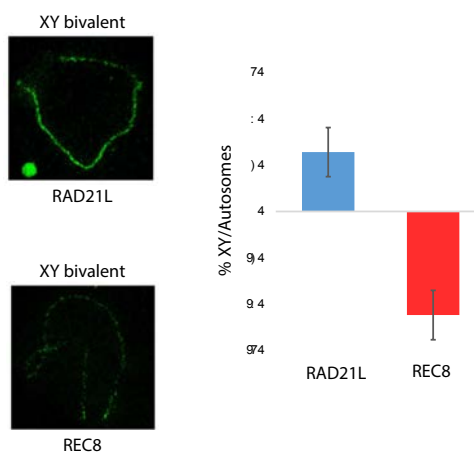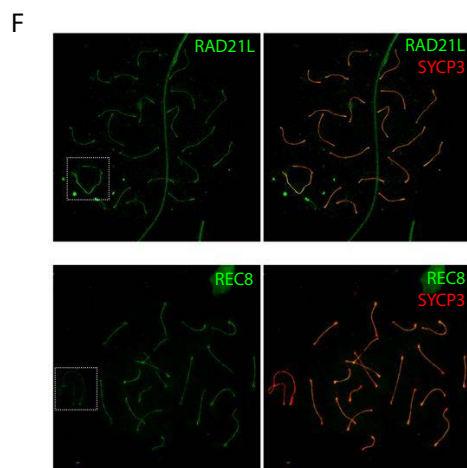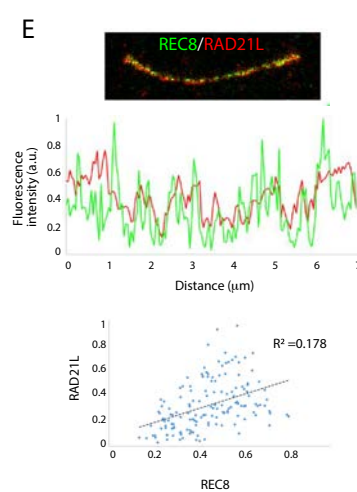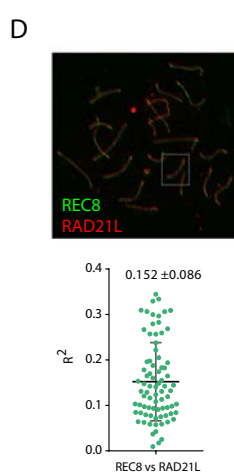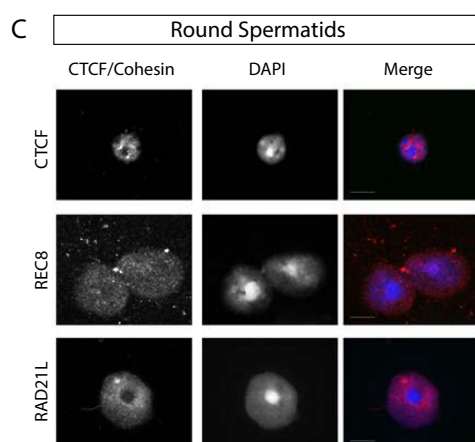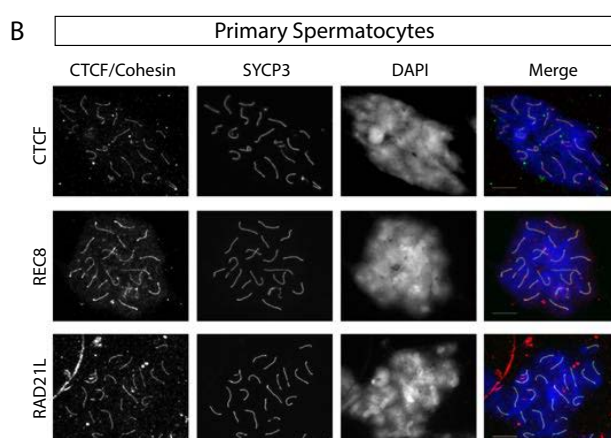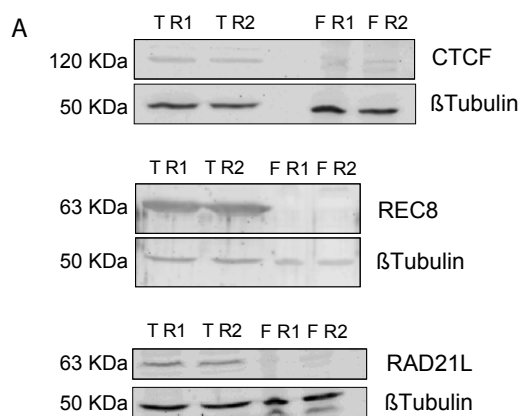
Supplementary Figure 7: Evolutionary strata and ampliconic regions in the mouse X chromosome, related to Figure 7. (A) Evolutionary strata (from Cortez et al., 2014) for the human and mouse X chromosomes. Strata were assigned to the mouse X using human synteny (https://ensembl.org/Homo_sapiens/Location/Synteny?r=X). White regions are breaks in synteny that usually correspond to amplicons (see panel B). (B) RNA-seq tag counts on the X chromosome in spermatogonia (Spg), pachynema/diplonema (P/D), round spermatids (RS) and sperm. Colours in the centre represent evolutionary strata (as in panel A), green represent amplicons (see Table S7 for coordinates). Peaks to the left represent counts for uniquely mapped RNA-seq reads. Peaks to the right are counts for both uniquely mapped and multi-mapped RNA-seq reads. As expected, in the ampliconic regions counts are higher when including multi-mapped reads.
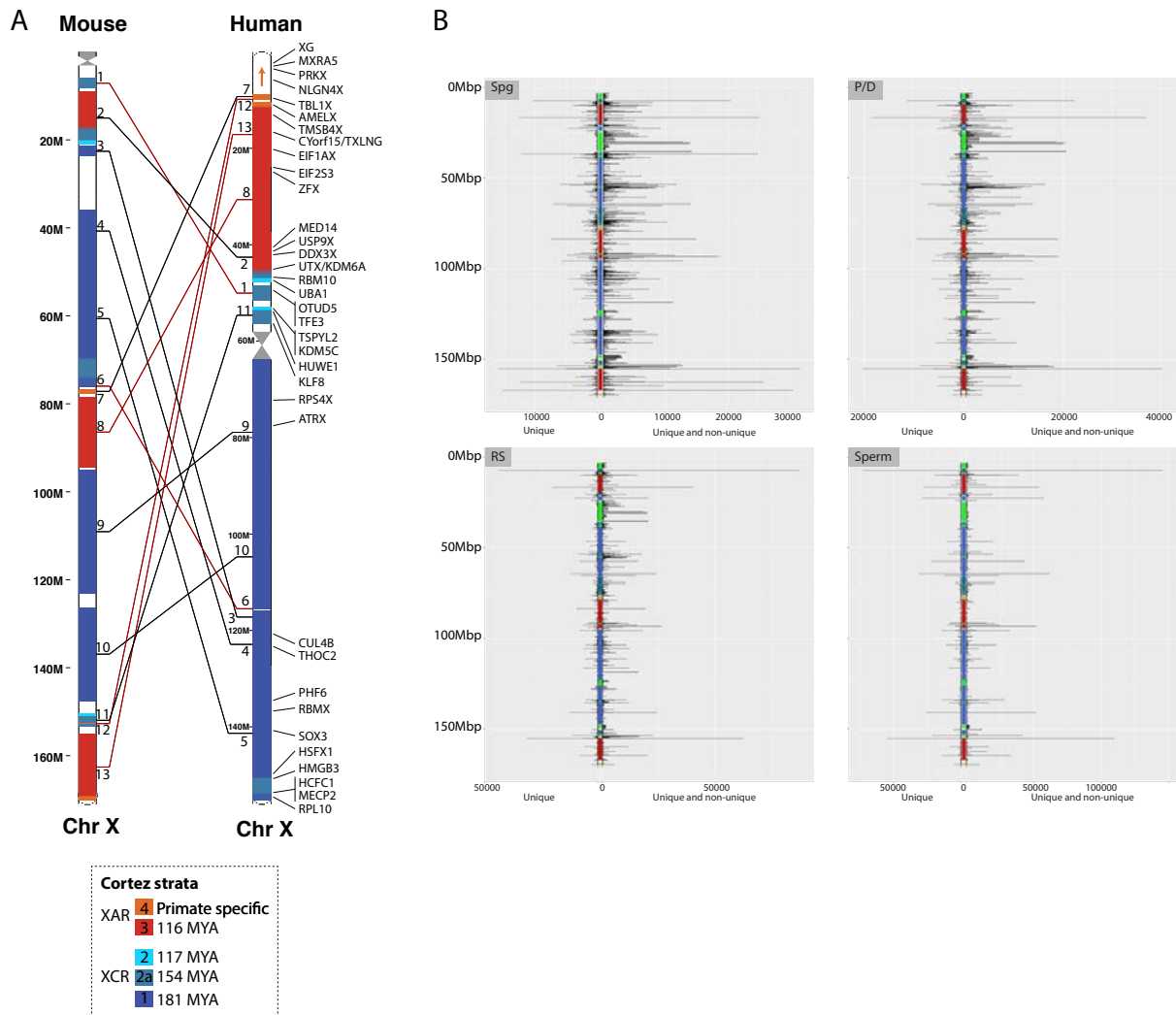
Supplementary Table 1: Hi-C quality metrics per individual libraries. Related to Figures 1 and 2.

Biological replicates	N° cells (million)	Library ID/Lane	Raw read pairs	Trimmed read pairs	Uniquely mapped read pairs	Self-circle	Dangling-ends	Error	Extra dangling-ends	Too short	Too large	Duplicated	Random breaks	Total valid read pairs
Fibroblast (rep1)	5	CB110ANXX_5_6_1	248031330	240260284	174084630	221384	553491	330946	5114580	8427269	12671	5818745	160097	154701718
Fibroblast (rep2)	5	CB110ANXX_6_7_1	192593621	187988387	132302339	159831	4134246	2030812	4129467	8309863	9193	9664537	889071	105293483
		CAV5EANXX_6_7_1	59889457	57753942	40530730	48732	1273768	609853	1263464	2552226	2885	957878	275008	33981882
Spermatogonia (rep1)	4.6	CBLJLANXX_1_0_1	278816401	267840097	193623984	320411	229587	70028	9096613	14213523	6277	14758971	53602	157495526
Spermatogonia (rep2)	3.1	CC5N8ANXX_7_0	1149936300	1098470448	193716658	324949	1063965	49219	4766585	12788396	5429	17528840	476662	159034223
L/Z (rep1)	1.7	CBLPLANXX_1_7_1	225472936	216917242	158331725	173222	91663	46047	7832006	10375259	5048	6302362	14855	134826221
P/D (rep1)	8.5	CAVGPANXX_3_18	24613697	23367598	16786120	41044	917421	296336	537549	1003614	873	324981	169589	13704277
		HHW7MBBXX_3_18	208519245	192672485	139504559	351389	7452993	2356644	4720624	8091386	7256	29233821	1395178	91722034
P/D (rep2)	7.2	CAUGDANXX_4_4_0	187070823	181953633	129656506	314503	5622403	770057	5877661	7183789	7981	6265771	1023401	104045257
		CAYRAANXX_4_4_1	63514173	61229210	44113539	106268	1910514	255197	2172753	2806677	2805	769256	366695	36383581
P/D (rep 3)	5.2	CBLPLANXX_2_15_1	278216841	265129919	192207789	229085	117948	57554	9069175	11492640	5924	7393339	18334	165268550
SpII (rep1)	6.9	HH3YTBBXX_6_13	92642190	90292818	68692253	99791	749197	662096	1514786	2490511	3827	7306689	63668	56465964
SpII (rep2)	8.8	HH52MBBXX_2_13	123339346	118739034	90276580	130940	998214	1220863	2008285	3271582	4770	10353127	83753	73149452
RS (rep1)	76.5	HHVVNBBXX_6_16	228741418	214721068	155162606	370068	4721597	1982027	4878668	9589424	6746	22008562	905986	114608474
		CAVGPANXX_3_19	41711076	40552575	27838653	73284	3683985	1520471	1250116	1920249	2025	1652498	567899	18219785
RS (rep2)	31.2	HHW7MBBXX_6_19	197560571	183760983	127434427	343429	16151547	7158102	5898111	8622094	8698	32749550	2406199	65912575
		CAV5EANXX_6_5_1	53504275	51471297	37297010	127329	7038796	559099	3029377	2269490	5666	1375399	1358556	23306483
Sperm (rep1)	44.4	CB110ANXX_4_5_1	201759632	196676968	143209990	487167	27084226	2182609	11763352	8692935	21478	16514284	5109160	81690561
		CBLPLANXX_3_4_1	248883003	240177238	172429305	155757	264201	44354	6149922	9676501	8814	6677087	98661	150537722
Sperm (rep2)	69.4	CC73FANXX_5_13	879788376	847261392	151996434	127487	566413	42930	7208022	8177371	5452	11958591	395193	1255517920

Supplementary Table 2: Hi-C quality metrics per cell type. Related to Figures 1 and 2.

INFO PER CELL TYPE	Fibroblasts	Spermatogonia	L/Z	P/D	SpII	Round Spermatids	Sperm
Raw read pairs	500,514,408	1,428,752,701	225,472,936	761,934,779	444,722,954	494,535,554	1,128,671,379
Trimmed read pairs	486,002,613	1,366,310,545	216,917,242	724,352,845	423,752,920	471,961,823	1,087,438,630
Mapped uniquely read pairs	346,917,699	387,340,642	158,331,725	522,268,513	314,131,439	335,780,080	324,425,739
Self-circle (% relative uniquely mapped)	0.124	0.167	0.109	0.204	0.192	0.304	0.087
Dangling-end (% relative uniquely mapped)	1.723	0.334	0.0579	3.251	2.071	15.839	0.263
Error (% relative uniquely mapped)	0.859	0.031	0.029	0.772	1.231	3.554	0.0270
Extra dangling-end (% relative uniquely mapped)	3.029	3.579	4.946	4.238	2.680	6.399	4.154
Too short (% relative uniquely mapped)	5.563	6.971	6.553	5.849	4.902	6.431	5.496
Too large (% relative uniquely mapped)	0.007	0.003	0.003	0.005	0.005	0.011	0.004
Duplicated (% relative uniquely mapped)	4.744	8.336	3.980	8.936	12.646	16.033	5.870
Random breaks (% relative uniquely mapped)	0.383	0.137	0.009	0.603	0.338	2.749	0.159
Total valid read pairs	293,977,083	316,529,749	134,826,221	411,123,699	244,223,890	189,129,404	276,055,642
Average genome coverage	16 179	17.420	7.420	22.626	13.441	10.409	15.193
Total valid (% relative to Raw)	58 735	22.154	59.797	53.958	54.916	38.244	24.458
Total valid (% relative to Trimmed)	60 489	23.168	62.156	56.757	57.633	40.073	25.386
Total valid (% relative to Mapped uniquely)	84 740	81.719	85.154	78.719	77.746	56.325	85.090

Supplementary Table 3: A/B compartments and TADs. NA: not analysed. Related to Figures 2 and 3.

cell type	Compartment mean size (bp)	TAD number	TAD mean size (bp)	% Robust TADs (score 10 and 9)
Fib	1000367	2002	1361638	72.33
Sg	755820	834	3268585	71.70
P/D	NA	294	9272109	79.59
L/Z	NA	305	8937705	74.25
SpII	760457	5004	544764	12.23
RS	869856	4649	586363	8.50
Sperm	933584	1042	2616123	14.87

Supplementary Table 4: BACs selected for the 3D-FISH analysis. Related to Figure 3.

Probe	Chr	Genome coordinates (bp), start-end	A/B
RP23-318P2	12	36,438,632-36,644,697	B
RP23-351M10	12	36,748,530-36,944,925	B
RP23-34D20	12	37,656,313-37,873,485	B
RP23-382L21	12	62,887,606-63,091,331	B
RP23-304F23	12	63,936,422-64,142,895	B
RP23-118G18	12	72,372,755-72,569,213	A
RP24-92G22	12	73,333,061-73,572,221	A
RP23-218K11	14	24,101,846-24,278,172	A
RP23-157K12	14	24,344,637-24,545,440	A
RP23-93L3	14	25,993,682-26,147,325	A

Supplementary Table 5: RNA-seq quality metrics per individual libraries. Related to Figure 4.

Cell type	Sample ID	Number of read pairs	Uniquely mapped read pairs (%)	Multi-mapped read pairs (%)	Unmapped reads (%)	Uniquely mapped read pairs on genes (%)
Spermatogonia (rep 1)	P0000_N701-S505	37543704	30516058 (81.28%)	2740144 (7.30%)	4287502 (11.42%)	26728331 (71.19%)
Spermatogonia (rep 2)	P0000_N701-S517	38010687	28074303 (73.86%)	3540476 (9.31%)	6395908 (16.83%)	20473593 (53.86%)
Spermatogonia (rep 3)	P0000_N702-S505	34953928	27053571 (77.40%)	3075743 (8.80%)	4824614 (13.80%)	22602831 (64.66%)
Spermatogonia (rep 4)	P0000_N702-S517	32262491	23839185 (73.89%)	4079114 (12.64%)	4344192 (13.46%)	19311631 (59.86%)
Pachynema/diplonema (rep 1)	P0000_N701-S502	40170411	32865101 (81.81%)	2622824 (6.53%)	4682486 (11.66%)	28179005 (70.15%)
Pachynema/diplonema (rep 2)	P0000_N701-S506	35087621	29488533 (84.04%)	2552231 (7.27%)	3046857 (8.68%)	25716482 (73.29%)
Pachynema/diplonema (rep 3)	P0000_N702-S502	29130117	21967847 (75.41%)	2704119 (9.29%)	4458151 (15.30%)	18279098 (62.75%)
Pachynema/diplonema (rep 4)	P0000_N702-S506	34320275	26319970 (76.69%)	3217903 (9.38%)	4782402 (13.93%)	22198845 (64.68%)
Round spermatids (rep 1)	P0000_N701-S503	51211991	41336093 (80.72%)	4260103 (8.32%)	5615795 (10.97%)	35803461 (69.91%)
Round spermatids (rep 2)	P0000_N701-S507	30636921	26367194 (86.06%)	1484709 (4.85%)	2785018 (9.09%)	23947716 (78.17%)
Round spermatids (rep 3)	P0000_N702-S503	38753330	32455484 (83.75%)	2827907 (7.30%)	3469939 (8.95%)	29049945 (74.96%)
Round spermatids (rep 3)	P0000_N702-S507	27825912	22246393 (79.95%)	2035509 (7.315%)	3544010 (12.74%)	19287752 (69.31%)
Sperm (rep 1)	P0000_N701-S504	30756175	24260179 (78.88%)	2931414 (9.53%)	3564582 (11.60%)	20342978 (66.14%)
Sperm (rep 2)	P0000_N701-S508	34156455	29708743 (86.98%)	1432018 (4.19%)	3015694 (8.83%)	27099974 (79.34%)
Sperm (rep 3)	P0000_N702-S504	33023282	28523977 (86.37%)	1404504 (4.25%)	3094801 (9.37%)	25838380 (78.24%)
Sperm (rep 4)	P0000_N702-S508	30854197	26588200 (86.17%)	1306126 (4.23%)	2959871 (9.60%)	23990832 (77.75%)

Supplementary Table 6: DEG during spermatogenesis. Related to Figure 4.

Pair-wise comparison	Type	DEG number	Protein coding genes (%)	lncRNA (%)	asRNA (%)	Pseudogenes (%)	smRNA (%)	Unannotated (%)
Spermatogonia versus pachynema/diplonema	Up-regulated	5091	83.5	4.11	2.38	2.97	0.35	6.7
	Down-regulated	3842	68.17	12.88	6.59	7.11	0.52	4.74
Spermatogonia versus round spermatids	Up-regulated	2574	92	2.25	1.01	1.05	3.26	0.43
	Down-regulated	1105	83.62	6.88	4.43	2.9	1.99	0.18
Spermatogonia versus sperm	Up-regulated	6956	80.39	5.64	2.66	4.87	0.33	6.11
	Down-regulated	6124	56.99	16.1	8.69	11.95	0.77	5.5
Pachynema/diplonema versus round spermatids	Up-regulated	350	86.86	7.14	2.29	0.86	0	2.86
	Down-regulated	1609	73.65	9.51	7.33	3.85	0	5.66
Pachynema/diplonema versus sperm	Up-regulated	2619	77.36	8.29	3.97	5.65	0.15	4.58
	Down-regulated	3206	67.22	12.45	6.99	8.3	0.22	4.83
Round spermatids versus sperm	Up-regulated	1636	77.93	7.76	3.61	8.19	0	2.51
	Down-regulated	1626	70.91	10.27	7.07	7.5	0	4.24

Supplementary Table 7: Amplicons in the X chromosome. Related to Figure 7.

Start (bp)	End (bp)	Amplicon #
3076875	5671026	1
8327424	8880750	2
8962134	9063861	3
24474308	35427093	4
37244976	38129967	5
53608032	53670408	6
54531208	55226876	7
70393901	70546385	8
73074345	73292976	9
74639371	74890297	10
77837901	78245921	11
91331690	92016333	12
94724569	94797933	13
102707865	102904882	14
103013782	103067153	15
118427227	118480737	16
123013511	126033906	17
135167624	135627705	18
147992993	149487784	19
152432477	152491860	20
154001590	154579360	21
154924012	155089793	22

Supplementary Table 8: TADs in the X chromosome. Related to Figure 7.

Cell type	TAD number	TAD mean size (bp)
Fib	101	1693564
Sg	62	2758871
P/D	26	6578846
L/Z	21	8145238
SpII	57	3000877
RS	69	2478986
Sperm	53	3227358

4.3 The impact of chromosomal fusions on 3D genome folding and recombination in the germ line

The Impact of Chromosomal Fusions on 3D Genome Folding and Recombination in the Germ Line

Covadonga Vara[1,2], Andreu Paytuví-Gallart[1,2,3], Yasmina Cuartero[4,5], Lucía Álvarez-González[1,2], Laia Marín-Gual[1,2], Francisca Garcia[6], Beatriu Florit-Sabater[1,2], Laia Capilla[2,7], Rosa Ana Sanchéz-Guillén[1,2,8], Zaida Sarrate[1], Riccardo Aiese Cigliano[3], Walter Sanseverino[3], Jacint Ventura[7], Marc A. Marti-Renom[4,5,9,10], François Le Dily[4,5], Aurora Ruiz-Herrera[1,2]*

[1]Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, 08193, Spain

[2]Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, 08193, Spain

[3]Sequentia Biotech, Carrer Comte d'Urgell 240, Barcelona, 08036, Spain

[4]Centre for Genomic Regulation, The Barcelona Institute for Science and Technology, Carrer del Doctor Aiguader 88, Barcelona, 08003, Spain

[5]CNAG-CRG, Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Baldiri Reixac 4, Barcelona, 08028, Spain

[6]Servei de Cultius Cel.lulars, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, 08193, Spain

[7]Department de Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, 08193, Spain

[8]Present address: Instituto de Ecología AC (INECOL), Red de Biología Evolutiva, Xalapa, Veracruz, Mexico

[9]Pompeu Fabra University, Doctor Aiguader 88, Barcelona, 08003, Spain

[10]ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

* corresponding author

4.3.1 Summary

Background: The spatial folding of chromosomes inside the nucleus has regulatory effects on gene expression, yet the impact of genome reshuffling on this organization remains unclear. Even more fragmented is our understanding of the heritability of structural variations in genome folding. Here, we take advantage of chromosome conformation capture followed by deep sequencing (Hi-C) in combination with SNP genotyping and analysis of crossover events to study how the higher-order chromatin organization, chromosomal spatial occupancy and recombination landscapes are affected by chromosomal fusions in the mammalian germ line.

Results: We analyzed a unique wild system of house mice characterized by recent chromosomal fusions. We demonstrate that these fusions alter the nuclear architecture during meiosis, impacting on the inter- and intra-chromosomal interactions ratio in meiotic (primary spermatocytes) and post-meiotic (round spermatids) cells. We detected an increased rate of heterologous interactions in primary spermatocytes, with alterations in both chromosome synapsis and axis length. These topology disturbances were associated with changes in genomic landscapes of recombination, resulting in detectable genomic footprints. Moreover, the reorganization of chromosomal nuclear occupancy in prophase I exposed chromosomal domains to novel regulatory environments. In round spermatids, heterologous interactions decreased genome-wide in mice with chromosomal fusions, highlighting the more spatially constrained genome conformation in post-meiotic cells.

Conclusions: Overall, we show the architectural genome-wide implications of structural variations during mouse spermatogenesis. As such, chromosomal fusions impact the dynamic three-dimensional genome topology of germ cells in two ways: (i) altering meiotic chromosomal nuclear occupancy and synapsis, and (ii) reshaping landscapes of recombination.

Keywords: Meiosis, recombination, genome structure, Hi-C, TADs, compartments, chromosomal fusions, house mouse

4.3.2 Background

Higher-order chromatin structure demarcates the limits of gene-regulatory domains (Dixon et al. 2012; Lieberman-Aiden et al. 2009; Rao et al. 2014). Thus, disturbances of domain architecture due to genome reshuffling (i.e., inversions, fusions or indels) represent a non-gradual change in gene regulation as shifting of domain boundaries expose genes to novel regulatory environments (Kraft et al. 2019; Laugsch et al. 2019). Models (Deakin et al. 2019; Farré et al. 2015) and growing experimental evidence (Bompadre and Andrey 2019; Franke and Gómez-Skarmeta 2018) suggest that indels and inversions can alter interactions between contiguous topological associated domains (TADs) which can lead to oncogene activation, morphological alterations and novel gene functions. However, the impact of balanced chromosomal reorganizations, such as Robertsonian (Rb) fusions (Robertson 1916), on genome architecture and its heritability are less explored. This is of particular relevance since Rb fusions represent the most common chromosomal rearrangement in nature (from plants to mammals) (Robinson 1995), and are linked to recurrent miscarriages, infertility and aneuploid offspring in humans (Scriven 2001; Wilch and Morton 2018). In fact, it has been long suggested, although not yet empirically demonstrated at the genome level, that the presence of chromosomal fusions in the germ line can alter segregation patterns [the so-called 'inter-chromosomal effect' (Lucchesi and Suzuki 1968)].

Germ cells are a unique cell model to test the genome-wide impact of chromosomal fusions. They have sequential developmental stages that involve dramatic and tightly regulated chromosomal movements and chromatin remodeling. These include changes in intra-/inter-chromosomal interaction ratios, distance-dependent interaction frequencies, genomic compartments, TADs, occupancy of insulator proteins (CTCF and cohesins) and gene expression (Alavattam et al. 2019; Patel et al. 2019; Paytuví-Gallart 2019; Vara et al. 2019b; Wang et al. 2019b). The delicate fine-tuning between chromatin remodeling, architectural proteins and cell-specific gene expression is crucial during

the first meiotic prophase (prophase I) when homologous chromosomes align, pair, synapse and recombine (Vara et al. 2019b).

Recombination has a dual role in sexual reproduction: (i) it assembles new combinations of allelic variants, contributing to the maintenance of genetic diversity and (ii) establishes physical associations between homologous chromosomes that enable faithful chromosomal segregation during meiosis. Importantly, recombination can be modulated not only by factors that control the formation of meiotic crossovers (COs) during early meiosis [e.g., chromosome axis length is determined by chromatin loop length (Ruiz-Herrera et al. 2017; Wang et al. 2019a, 2015b)], but also by large-scale structural reorganizations that can dramatically alter the genomic landscape (Capilla et al. 2014; Farré et al. 2013; Torgasheva et al. 2013; Ullastres et al. 2014). Yet, the impact of large-scale genome reshuffling (e.g., chromosomal fusions) on the three-dimensional genome topology in germ cells and its implications for recombination remain unknown.

Here, we take advantage of chromosome conformation capture followed by deep sequencing (Hi-C) in combination with cytological analysis of COs events and SNP genotyping to study how genome folding and recombination landscapes are affected by chromosomal fusions in the mammalian germ line. We analyzed a unique wild system of the Western European house mouse (*Mus musculus domesticus*) from the northeast of the Iberian Peninsula, the so-called 'Barcelona Rb system' (BRbS), characterized by a recent evolutionary origin of chromosomal fusions (Capilla et al. 2014; Medarde et al. 2012; Piálek et al. 2005; Vara et al. 2019a). The standard karyotype of *M. m. domesticus* consists of 40 acrocentric chromosomes, contrasting BRbS mice, which have a variety of diploid numbers (2n) ranging from 2n=27 to 2n=40 (Medarde et al. 2012). This system is characterized by the presence of different Rb fusions distributed in non-geographically coincident clines leading to a progressive reduction in diploid numbers towards the center of the range (Medarde et al. 2012).This natural model permits interrogation of the impact Rb fusions have on chromatin remodeling and fine-scale recombination in the germ line. In particular, we studied how chromosomal fusions alter the nuclear architecture at different hierarchy levels in meiotic (i.e., primary spermatocytes) and post-meiotic cells (i.e., round spermatids), and discuss the implications for evolution and fertility.

4.3.3 Results

4.3.3.1 Variation in recombination rates in natural populations of house mice

We analyzed the variation in recombination rates in wild-caught BRbS mice and the potential impact of Rb fusions on these patterns. We conducted an integrative approach that combined the cytological mapping of CO events directly on male germ cells (reflecting recombination at the Mbp scale) (**Figures 4.10, 4.11 and Figure S1**) together with estimates of linkage disequilibrium based on SNP genotyping (recombination at the kbp scale) (**Figure S2**). We first experimentally determined the number (frequency) of COs along chromosomal axes in 45 wild mice by the immunodetection of the recombination protein MLH1 (a marker of COs) on pachytene bivalents (**Figure 4.10 A-C, Table S1**) (see **section 3.2.7**). We also included three laboratory mice (strain C57BL/6J) for comparison. The CO survey in wild-caught mice included 15 all-acrocentric individuals (two standard populations, 2n=40) and 30 mice with Rb fusions (three populations, 2n=39-28), allowing for microscopy visualization of a total of 1,468 spermatocytes (**Figure 4.10 B-D, Table S1**). Of Rb mice, the population Sant Sadurní d'Anoia was characterized by a low number (from one to three) of Rb fusions, whereas the Castelldefells and Viladecans populations presented from four to six Rb fusions (2n=33-28) (**Table S1**).

The observed average number of COs ranged between individuals from 20.16 ($\pm$1.18) MLH1 foci per cell to 22.12 ($\pm$4.56) in wild-caught standard mice, and from 18.13 ($\pm$1.78) to 21.82 ($\pm$2.21) in Rb mice (**Figure 4.10 D**). A population-level analysis of COs detected differences between and within populations (**Figure 4.10 D**). Mice from Rb populations with four or more fusions presented broader inter-individual variability regarding the number of MLH1 foci numbers per cell (**Figure**
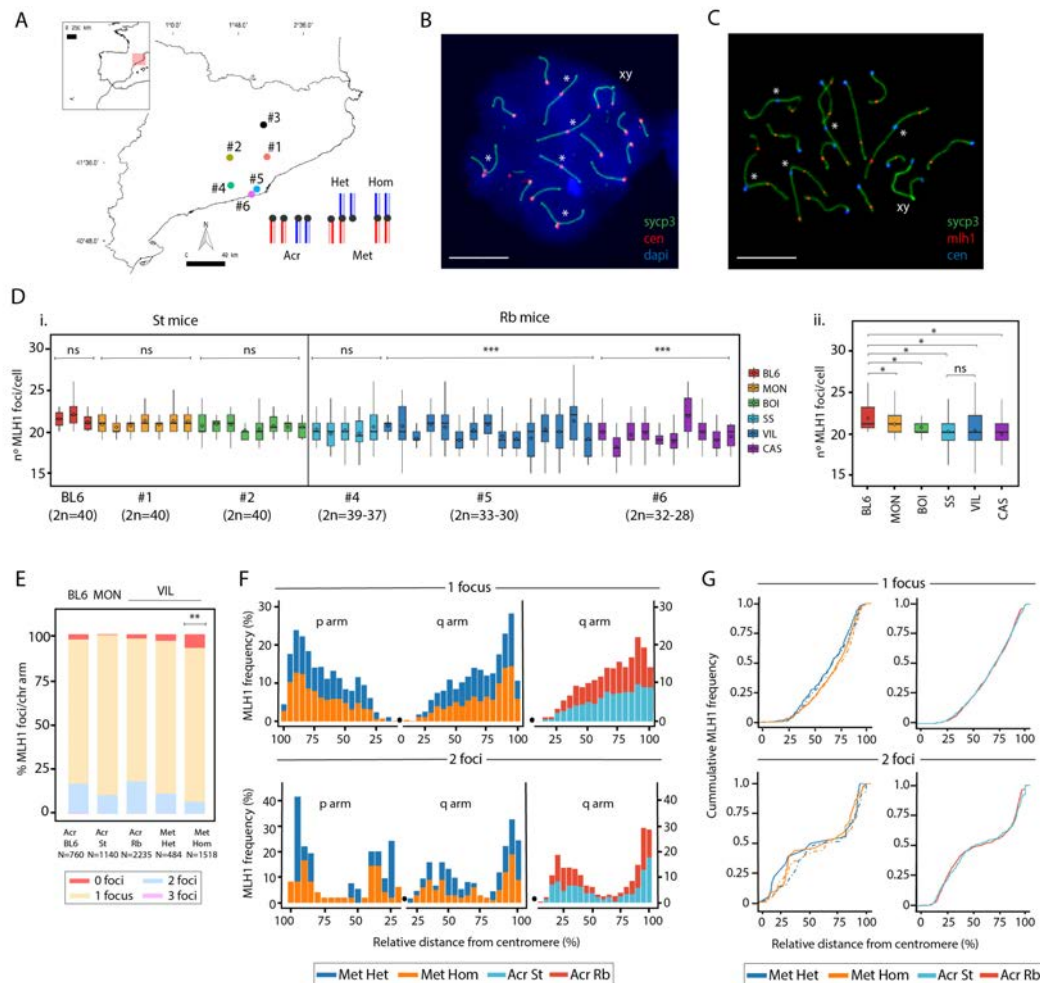
Figure 4.10: Genomic landscapes of recombination. **(A)** BRbS populations sampled. See Table S1 for population assignment. Inset: types of chromosomal fusions. Chromosome type legend: Acr, all acrocentric chromosomes of standard mice; Met Het, Rb chromosomes in the heterozygous state of Rb mice; Met Hom, Rb chromosomes in the homozygous state of Rb mice. **(B)** Immunofluorescence of a spermatocyte at pachytene from a Rb mouse (2n=32): SYCP3 (green), centromeres (red) and DNA (blue). Asterisks indicate Rb fused chromosomes. Scale bars = 10 μm. **(C)** Immunofluorescence of a spermatocyte at pachytene: SYCP3 (green), MLH1 (red) and centromeres (blue). Asterisks indicate Rb fused chromosomes. **(D)** Boxplots representing the number of MLH1 foci/cell per specimen represented (i) individually (colors correspond to panel A) and (ii) per population. (i) Three laboratory mice (BL6) are included for comparison. Diamonds indicate mean values with diploid numbers shown for each population. P-values (Kruskal–Wallis, ***$p<0.0001$) represent intra-population differences (CAS, Castelldefels; BOI, Castellfollit del Boix; MON, Caldes de Montbuí; SS, Sant Sadurní d'Anoia; and VIL, Viladecans). N.s.: no significant. (ii) Mean numbers of MLH1 foci/cell in laboratory mice (BL6) and wild-caught St and Rb mice represented per population. P-values (Dunn's test, *$p<0.05$) represent differences between populations. **(E)** Percentage of chromosomal arms showing different number (0, 1, 2 or 3) of MLH1 foci per chromosome (Chi square, **$p<0.001$). N = number of chromosomes analyzed. **(F)** Distribution of MLH1 foci along individual chromosomal arms with one (upper panels) and two MLH1 foci (lower panels). The X-axis represents the positions on the chromosomal axes from the centromeric end (left black dots) to the distal telomere (right). The Y-axis indicates the frequency of MLH1 foci in each 10% interval of chromosome arm length. Number of chromosomes analyzed per chromosome type: Acr St: 1,254 chromosomes; Acr Rb: 1,774 chromosomes; Met Het: 178 chromosomes; Met Hom: 563 chromosomes. **(G)** Cumulative frequency plots representing the chromosomal arms analyzed in panel 1F. Chromosome type legend: Acr St, all acrocentric chromosomes of standard mice (bivalents); Acr Rb, acrocentric chromosomes of Rb mice (bivalents); Met Het, Rb chromosomes in heterozygous state of Rb mice (trivalents); Met Hom, Rb chromosomes in homozygous state of Rb mice (quadrivalent).

4.10 D). In fact, mean numbers of COs per cell were positively correlated with diploid numbers (Spearman, $p<0.001$), and therefore negatively with the number of Rb fusions (Spearman, $p<0.0001$) (**Figure S1 A**). Despite fewer CO events, Rb mice presented significantly more inter-individual CO co-variation than standard mice (Mann-Whitney test, $p<0.0001$) (**Figure S1 B**). As such, mice with Rb fusions showed significantly lower mean numbers of COs per cell (20.00 ± 2.02) when compared to wild-caught standard (21.06 ± 1.92) and C57BL/6J mice (21.16 ± 1.76) (Dunn's test, $p<0.001$) (**Figure

4.10 D), hence confirming previous surveys conducted on distinct European Rb systems (Capilla et al. 2014; Merico et al. 2003; Merico et al. 2013). Mirroring descriptions from other mammals (Ruiz-Herrera et al. 2017; Wang et al. 2019a), both the mean numbers of COs and meiotic double strand breaks (DSBs) (here exemplified as RAD51 foci) were correlated in both standard (C57BL/6J strain and wild-caught mice) and Rb mice (Spearman, p<0.05, **Figure S1 C**). The variation in RAD51 foci numbers between laboratory mice and BRbS mice are expected based on CO counts, as these to processes are interrelated (Ruiz-Herrera et al. 2017; Wang et al. 2019a). As such, differences in DSB numbers between standard and Rb mice, together with the observation of a significantly high COs/DSBs ratio in Rb mice (**Figure S1 D**), suggests that a higher proportion of DSBs are resolved as COs when compared to standard mice (**Figure S1 E**).

4.3.3.2 Chromosomal fusions reshape genomic landscapes of recombination

Since each bivalent requires one CO to ensure proper chromosomal segregation (and hence fertility) during the first meiotic division [the so-called 'obligatory CO', (Petes 2001)], we quantitatively assessed whether the number of COs per chromosomal arm (0, 1, 2 or 3) was altered in Rb mice when compared to wild-caught standard and laboratory mice. We analyzed six standard mice (2n=40) from Caldes de Montbuí population and seven Rb mice (2n=31-32) from the Viladecans population (**Figure 4.10 E, Figure S1 D and Table S1**). We also included three control BL6 mice. The chromosomes were assessed in five categories: (i) acrocentric chromosomes of laboratory mice (Acr BL6), (ii) acrocentric chromosomes of wild-caught standard mice (Acr St), (iii) acrocentric chromosomes of Rb mice (Acr Rb), (iv) Rb chromosomes in heterozygous state of Rb mice (Met Het) and (v) Rb chromosomes in homozygous state of Rb mice (Met Hom) (**Figure 4.10 E**). In Rb mice from Viladecans, our analysis indicated that irrespective of the chromosomal complement (2n=31 or 2n=32), the CO distribution per chromosome arm had the same pattern (**Figure S1 F**). As a result, data from Viladecans mice were pooled together for subsequent downstream analysis of CO distribution.

We detected that the frequency of chromosome arms with zero CO were low (below 3%) in standard mice (C57BL/6J mice and wild-caught standard mice), mirroring previous observations (Capilla et al. 2014). However, in Rb mice with fusions in the heterozygous state we detected an increased proportion of chromosomal arms with zero CO (3.73%), and a statistically significant increase in Rb mice with fusions in homozygous state (7.71%) (Chi-square, p<0.0001, **Figure 4.10 E and Figure S1**). Moreover, in Rb mice with homozygous fusions we detected a significant decrease in the frequency of chromosomal arms with two COs (6.66%), when compared with their non-fused counterparts (i.e., acrocentric chromosomes) (17.8%) or Rb mice with heterozygous fusions (11.2%) (Chi-square, p<0.0001, **Figure 4.10 E**). These observations suggest that the reduced CO frequency observed in Rb mice is due to a reduction of chromosome arms with two COs, especially when the Rb fusions are homozygous (Figure **Figure 4.10 E**).

We next examined how chromosomal CO distribution was affected by Rb fusions. CO distribution along acrocentric chromosomes (i.e., non-fused) was similar in both standard (Caldes de Montbuí) and Rb mice (Viladecans). However, we detected differences between fused chromosomes in the homozygous versus heterozygous state (**Figure 4.10 F-G**). Single CO events in heterozygous fusions were located towards medial-distal regions (from 25% of the chromosome arm length onwards) (**Figure 4.10 F-G**), whereas there was a displacement in homozygous fusions towards distal regions (80-90% of the chromosome arm length), especially in q-arms. On chromosome arms with two COs, a bimodal distribution was detected in both standard and Rb mice, although there were again contrasting patterns between fused chromosomes in the homozygous and heterozygous states. On fused chromosomes in heterozygous state the majority of COs were localized either in proximal (15-20% of the chromosome arm length) or very distal (90-95% of the chromosome length) regions, especially on p-arms (**Figure 4.10 F-G**). In contrast, for chromosomes in homozygous state COs appeared away from the centromere, at medial regions from 25% of the chromosome arm length onwards. Our results indicate that chromosomal fusions reshape recombination landscapes by both

reducing the number of COs per chromosome arm and by displacing recombination events away from the centromere.

Perturbations in obligatory CO frequencies and overall CO distribution can impair meiosis and affect reproductive fitness. As such, we analyzed sperm viability. Rb mice presented a significantly higher fraction of immobile sperm (66.1%±21.2 vs. 31.7%±8.8, Mann-Whitney test, p=0.006) and increased sperm mortality (72.9%±18.5 vs. 32.1%±16.2, Mann-Whitney test, p= 0.005) when compared with standard mice.

To assess the fine scale genomic impact of both the overall reduction and chromosomal re-distribution of COs, we analyzed the landscape of genomic divergence (expressed as F_{ST} values) and estimated recombination rates based on linkage disequilibrium (expressed as $4N_{e}r/kbp$) (see Methods). In accordance with the variation in recombination rates in wild-caught BRbS mice, we also detected differences in molecular diversity between populations (**Tables S2, S3 and S4**). Both principal component analysis and estimations of population structure revealed heterogeneity of genetic clusters according to populations (**Figures S2 A and B**). Although estimates of observed and expected heterozygosity were similar between populations, standard mice presented higher allelic richness than in Rb mice (**Table S2**). Rb mice presented higher F_{ST} estimates than standard mice, both when comparing divergence between standard and Rb populations and between Rb populations (Tukey-Kramer test, p<0.005) (**Tables S3 and S4**). Consistent with the cytological analysis, we observed an overall reduction of recombination rates (expressed as $4N_{e}r/kbp$) on Rb mice when compared to BRbS standard mice (Mann-Whitney test, p<0.001; **Figures S2 C and D**).

4.3.3.3 Both axes length and CO chromosome distribution are modulated by chromosomal fusions

Variation in both chromatin loop length and chromosomal axis length can alter CO frequencies (Ruiz-Herrera et al. 2017; Wang et al. 2019a, 2015b); therefore we sought to understand how these characteristics were affected by chromosomal fusions. We did not observed differences in the total axis length per cell in Rb mice when compared to standard mice (**Figure S3**). However, when analyzing chromosomes individually in Rb mice, fused chromosomes arm length was longer than the acrocentrics (Mann-Whitney test, p<0.0001) (**Figure 4.11A**). Accordingly, fused chromosomes had significantly lower CO density than acrocentrics (Mann-Whitney test, p<0.0001), irrespectively of their state (homozygous or heterozygous) (**Figure 4.11A**). Remarkably, heterozygous fused chromosomes were significantly longer than when in a homozygous state (Dunn's test, p<0.0001) (**Figure 4.11 A and Figure S3 B**). These observations suggest that Rb fusions affect both chromosomal axis length and CO chromosome distribution. Likewise, CO formation is established in a chromosome-specific manner, mirroring recent observations in all-acrocentric mice (Wang et al. 2015b).

4.3.3.4 Fused chromosomes perturb pairing when heterozygous and centromere misalignment when homozygous, affecting both CO distribution

Meiotic DSBs are repaired in the context of the chromosomal axes as homologous chromosomes pair and synapse (Baudat et al. 2013; Lam and Keeney 2015). Thus, disturbances of homologous pairing during prophase I are expected to affect CO patterns. Concomitant with this view, perturbed pairing in Rb mice had an effect on CO distribution (**Figure 4.11 B-E**). We analyzed five Rb mice from the Viladecans population, categorizing three different pairing states in heterozygous fusions: (i) fully synapsed, (ii) open (presenting a synaptonemal complex bridge) and (iii) asynapsed chromosomes (**Figure 4.11 B-C and Figure S3 C**). Disturbed synapsis influenced CO distribution per chromosome arm, while the average number of CO events per arm was similar between individuals (**Figure 4.11 C and Figure S3 C**). That is, while fully synapsed heterozygous fusions generally presented one, homogeneously distributed, CO event per arm, asynapsed heterozygous fusions presented a bimodal CO distribution, with COs either in intermediate positions (40-50% of the chromosome
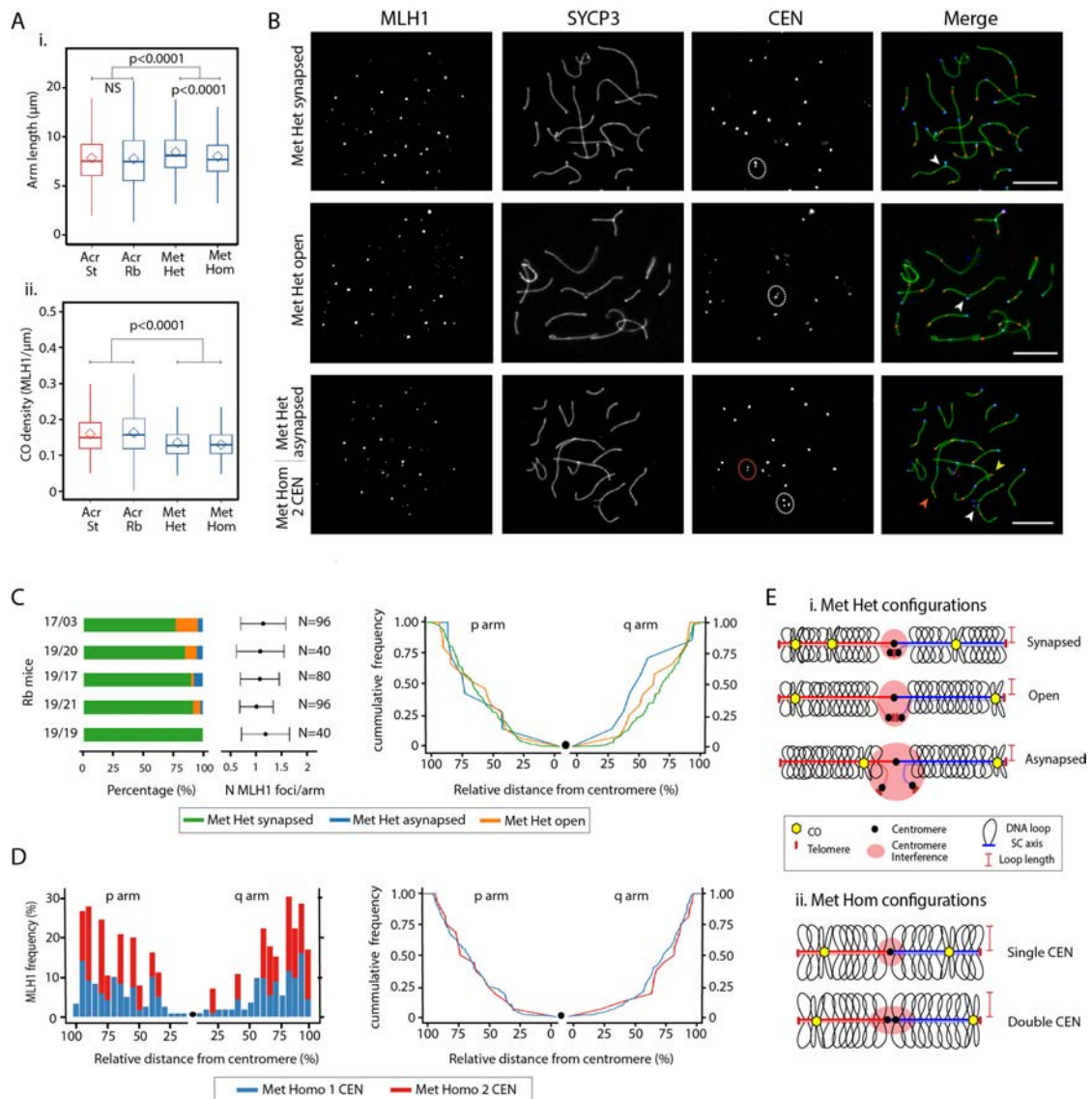
Figure 4.11: Effects of Rb fusions on recombination in the synaptonemal complex. (A) (i) Axis length analysis in standard (St) and Robertsonian (Rb) mice from BRbS (see text and Table S1 for further details) according to chromosome types (Dunn's test, p<0.001; NS: non-significant). (ii) Analysis of CO density in the different chromosome types (Mann-Whitney test, p<0.0001). **(B)** Immunofluorescence of primary spermatocytes at pachynema stage, labelling the synaptonemal complex with SYCP3 (green), the centromeres with CEN (blue) and MLH1 (red). White dashed circles: centromeric signals in heterozygous fusions. Red dashed: double centromere signals in homozygous fusions. White arrowheads: Met Het chromosomes. Red arrowhead: Met Hom with double centromeric signals. Yellow arrowhead: Met Hom with a single centromeric signal. Scale bar= 10 μm. **(C)** Synapsis and recombination patterns found in mice with Rb fusions in heterozygosis (see Table S1 for mice codes). Left panel: Percentage of heterozygous chromosomes according to synapsis pattern per each of the Rb mice analyzed (see Table S1 for further details). Middle panel: Representation of the mean number of MLH per arm corresponding to the analyzed mice. N = number of cells analyzed per individual. Right panel: Cumulative frequencies of MLH1 distributions in heterozygous metacentrics in different synaptic states. **(D)** Double centromeric signals and recombination. Left panel: Distribution of MLH1 foci along individual chromosomal arms in homozygous metacentrics with a single (blue) or double (red) centromeric signal. The X-axis represents the positions on the chromosomal axes from the centromeric end (black dot) to the distal telomere. The Y-axis indicates the frequency of MLH1 foci in each 10% interval of chromosomal length. Right panel: Cumulative frequencies of MLH1 distributions in homozygous metacentrics with a single or a double centromeric signal. **(E)** Diagram depicting chromosomal axis (SC) length, DNA loops and recombination patterns observed in fused chromosomes at pachytene in Rb mice. Met Het can present three states: synapsed, open and asynapsed. COs are closer to the centromere in asynapsed chromosomes. Open chromosomes present distal COs, whereas synapsed chromosomes can present more than two COs per arm and in interstitial to distal positions. Regardless of the synapsis state, Met Het chromosomes present with longer axes than Met Hom. Conversely, DNA loop lengths in Met Het are shorter than Met Hom.

length) or in distal regions (80-90% of the chromosome length) (**Figure 4.11 C and Figure S3 C**). However, synaptic disturbances in heterozygous chromosomes (**Figures 4.11 and 4.13**) did not

significantly reduce the number of COs per arm when compared to acrocentrics (**Figure 4.10 F**), most probably due to variation in chromosomal axis length (**Figure** 4.11 A). Interestingly, we observed that 28.5% of the CO events detected in asynapsed chromosomes where located at the border of synapsed and unsynapsed regions (30% of the chromosome length, **Figure S3 C**). These COs could prevent further asynapsis, as previously suggested for hexavalent meiotic configurations in Rb mice (Ribagorda et al. 2019).

Both the reduced number of COs per arm and the displacement away form the centromere observed in homozygous fusions compared to heterozygous (**Figure 4.10 E and Figure S1 F**) was unexpected based on the chromosomal speciation theory (Rieseberg 2001; White 1978), thus we sought to understand the mechanistic factors behind this pattern. Remarkably, we detected that homozygous chromosomal fusions were associated with the presence of double centromeric signals at variable frequency among individuals (from 13.33% to 39.02%, **Figures 4.11 D and Figure S3 D**). We then investigated whether these double centromeric signals correlated with CO chromosomal distribution by analyzing mice from Viladecans with the same number of fusions (4 fusions in homozygosis and one fusion in heterozygosis). Notwithstanding cell-to-cell variability, the presence of a double centromere signal was associated with displacement of COs towards distal regions of the chromosome (**Figure 4.11 D**). This extended centromeric recombination suppression, coupled with shorter chromosomal axes, likely explain the significant reduction of COs per arm on homozygous Rb chromosomes when compared to heterozygous (**Figure 4.10 E and Figure S1 F**).

Overall, we observed that Rb fusions had an effect on CO chromosomal distribution and chromosomal axis length depending on if found in a heterozygous or homozygous state. With consistent genome size, the longer chromosomal axes in heterozygous fusions would be attributed to shorter loops (**Figure 4.11 E**). Despite the presence of different levels of asynapsis, heterozygous fusions present longer axes allowing the formation of the obligatory CO per arm necessary for faithful chromosomal segregation. The contrasting pattern found in homozygous fusions (short chromosomal axis, high frequency of arms with zero CO and centromere interference) suggests that chromosome architecture could play a major role in reshaping CO distribution.

4.3.3.5 Chromosomal fusions reorganize spatial chromosome occupancy

We then analyzed whether Rb fusions impact the three-dimensional genome folding in germ cells. Using FACS, we isolated highly enriched meiotic (primary spermatocytes at pachytene/diplotene stage, P/D) and post-meiotic cells (round spermatids, RS) from wild-caught Rb mice (derived from the same population) with a high number of fusions (**Table S1, Figures S4 A and B**). For each germ cell fraction we performed in situ Hi-C (Rao et al. 2014) to generate genome-wide Hi-C maps for primary spermatocytes and round spermatids. A Rb somatic profile was derived from a wild-caught Rb male primary fibroblast cell line. Germ cell and somatic data for all-acrocentric (standard) mice was previously published (Vara et al. 2019b). After filtering, an average of 254.9 million valid interactions per cell type were detected (**Tables S5 and S6**) with high correlation values between biological replicates (from 0.96 to 0.92) (**Figure S4 C**).

Genome-wide interaction maps confirmed the presence of six Rb fusions (3.8, 4.10, 5.15, 6.10, 9.11 and 12.13) in all cell types analyzed from Rb mice (**Figure 4.12 A, Table S1**). All chromosomes involved in Rb fusions presented higher inter-chromosomal interaction values (interaction ratio 3.0) in Rb mice, compared to standard mice, in all cell types (**Figure 4.12 B, Table S1**). However, we also detected different patterns of inter-chromosomal interaction ratios between non-fused chromosomes. In fibroblasts few non-fused chromosomes (e.g., chromosomes 1 and 2) had higher values of heterologous interactions in Rb mice when compared to standard mice (**Figure 4.12 B and C**). However, the spatial genome architecture was highly reorganized in P/D, affecting nearly all chromosomes (fused and non-fused) (**Figure 4.12 B**).

In P/D all chromosomes showed high values of heterologous interactions in Rb mice (interaction ratio $\leqslant$ 1.5) for all chromosomes when compared to standard mice (**Figure 4.12 B**), suggesting a

genome-wide re-distribution of the spatial disposition of chromosomes inside nuclei. The presence of heterologous associations was further demonstrated by the immunodetection of the centromeric constitutive heterochromatin (exemplified as H3K9Me3 signals) on pachytene bivalents in a subset of Rb mice (**Figure 4.13 A**). In fact, heterologous interactions were dependent on Rb fusions pairing states (i.e., fully synapsed, open or asynapsed) (**Figure 4.13 B**). Specifically, we observed more centromeric associations of acrocentric chromosomes when Rb fusions were fully synapsed (**Figure 4.13 C**). Additionally, the sex body (the X and the Y chromosome) was generally isolated from autosomes (**Figure 4.13 B**). When heterozygous Rb fusions failed to fully synapse, there was a general disruption of inter-chromosome associations (**Figure 4.13 C**) and the sex body presented abnormal signals of heterochromatinization being frequently associated with fused chromosomes (**Figure 4.13 B**). Interestingly, we detected full heterochromatinization of the sex body in more than 90% of cells with asynapsed heterozygous Rb fusions (**Figure 4.13 B**), which could contribute to the increased X/autosome heterologous interactions in P/D Rb mice (**Figure 4.12 E**). In contrast, heterologous interactions were reduced in RS for all chromosomes not involved in fusions (ratios <1, **Figure 4.12 B**), a pattern confirmed by the analysis of inter-/intra chromosomal interactions (**Figure 4.12 C**). In fibroblasts and P/D Rb mice showed higher inter-/intra chromosomal interaction ratios than all-acrocentric mice (Mann-Whitney test, $p<2.2e\text{-}16$), a pattern that was clearly reversed in RS with higher interaction ratios in standard mice than in Rb mice for all chromosomes (Mann-Whitney test, $p<2.2e\text{-}16$) (**Figure 4.12 C**).

Chromosome-specific interaction maps also revealed altered three-dimensional chromosome folding in Rb mice, especially in prophase I (**Figure 4.12 D**). For non-fused chromosomes (e.g., chromosomes 1 and 2, which represented a pattern consistent across all chromosomes) interaction patterns were disrupted during P/D in Rb mice (**Figure 4.12 D**), losing the inter-chromosome interaction pattern previously described for standard mice (Patel et al. 2019; Vara et al. 2019b). These results suggest the presence of genome-wide conformational changes triggered by the presence of Rb fusions, which was supported by the cytological analysis of chromosome associations (**Figure 4.13 C**). For fused chromosomes (e.g., chromosomes 3 and 8, being the pattern consistent across all chromosomes), we not only detected changes in mid-chromosome interactions, but also high interaction values in pericentromeric regions (**Figure 4.12 D and F**).

Overall, our data suggests that the presence of Rb fusions reorganize chromosomal nuclear occupancy genome-wide (**Figure 4.13 D and E**), increasing the rate of heterologous interactions in primary spermatocytes. Moreover, our results suggest that pairing disturbances of heterozygous fusions during prophase I have a direct impact in chromosome spatial distribution (**Figure 4.13 E**). As such, when heterozygous fusions are synapsed, more permissiveness for centromeric associations between acrocentric chromosomes is observed and the sex body is generally isolated at the periphery of nuclei. Conversely, when heterozygous Rb fusions fail to fully synapse, there is a general disruption of inter-chromosome associations and the sex body present abnormal signals of heterochromatization and it is frequently associated with fused chromosomes (**Figure 4.13 E**).
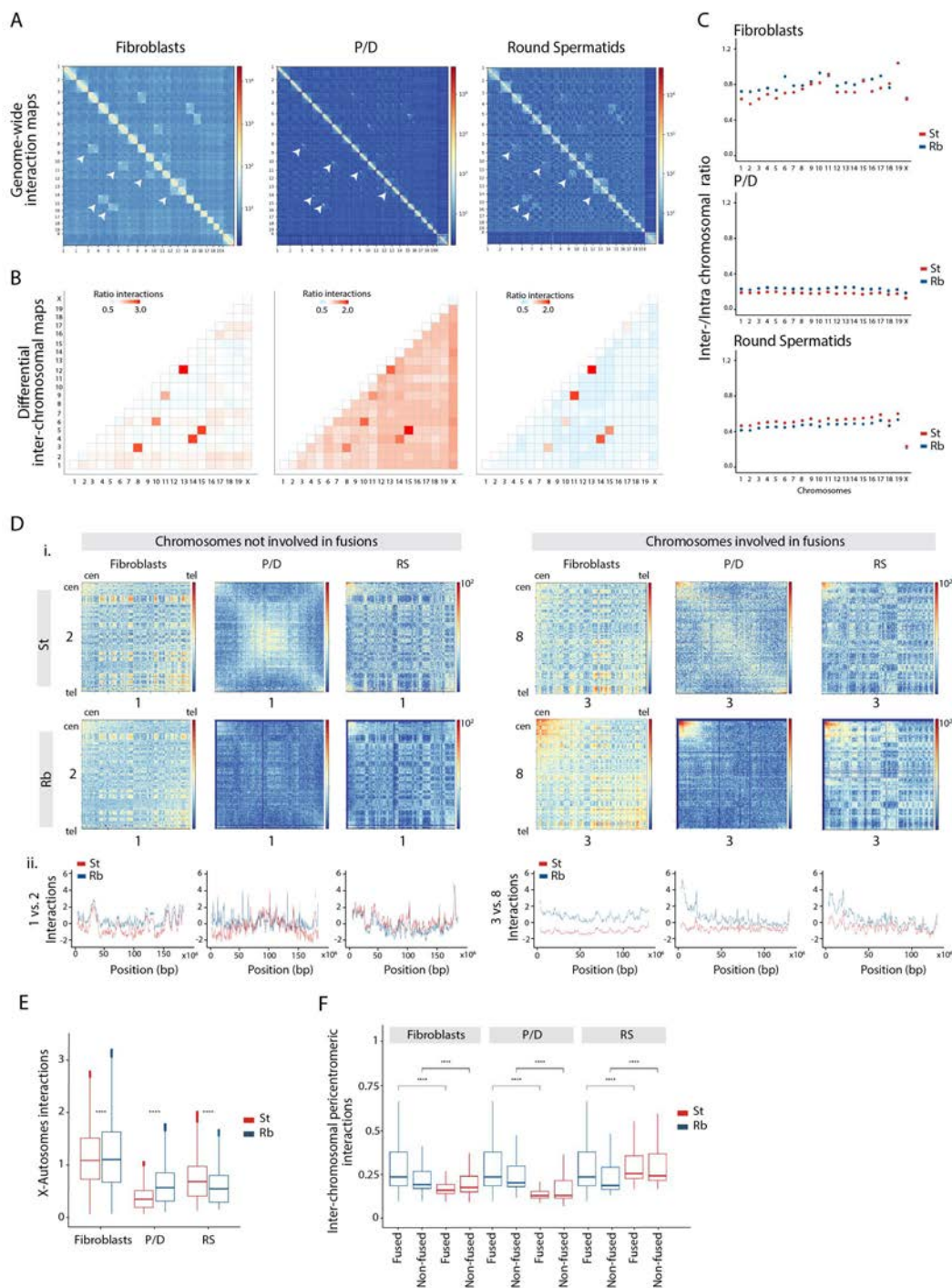
Figure 4.12: Effect of Rb fusions in the higher-order chromatin structure. (A) Genome-wide ICE-corrected heatmaps (500kbp) for the cell types analyzed (fibroblasts, pachynema/diplonema – P/D and round spermatids – RS) in Rb mice. Chromosomes involved in Rb fusions emerge as regions with high inter-chromosomal interaction in all cell types (arrow heads). These are the following: 3.8, 4.14, 5.15, 6.10, 9.11 and 12.13. **(B)** Differential heatmaps depicting genome-wide inter-chromosomal interaction ratio between standard and Rb mice. Chromosomes in red indicate higher interactions in Rb than in standard mice whereas chromosomes in blue indicate higher interactions in standard than in Rb mice. As expected, all chromosomes involved in fusions present high interaction ratios in all the three cell types analyzed. **(C)** Inter-/Intra-chromosome interaction ratio for the cell types analyzed (fibroblasts, pachynema/diplonema – P/D and round spermatids – RS) in standard (St) and Rb mice. In fibroblast (upper panel) values are the same for chromosome 19 in both St and Rb mice. Continues in the next page.

Figure 4.12: (From previous page). **(D)** Interaction patterns between chromosomes involved and not involved in fusions. **(i)** Interaction heatmaps representing chromosomes 1 and 2 (not involved in fusions) and chromosomes 3 and 8 (fused in heterozygosis) in both standard mice and Rb mice. In chromosomes 3 and 8, the fusion becomes evident in interaction maps from Rb mice, with high interaction in the pericentromeric region of the chromosomes (0 to 3 Mbps from the centromere). The observed scaling is consistent across all chromosomes. **(ii)** Interaction plots for fibroblasts, P/D and RS for chromosomes not involved (1 and 2) and involved in fusions (3 and 8). The observed scaling is consistent across all chromosomes. **(E)** Boxplot showing the number of interactions between the X chromosome and autosomes detected in standard and Rb mice per each cell type (Mann-Whitney test, ****p<0.0001). **(F)** Boxplots depicting inter-chromosomal interactions per million at pericentromeric regions (from the centromere up to 3.5 Mbp) between standard and Rb mice (Mann-Whitney test, ****p<0.0001). For each cell type two groups of chromosomes were compared: chromosomes involved in Rb fusions (3.8; 4.14; 5.15; 6.10; 9.11 and 12.13) and chromosomes not fused (1, 2, 7, 10, 16, 17, 18, 19 and X).
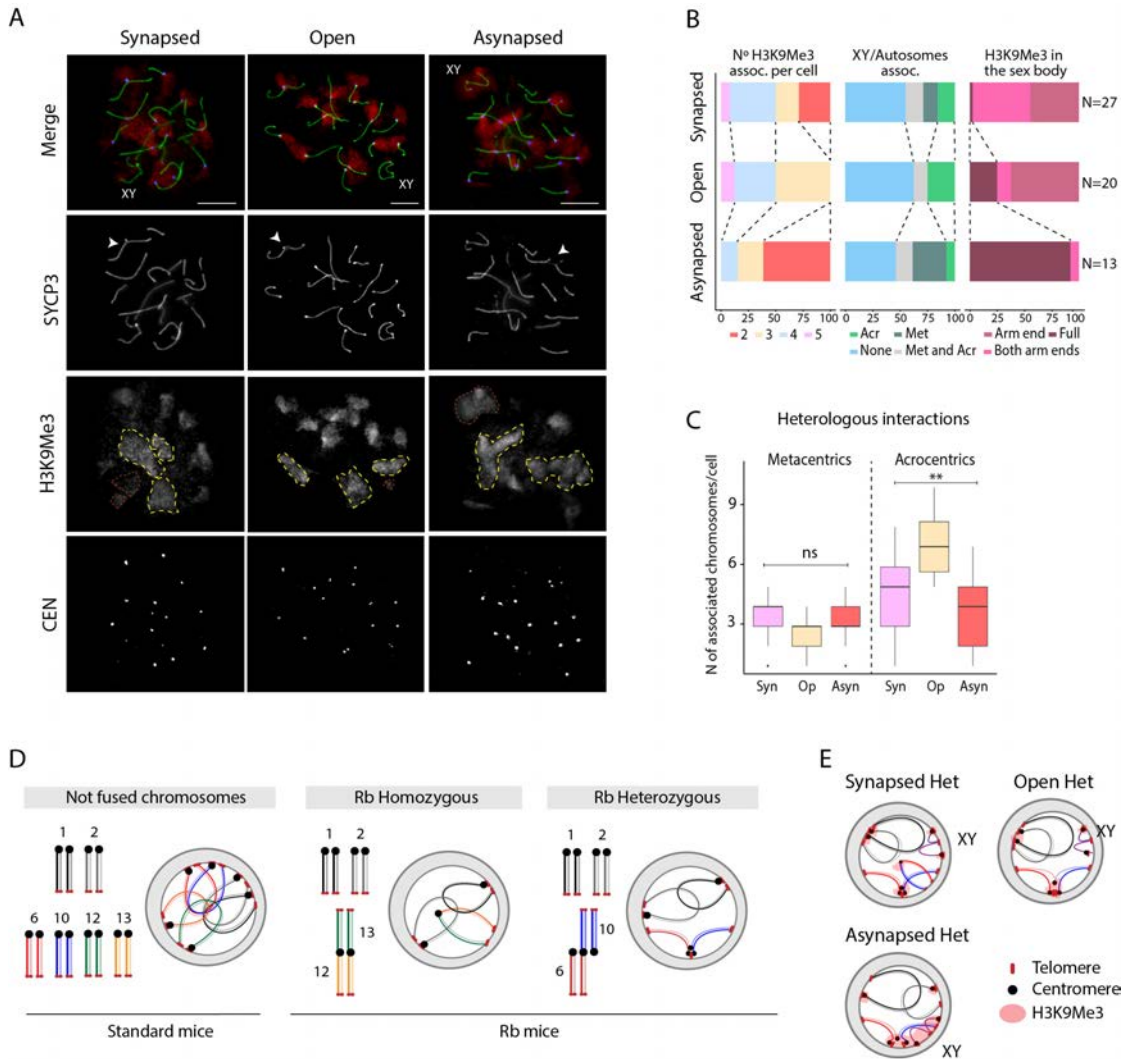
Figure 4.13: Inter-chromosomal associations. (A) Examples of immunofluorescence on primary spermatocytes at pachytene, labelling the synaptonemal complex with SYCP3 (green), the centromeres (CEN, blue) and H3K9Me3 (red). SYCP3 staining allowed for the detection of the different heterozygous Rb fusion states: synapsed, open and asynapsed (white arrows). H3K9Me3 shows associations between different chromosomes (yellow dashed lines) and differential distribution in XY (orange dashed lines): in both arms (as shown in the synapsed example), in one arm end (as shown in the open example) or fully around the sex body (as shown in the asynapsed example). The sex body is indicated as XY. Scale bar=10 μm. **(B)** Analysis of H3K9Me3 associations according to heterozygous Rb fusion states (synapsed, open and asynapsed) (left panel), sex chromosomes/autosomes associations (central panel) and only the sex body (right panel). Dashed black lines indicate exemplary differences in the proportion of analyzed cells. **(C)** Number of associated chromosomes (metacentrics or acrocentrics) detected per cell depending on the heterozygosis Rb fusion state (Kruskal-Wallis, $p \leqslant 0.005$). **(D)** Schematic representation of chromosome organization in P/D according to the presence of Rb fusions. In standard mice, all chromosomes are acrocentric and are attached to the nuclear lamina. When Rb fusions are present, chromosome organization is disrupted, affecting chromosome disposition inside the nucleus, either in homozygous or heterozygous state. **(E)** Schematic representation of the centromeric associations detected with the H3K9Me3 signal in addition to the XY disposition according to heterozygous Rb fusion states (i.e., synapsed, asynapsed or open).

4.3.3.6 Higher-order chromatin remodeling result from chromosomal fusions

We further investigated whether the presence of Rb fusions had an effect on chromatin remodeling at the finer scale (e.g., compartments). In somatic cells, the genome-wide analysis of differential Hi-C matrices showed higher interactions at shorter genomic distances in standard mice than in Rb mice (**Figure 4.14 A**). This interaction pattern, which is dependent on genomic distance, was switched at genomic distances of 6 Mbp in fibroblast (**Figure 4.14 B**). This was concomitant with slight changes in A/B compartments (**Figure 4.14 C and D**), with Rb mice presenting smaller mean compartment size (0.85 Mbp) than standard mice (1 Mbp).

Mirroring previous studies (Patel et al. 2019; Vara et al. 2019b), our genome-wide analysis showed that most compartments were mostly lost in primary spermatocytes (P/D) in Rb mice (**Figure 4.14 C**), when homologous chromosomes condense, align, pair, synapse and recombine. Consistent with this absence of compartments during prophase I, eigenvector values were close to 0 (**Figure 4.14 C**) and inter-/intra-chromosomal interaction ratios reached a minimum for all chromosomes (**Figure 4.12 C**). Importantly, the analysis of differential Hi-C matrices and intra-chromosomal interaction ratios revealed that, on average, probabilities of interactions were higher in Rb mice at genomic distances larger than 10 Mbp (**Figure 4.14 B**). Local adjustments in chromatin packing density due to differences in loop positioning and size, together with variations in chromosomal axis length (as revealed by our cytological observations), likely give rise to the differences between standard and Rb mice Hi-C maps.

Robertsonian mice presented a distinctive pattern in post-meiotic cells, with high interactions at shorter genomic distances ($\leqslant$5 Mbp) (**Figure 4.14 A and B**). The effect of highly condensed chromatin in round spermatids of Rb mice would increase short-range contacts (relative to long-range contacts), as we observe. Although A/B compartments re-appeared in RS, they were present as a blurry plaid pattern of larger mean size (0.93 Mbps in Rb mice) than in fibroblasts. Interestingly, the proportion of genomic bins with the same compartment status (A or B) was higher between standard and Rb mice in RS (95% bins conserved, r^2=0.94) than in fibroblasts (88.24% bins conserved, r^2=0.80) (**Figure 4.14 D**). These results, together with the low inter-/intra-chromosomal interaction values detected in Rb mice (**Figure 4.12 C**), suggest that Rb fusions induce physical constrains in RS most likely due to the spatial organization of chromosomes around the chromocenter (Brinkley et al. 1986; Hoyer-Fender et al. 2000), thus favoring local interactions.

4.3.3.7 Chromosomal fusions induce TAD reorganization

Chromosome fusions also affected chromatin remodeling at the sub-megabase scale as revealed by the number of TADs detected and the robustness of their boundaries at a 50Kbp resolution (**Figure 4.14 E-I and Figure S5**). Similar to previous observations in standard somatic cells (Vara et al. 2019b), TADs were well defined in Rb fibroblasts (**Figure 4.14 E-F**), with a total of 2,391 detected. With an average length of 1.14 Mbps, TADs in Rb fibroblasts had a higher variance of insulation scores than in standard fibroblasts (**Figure 4.14 E**). Nevertheless, the majority (70%) of TADs were stable (**Figure 4.14 H**), with general preservation of border conformation, as revealed by the TAD meta-border plot (**Figure 4.14 I**) that shows the specific border interactions of loop domain TADs (Barrington et al. 2019; Rao et al. 2014).

In primary spermatocytes there was a substantial reduction in the variance of TAD insulation score, and reduction of total TADs detected, when compared to somatic cells (288 in Rb and 294 in standard). The TAD insulation score variance was higher in Rb mice (**Figure 4.14 E**). Moreover, meta-border plots show asymetric TAD borders (**Figure 4.14 I**), more characteristic of stripe TAD domains (Barrington et al. 2019; Vian et al. 2018). Variance of TAD insulation scores was partially recovered in RS, being higher in Rb mice than standard. In RS, Rb mice had fewer but larger (n=3,805, 0.71 Mbp average length) TADs than standards (n=4,649, 0.59 Mbp average length). Only 40% of TADs were stable in RS (**Figure 4.14 H**). The presence of weaker TAD borders in both standard

and Rb RS (mean TAD insulator score=5.46) when compared to fibroblasts (mean TAD insulator score=8.8), could explain the difference in TADs observed between these cell types, with TADs more prone to reorganize in RS (**Figure 4.14 F-G**). In fact, rearranged, split and merged TADs presented significantly lower TAD border scores than stable TADs (Mann-Whitney test, p=2e-14). Meta-border plots showed a blurred pattern, consistent with loop domain TAD borders and weak TAD insulator scores (**Figure 4.14 I**).
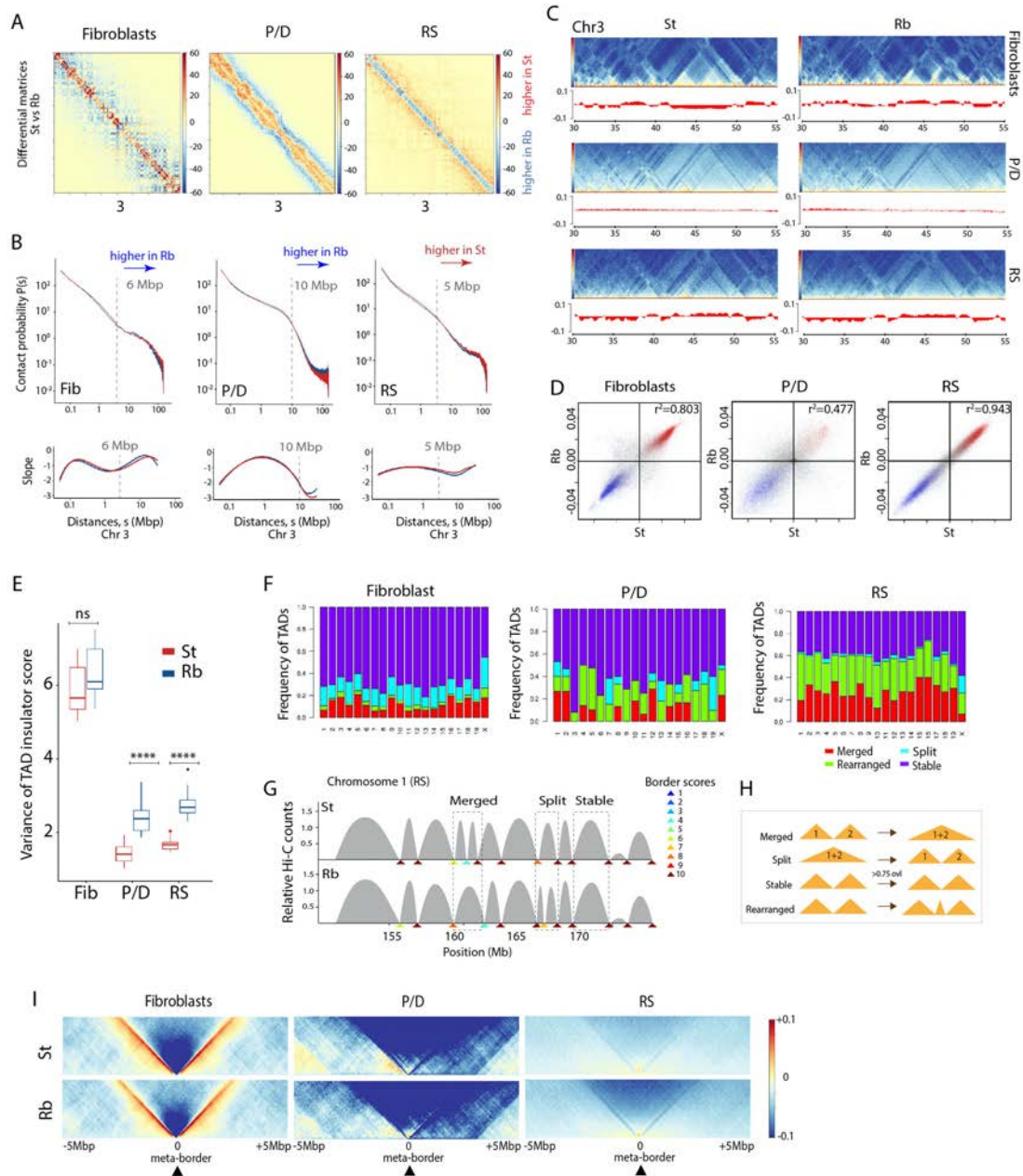
Figure 4.14: Variance in fine-scale compartmentalization. (A) Differential ICE-corrected Hi-C matrices (log2 of fold change using Rb mice as a reference when compared with standards) for chromosome 3 in fibroblasts, P/D an RS, at a 50 kbp resolution. The observed scaling is consistent across all chromosomes. Red indicates higher number of interactions in standard mice when compared to Rb whereas blue represents higher number of interactions in Rb. **(B)** Contact probability P(s) as a function of genomic distance in all cell types for chromosome 3. The observed scaling is consistent across all chromosomes. Dashed purple line delineates the interactions pattern shift between St and Rb mice. Lower panel: Plot of the slope of the P(s) curves shown above with the genomic position (gray line) of pattern shift. **(C)** Chromosome 3 region-specific ICE-corrected heatmaps at 50 kbp (from 30 to 55 Mbp), depicting compartment signal (1st eigenvector) for all cell types. Continues in the next page

Figure 4.14: (From previous page). **(D)** Conservation of A/B compartments between standard and Rb mice. Genome-wide pair-wise representation of eigenvectors (including all chromosomes) in the three cell types analyzed. Each dot represents a 50Kbp bin in the genome. Bins representing A compartment conservation are depicted in red, whereas in blue are depicted bins with B compartment conservation. Bins with unclear signal or compartment switching are represented in grey. Correlation values are represented for each pair-wise comparison (Spearman, p<0.0001). **(E)** Variance of TAD insulator score between standard and Rb mice in all cell types (Mann-Whitney test, ****$p \leqslant 0.0001$, ns p>0.05). **(F)** Frequency of TAD reorganizations between standard and Rb for fibroblasts, P/D and RS. **(G)** Example of TAD border alignments along chromosome 1 of round spermatids (from 150 to 175 Mbp). Examples of merged, split and stable TADs are indicated. TAD border scores are also shown, informing of the TAD boundary strength. **(H)** Schematic representation of TAD reorganization. Merged TADs are the result of fusing two different TADs. Split TADs are the those in which one TAD is divided into two TADs. TADs are considered stable when there is an overlap above 75%. When TADs are found in different organization they are considered rearranged. **(I)** Meta-plots for all TAD boundaries detected in Rb mice: fibroblasts (n=2,378), P/D (n=288) and RS (n=3,798). Data on standard mice was extracted from Vara and colleagues (Vara et al. 2019b).

4.3.3.8 Olfactory receptor genes enrichment in meiotic-specific inter-chromosomal interactions

The increased rate of heterologous interactions detected in Rb primary spermatocytes resulted in the emergence of new inter-chromosomal interactions (n=249, representing 0.5% of the mouse genome) that involved all chromosomes (e.g., chromosomes 2, 3 and 7, **Figure 4.15 A**). To explore the potential role of Rb-specific inter-chromosomal interactions we analyzed the gene ontology of the 2,000 genes contained in these regions. Importantly, we detected enrichment for sensory perception genes (**Figure 4.15 B**), specifically, olfactory receptors (OR) (n=118) and vomeronasal receptors (VR) (n=92). These OR/VR genes with heterologous interaction were located in 11 out of the 94 OR clusters previously described in the mouse genome (Degl'Innocenti et al. 2019) (**Figure 4.15 C**). These included mouse chromosomes 4, 6, 7, 9, 10, 14, 16, 18 and 19. Interestingly, we detected that individual OR clusters interacted with a wide range of regions from different chromosomes, being the interactions either chromosome-specific (e.g. the cluster on chromosome 19 interacts solely with chromosomes 3) or multi-chromosomal (e.g. interactions between two clusters on chromosome 10 with multiple regions of chromosomes 1, 2, 3, 4, 7, 8, and 9) (**Figure 4.15 D**). Furthermore, we detected significantly more repeats in OR gene clusters when compared with the rest of the mouse genome (**Figure S6 A**) (Wilcoxon test, p≤.0001), particularly LINE/L1, and LTR/ERV family retrotransposons (**Figure S6 B**).
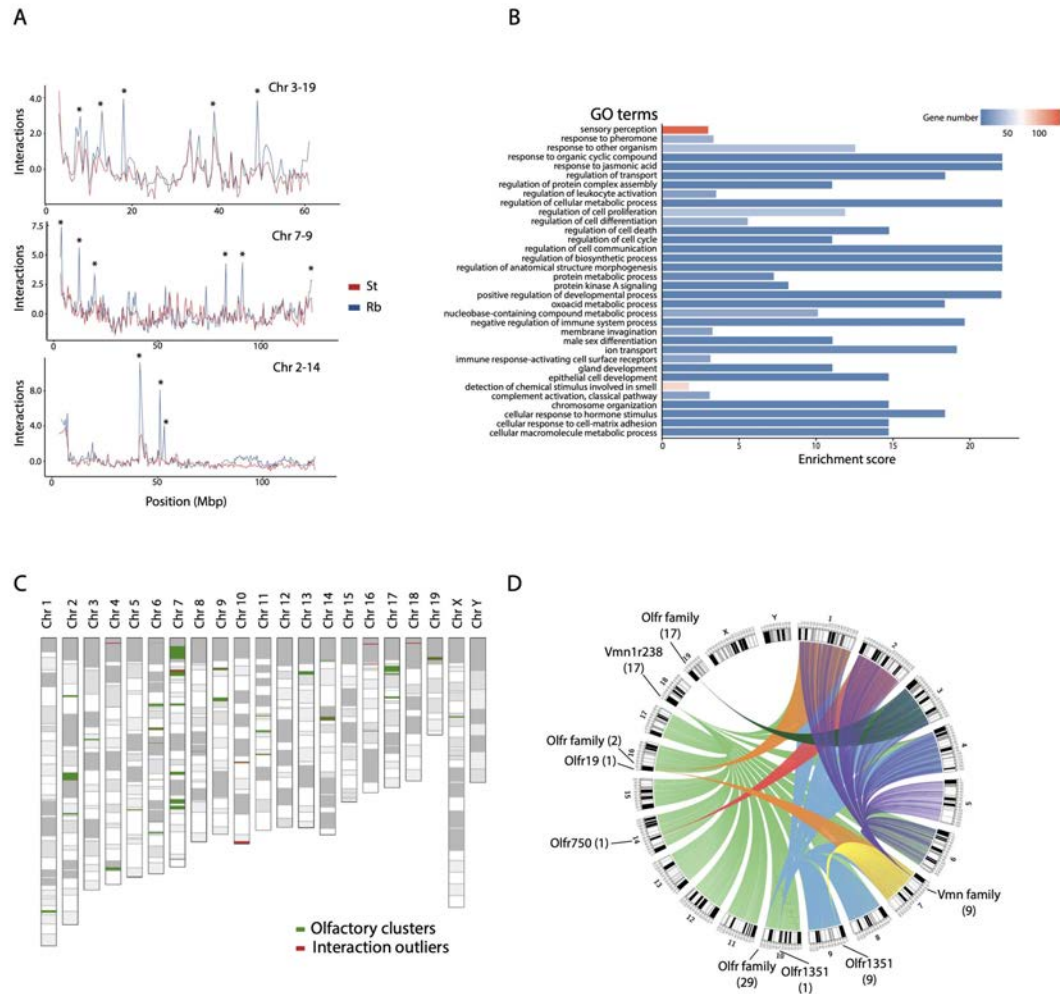
Figure 4.15: Inter-chromosomal interaction and olfactory receptors.(A) Interaction profiles of pairs of chromosomes in Rb P/D (3 and 19, 7 and 9, 2 and 14). In all the three examples inter-chromosomal interactions are depicted by asterisks. **(B)** Gene ontology analysis of genes in newly detected inter-chromosomal interactions in Rb P/D. Data shows that the inter-chromosomal interactions regions are significantly enriched with sensory perception. **(C)** Mouse ideogram showing the localization of olfactory clusters (green) described in literature (Degl'Innocenti et al. 2019) and the inter-chromosomal interactions detected in this work (red). **(D)** Circus plot representing the inter-chromosomal interactions related with sensory perception, which are mostly genes from the olfactory (Olfr) and vomeronasal (Vmn) receptor family. The number of genes found in each region and the gene family are shown.

4.3.4 Discussion

Here we provide evidence that chromosomal fusions impact three-dimensional genome topology and meiotic recombination, highlighting the implications of large-scale genome reorganizations on genome function and fertility. Our results show that chromosomal fusions pose important mechanistic constrains in the nuclear architecture of germ cells, affecting heterologous interactions, chromosomal synapsis and meiotic recombination. This was reflected at different hierarchical levels: (i) chromosomal nuclear occupancy, (ii) inter- and intra-chromosomal interactions, (iii) chromosomal axis length, (iv) chromatin loop size and (v) topological domains.

Our high-throughput analyses in combination with cytological observations show that disturbances in nuclear chromosomal occupancy occur genome-wide, affecting not only chromosomes involved in Rb fusions but also not-fused chromosomes. The pattern observed in primary spermatocytes is especially relevant as fundamental cellular processes take place, such as synapsis and pairing of homologous chromosomes and the formation and repair of DSBs. It is also in primary spermatocytes

where DNA loops are organized along the chromosomal axis, with loop size and axis length inversely correlated (Ruiz-Herrera et al. 2017; Wang et al. 2015b; Zickler and Kleckner 2015), which can co-vary within gametes (Wang et al. 2019a). Although we can only speculate on the mechanisms behind the observed patterns, Rb fusions alter the distribution of telomeres and centromeres attached to the nuclear lamina, ultimately inducing ectopic heterologous interactions. In standard mice all chromosomes are acrocentric, so telomeres and centromeres are attached to the nuclear lamina, globally disposed during prophase I. But the presence of Rb chromosomes poses mechanistic constrains with centromeres locating either towards the center of the nucleus or at the periphery, as suggested by previous microscopic observations (Berríos et al. 2014). Such alteration in nuclear occupancy, together with the pairing disturbances observed with fusions in heterozygous state would trigger stochastic physical interactions and expose chromosomal domains to novel regulatory environments (i.e., OR gene clusters). Remarkably, this disruptive pattern also affected chromosome axis length and, as a result, higher-order chromatin remodeling. Because chromatin is organized into DNA loops that emerge out of meiotic chromosomal axes (Zickler and Kleckner 1999), variations in axis length alter the number and size of these loops (Wang et al. 2019a), which impacts medium and long-range interactions. Likewise, because the assembly of chromatin loops into chromosomal axes, and the formation and repair of meiotic DSBs (and subsequent COs) are interconnected (Capilla et al. 2016; Ruiz-Herrera et al. 2017; Wang et al. 2019a), the remodeling of chromosomal interactions translated into a reorganized recombination landscape in Rb mice. Thus, Rb fusions pose mechanical constrains on the spatial genome architecture that affect not only the hierarchical three-dimensional organization of the genome, but also the distribution of CO events genome-wide. The molecular mechanisms behind this phenomenon remain unknown.

Importantly, the re-distribution of COs across chromosomal arms was consistent with observation of low recombination rates at a fine scale (Kbp), and high values of genomic divergence. This provides novel evidence that disturbances in CO distribution due to chromosome fusions result in detectable genomic footprints. According to the 'suppressed recombination' model (Faria and Navarro 2010; Farré et al. 2013) a reduction in recombination is expected within reorganized regions in the heterokaryotype (e.g. heterozygous Rb fusions). Although this was consistent with our observations (the proportion of chromosome arms with zero CO was slightly higher in heterozygous Rb fusions than non-fused chromosomes), we also detected an even stronger reduction of recombination in homozygous Rb fused chromosomes. Decreased recombination in homozygous Rb fusions is counterintuitive, but could be explained by the presence of double centromeric signals that represented misaligned centromeres (Borodin et al. 2008; Cappelletti et al. 2019). Since centromeres can reduce recombination rates [so-called 'centromere interference', (Lynn et al. 2004)], the presence of double centromeres could magnify this effect, interfering with both the formation of COs during prophase and subsequently, normal chromatid segregation in metaphase I. In fact, previous studies in BRbS reported heterochromatinization disturbances at the centromeres of fused chromosomes (Capilla et al. 2014), and a higher frequency of apoptotic spermatogenetic cells in homozygous mice than in heterozygous mice (Medarde et al. 2015). In this context, Rb fusions in homozygosis could present high centromere interference responsible for CO reduction.

Remarkably, we also detected that the frequency of chromosome arms with zero COs was not significantly altered in mice with Rb in heterozygosis, but chromosomal axes were longer when compared to non-fused chromosomes, thus affecting CO density. Conversely, the cytological data suggested the presence of shorter chromosomal axes in homozygous Rb fusions (and therefore, longer DNA loops). This was consistent with the Hi-C data, suggesting that Rb fusions induce variance in higher-order chromatin organization. This adds to initial reports on global modulation of chromosomal CO frequencies by both chromatin loop size and chromosomal axis length (Ruiz-Herrera et al. 2017; Wang et al. 2019a, 2015b), further suggesting that recombination landscapes can be altered within cells by Rb fusions affecting both the loop/axes length ratio and the spatial chromosome occupancy.

Likewise, Rb fusions not only pose restrictions for chromosomal interactions in primary spermato-cytes but also in post-meiotic cells. Chromosomes are highly condensed in round spermatids with

all centromeres associated in the center of the cell forming the chromocenter (Vara et al. 2019b). As heterologous contacts were reduced in Rb mice we hypothesize that chromosomal fusions restrict interactions between non-fused chromosomes in round spermatids. This restricted and condensed pattern favors intra-chromosomal contacts, resulting in reorganization of TADs in post-meiotic cells.

Our observations have important evolutionary and developmental implications. The dynamic three-dimensional genome topology of germ cells can be affected by chromosomal fusions in two ways: (i) altering chromosomal nuclear occupancy, and (i) reshaping landscapes of recombination. The redistribution of chromosomal nuclear occupancy in spermatocytes that result from Rb fusions brings new genomic regions into close proximity that are distanced in standard mice. In the case of acrocentric chromosomes, heterologous association of pericentromeric regions can predispose to the occurrence of new Rb fusions if DSBs occur and are ill repaired. As such, changes in chromosomal nuclear occupancy could explain the occurrence of Rb fusions detected not only in natural populations of rodents (Matveevsky et al. 2020; Piálek et al. 2005), but also in other mammalian species like shrews (Fedyk et al. 2019). Moreover, the rearranged nuclear architecture in Rb mice would expose chromosomal domains to novel regulatory environments, potentially affecting gene expression and/or regulation, as initially proposed by the integrative breakage model of genome architecture (Deakin et al. 2019; Farré et al. 2015). New chromosomal interactions resulting from chromosomal fusions may rewire or attenuate gene networks, providing new grounds for evolutionary novelty in the long round. This appears the case of olfactory receptor family clusters detected in meiotic-specific inter-chromosomal interactions in Rb mice. As this gene family is expressed in the male germ line (Vara et al. 2019b), with a function in spermatogenesis and fertilization (Fedyk et al. 2019; Flegel et al. 2016; Matveevsky et al. 2020), altered regulation of their expression could have an adaptive role. In fact, inter-chromosomal interactions between OR genes have been previously described in sensory neurons (Monahan et al. 2019), suggesting that associations between multiple chromosomes can selectively regulate transcription of individual OR genes. Further functional studies will be needed to support this hypothesis in the germ line.

Importantly, disturbances of CO distributions can be regulated individually at the chromosome level to allow proper disjunction of chromosomes during meiosis. Under this scenario, a decrease in meiotic recombination due to the presence Rb fusions might induce subfertility, as revealed by our observations of significant differences in sperm motility and vitality between Rb and standard mice. This, together with previous reports on variation of the sperm head morphology and spermatogenic activity in Rb mice (Medarde et al. 2013; Sans-Fuentes et al. 2010), suggest that Rb fusions have an effect on mouse fertility, although mild enough to allow their persistence within populations. In this context, our observations of topological disturbances in prophase I provides another layer of complexity, setting the grounds for exploring the mechanisms responsible for the origin and maintenance of balanced chromosomal fusions in the germ line.

4.3.5 Conclusions

We demonstrate that chromosomal fusions affect the three-dimensional genome-wide topology in germ cells, ultimately reshaping recombination landscapes. The modulation of recombination implies a close interplay between different factors that are involved in chromatin remodeling, centromere interference, DNA loop size and chromosomal axis length. Understanding how such changes in genome organization affect gene expression and regulation is an important dimension of further understanding the effect of genome reshuffling on evolution and fertility. We anticipate that our results will provide impetus for exploration of the functional and structural basis of genomes in a broad context, reinforcing the link between the three-dimensional genome architecture, genome integrity and fertility.

4.3.6 Methods

Animals and cell lines. We sampled a total of 63 wild-caught house mice (*M. m. domesticus*) from 6 populations, covering the extent of the BRbS (**Figure 4.10 A and Table S1**). All animals included in the study were previously karyotyped (Capilla et al. 2014; Medarde et al. 2012; Sánchez-Guillén et al. 2015) confirming the presence of Rb fusions (**Table S1**). Three males from the laboratory strain C57BL6 (BL6) were also included in the recombination analysis. Animals were housed and treated in strict accordance with ethical guidelines approved by the Universitat Autònoma de Barcelona (Spain).

The BRbS system mice included in the present study consisted of three populations without Rb fusions (2n=40; Caldes de Montbui, Castellfollit del Boix and Olost) and three populations with Rb fusions (2n=39-28; Sant Sadurní d'Anoia, Castelldefels and Viladecans). Mice from Rb populations are characterized by having between one and six Rb fusions involving 12 different chromosomes [Rb(3.8), Rb(4.14), Rb(5.15), Rb(6.10), Rb(9.11), and Rb(12.13)], either in heterozygous or homozygous states (**Table S1**) (Medarde et al. 2012; Vara et al. 2019a). The BRbS is characterized by Rb fusions present as chromosomal polymorphisms, thus not fixed within populations.

Moreover, a primary fibroblast cell line derived form a male mouse (#954, 2n=30) previously established in our lab (Sánchez-Guillén et al. 2015) was used as a somatic control in the Hi-C experiments. Cells were cultured in DMEM medium supplemented with 10% fetal bovine serum and 1% PenStrep at 37ºC and 5% CO_2.

Spermatocyte spreads and immunofluorescence. Direct analyses of recombination are normally based on the detection of either COs or their final products, the chiasmata, being the former visible cytogenetically in meiocytes in later stages of the first meiotic division (i.e., pachytene chromosomes) (Capilla et al. 2016; Ruiz-Herrera et al. 2017). Here, we analyzed physical location of COs along the axes of chromosomes using the immunofluorescence staining technique to detect MLH1, a protein that localizes type I (interfering) COs along with one of the proteins involved in synaptonemal complex formation (the synaptonemal complex protein 3, SYCP3). The chromosomal distribution of MLH1 foci can be considered a proxy of the number of COs (Froenicke et al. 2002; Ruiz-Herrera et al. 2017; Segura et al. 2013). The position of the centromeres along the chromosomal axes was visualized by staining centromeric proteins using the sclerodactyly and telangiectasia (CREST) serum (**Figure 4.10 A and Table S1**). Additionally, we also detected RAD51 (marker of programmed DSBs) at early stages of prophase I and H3K9Me3 (marker of centromeric constitutive heterochromatin). Spermatocyte spreading and immunofluorescence was performed as previously described (Segura et al. 2013).

Sperm analysis. Analyses were conducted in a subset of 15 of male mice from the BRbS (**Table S1**). This included 6 standard mice from the Castellfollit del Boix population and 9 mice with Rb fusions from Viladecans and Castelldefells population. Briefly, right epididymis was obtained for each specimen, and the caudal portion was drained to acquire epididymal spermatozoa. The process was supervised with an Olympus SZ30 stereoscope microscope. Spermatozoa were processed following the protocols for the examination and processing of human semen samples described by the World Health Organization (2010). Around 1,000 spermatozoa were analyzed per individual using an Olympus CH30 microscope.

For each sample, a slide with a drop of fresh spermatozoa suspension was analyzed. Depending on the characteristics of the spermatozoa movement, they were classified into the following categories: (i) fast progressive and linear motility, (ii) slow progressive motility by swinging or doing circular movements, (iii) non progressive motility but movement on the head/tail and (iv) no motility. In order to determine sperm vitality, a drop of eosin Y (0.5%, Sigma-Aldrich) and a drop of sperm suspension were mixed on a slide. After 2 minutes, the preparation was assessed under the

microscope. This method allows the identification of live (not stained) or dead spermatozoa (stained by the inability to expel eosin).

Image and COs data analysis. Spermatocyte preparations were visualized and captured using a Zeiss Axioskop epifluorescence microscope equipped with the appropriate filters and a charged coupled device camera (ProgRes® CS10plus, Jenoptik).

For the analysis of COs and centromere position only pachytene spermatocytes were considered. Only axis-associated MLH1 foci were counted considering the number of MLH1 foci per arm, chromosome and per cell. SC length (expressed in µm) was calculated as the mean length of all autosomal SCs per cell in individual mice. Thus, we measured the SC length of each chromosome and the physical distance between COs. For the analysis of RAD51 only leptotene and zygotene spermatocytes were considered.

The Micromeasure 3.3 software (Reeves 2001) was used for the analysis of chromosome-specific recombination maps based on the distances between adjacent MLH1 foci as previously described (Ruiz-Herrera et al. 2017; Segura et al. 2013). For each chromosome, the position of each MLH1 focus was recorded as a relative position (percentage of the synaptonemal complex total length) from the centromere, identified by the immunofluorescence signal of the centromeres in each cell. A discrimination of heterozygous chromosome states was accounted for, in parallel for the presence of double centromere signals in homozygous metacentrics. In order to minimize possible karyotype effects, analyses on CO chromosomal distribution, asynapsis and double centromeres were conducted in animals belonging to one standard population (Caldes de Montbui) and one Rb population (Viladecans).

CO frequency and cumulative frequency plots were constructed for each chromosome types (acrocentric, Rb in heterozygous state and Rb in homozygous state). As CO data were not normally distributed, analysis of variation in the number and position of MLH1 foci along chromosomes among different groups were assessed using non-parametric tests, as previously described (Capilla et al. 2014; Merico et al. 2013). In this case, the Mann-Whitney test was used in comparisons between groups, Kruskal-Wallis to test intra-group differences (either by population or upon fusions) and Dunn's test for multiple comparisons. Data were expressed as the mean±standard deviation (SD). Additionally, Chi-square was used to compare arm proportions of different number of MLH1 foci per arm type. Moreover, Spearman correlations were calculated to test relationships among MLH1 and RAD51 foci, MLH1 foci and diploid number and MLH1 foci and the number and state of chromosomal fusion. Spearman correlation analysis was also performed for MLH1 foci and chromosome arm length. A value of $p \leqslant 0.05$ was considered statistically significant.

SNP genotyping. Analysis of genomic divergence and recombination rate at the Mbp scale were conducted using genotyping data from a subset of 34 mice from two standard populations (2n=40; Castellfollit del Boix and Olost) and two Rb populations (2n=28-39; Sant Sadurní d'Anoia and Castelldefels) retrieved from (Vara et al. 2019a) (**Table S1**). Data consisted of the Mouse Universal Genotyping Array (MegaMUGA), which consisted of 77,808 evenly distributed SNP markers built on the Illumina Infinium platform (Morgan et al. 2016b). SNPs were filtered to remove markers with missing values >5% threshold using PLINK version 1.9 (Purcell et al. 2007). This resulted in a final data set of 63,344 informative SNPs distributed across all chromosomes, with the exception of the 8 and the Y chromosomes. These markers were considered for subsequent analyses of genome-wide screening of divergence and estimates of recombination rate.

The ADMIXTURE software (Alexander et al. 2009) was used to estimate individual ancestry and admixture proportions assuming K populations based on a maximum likelihood method. Analyses were run only for SNPs with a greater than 95% genotype call. The numbers of clusters (K) were evaluated applying Evanno's ∆K (Evanno et al. 2005) with three different K values (K=3, 4, 5) showing the lowest likelihood values (data non shown). ADMIXTURE analyses were plotted the R

package Pophelper v2.3.0 (Francis 2017). In addition, multiple dimensional scaling analysis was performed using PLINK, by first generating a genome file (–genome flag) from the vcf file containing the SNPs, and then the mds file that was plotted in R environment.

Genome-wide screening of genetic divergence and diversity. We estimated number of alleles (N_a), allelic richness (A_r), observed heterozygosity (H_O), expected heterozygosity (H_S), inbreeding coefficient (F_{IS}), and nucleotide diversity (pi). Na and Ar were estimated using the hierfstat package v0.04-22 (Goudet 2005) implemented in R. Allelic richness was refracted for a minimum of alleles of 22 (or 11 diploid samples), which was the lowest observed sample size between the three groups. Ho and Hs were calculated using PLINK v1.90b6.12 (Purcell et al. 2007); and F_{IS} and pi with VCFtools 162v.0.1.16. 1,000 bootstraps were performed for pairwise F_{ST} (Weir and Cockerham 1984) estimations with the R's StAMPP v1.6.1 package (Pembleton et al. 2013).

Pairwise F_{ST} comparisons were conducted between populations genome-wide and considering chromosomes involved in fusion (3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14 and 15) and not involved in fusions (1, 2, 7, 10, 16, 17, 18, 19 and X). Estimated F_{ST} values were adjusted with the Bonferroni correction for minimizing type I errors. Tukey-Kramer tests [JMP package version 5.1.2; SAS Institute Inc. (SAS Institute Inc 2019)] were used to analyze differences between groups.

Estimates of recombination rates. The program LDhelmet (Chan et al. 2012) was applied for the estimation of recombination rates a fine scale (kbp). As LDhelmet has a 25 diploid samples limit (50 haplotypes), we sampled a random subset of 25 individuals from the 34 individuals included in the SNP analysis, using the vcftools "—max-indv" option and by chromosome using "—recode" and "—chr" (Danecek et al. 2011). LDhelmet estimates the recombination rates from phased chromosomes or haplotypes from a population, thus we phased our data using the software SHAPEIT (Delaneau et al. 2012) by using the option –rho 0.001. Once phased, each chromosome vcf file was split into two groups, according to standard or Rb samples. Each file was then transformed to LDhelmet input snps and pos files with the "—ldhelmet" flag. The likelihood tables were generated using LDpop (Alexander and Machiela 2020), and then transformed to LDhelmet format following the software's manual indications. We performed the analysis per chromosome, based on the SNP data. Estimations of the population-scaled recombination rate $\rho = 4N_e r$ were obtained using the parameters recommended by software's developers, where Ne is the effective population size and r the genetic map distance across the region analyzed. Using this approach, we established recombination rates in windows of 50 SNPs across the mouse genome considering two groups: standard and Rb mice. Mann-Whitney test were used to analyze differences between groups.

Fluorescence Activated Cell Sorting (FACS) of mouse male germ cells. Testis cell disaggregation and FACS was conducted as previously described (Vara et al. 2019b). Briefly, germ cells at a concentration of 1 million per 500 µl were incubated in formaldehyde (1%) for 10 min prior to FACS. Glycine (0.125 M) was added and incubated with agitation at room temperature for 5 min and then at 4ºC for 15 min. Cells were then centrifuged for 10 min at 290 x*g* at 4ºC and resuspended in 3 ml 1x PBS with Hoechst staining.

Germ cells were sorted using a BD InfluxTM (BD Biosciences) coupled with an ultraviolet laser (355 nm). Subsequently, two main germ cell populations (P/D and RS) were isolated by plotting Hoechst Blue (UV355-460/50) vs. Hoechst red (UV355-670/30) emissions to discriminate cells by both their DNA content and their complexity. Cell populations were collected after sorting in 1x PBS and centrifuged for 5 min at 1,800 x*g*. The supernatant was discarded, and cell pellets were flash-frozen at -80ºC until use. Sorting experiments lasted between 3 to 6 hours to collect between 0.2×10^6 and 3.2×10^6 cells, depending on the germ cell population.

Cell enrichment of each flow-sorted population was evaluated by immunofluorescence using specific meiotic proteins and DAPI morphology. For primary spermatocytes, prophase-I stages (leptonema,

zygonema, pachynema and diplonema) were identified based on SYCP3 (1:400) and γH2AX patterns (1:300). Cell enrichment of round spermatids was determined based on nucleus morphology and DAPI pattern as previously described (Vara et al. 2019b). Cells were fixed on slides and then mounted with DAPI diluted in Vectashield (Vector Laboratories). Slides were analyzed using a fluorescence microscopy (Axiophot, Zeiss) coupled with a ProgRes ® CS10plus, Jenoptik camera. Representative images were captured with ACO XY (A. Coloma, Open Microscopy). Between 50 and 100 cells were counted for each flow-sorted population. Only sorted populations with enrichment above 80% were considered for subsequent experiments.

***In nuclei* Hi-C.**   The generation of Hi-C libraries was conducted following Vara and collaborators (Vara et al. 2019b). Rb mice included in the Hi-C analysis were selected based on their karyotype characteristics (high number of Rb fusions) and availability of testis material. All mice included in the Hi-C experiments belong to the same population (Viladecans) and were included in the recombination analysis, showing similar patterns of CO distribution (**Figure S1 F and Table S1**). Two replicates for cell type were obtained from a total of 3.4 x 106 primary spermatocytes at P/D stage, 12.8 x 106 round spermatids previously isolated by FACs. In addition a total of 10 x 10^6 Rb fibroblasts (two biological replicates) was also included. Libraries were submitted for Illumina sequencing (paired-end 75bp each side on HiSeq 2500, v4).

Hi-C data processing, binning and normalization.   Quality check and trimming step of raw data was carried out using BBDuk (version 10/2015) (Bushnell, 2014). Setting a minimum read length of 35 bp and a minimum Phred quality score of 20, adapters and low-quality reads were removed while preserving their longest high-quality regions. After the quality check, the reads were processed with TADbit (version 0.2.0.23) (Serra et al. 2017), which makes use of the GEM (version 1.7.1) mapper (Marco-Sola et al. 2012) to iteratively map them against the mouse genome (version mm10). Reads were mapped from 15 bp towards using a step size of 5 bp. The filters used to remove possible artefacts were the following: "self-circle", "dangling-end", "error", "extra dangling-end", "too short", "too large", "duplicated", and "random breaks". The maximum molecule length parameter was set at 2 times the 99.9 percentile of the insert size distribution, returned by the "insert_size" from TADbit. The maximum distance of a read to a cleavage site was set to the 99.9 percentile of the insert size distribution. An in-house script was used for binning and data normalization. This script imported the "HiC_data" module of TADbit, read the map files generated after the artefacts filtering step, binned the reads into a square matrix of 50 Kbp, and stored the matrix into a file in NPZ format (raw matrix). Afterwards, HiCExplorer (version 3.3) (Wolff et al., 2018) was used to normalize with the ICE (Iterative Correction and Eigenvector decomposition). The normalized matrices of standard and Rb were then compared by the log2ratio method using "hicCompareMatrix" from "HiCExplorer" to obtain the differential matrices.

Pair-wise comparisons between biological replicates derived from the Hi-C experiments were performed using HiCRep as previously described (Vara et al. 2019b). The Y chromosome was excluded from the analysis due to the lower number of interactions detected in our analysis (less than 1% of the overall detected interactions) and the highly repetitive nature of this chromosome. The correlation between 2 replicates was defined as the mean of the 20 correlation scores.

Averaged contact probability P(s).   ICE-normalized matrices were scaled with a factor of 1/sum(matrix). The resulting matrices were then input to "hicPlotDistVsCounts" from the HiCExplorer package in order to obtain the contact probability P(s).

Inter-chromosome/intra-chromosome interaction ratio.   ICE-normalized data stored in matrices were exported with HiCExplorer to the GInteractions format, which consists of 7 columns: chromosome, start and end from bin 1, chromosome, start and end from bin 2, and the amount of interaction.

The GInteractions tables were imported in R for further quantification of inter-chromosome and intra-chromosome interactions and plotting.

Inter-chromosomal interactions analysis.   Using R (version 3.6.1), the GInteraction tables were subset by chromosome so the analysis of each chromosome inter-chromosomal interactions could be done individually. Then, the mean of interactions of a given chromosome with others was calculated as previously described (Vara et al. 2019b). Finally, data was plotted as a heatmap where red stands for more interaction in Rb mice and blue in standard mice. The GInteraction tables subset by chromosome were then plotted as interaction profiles considering the interaction for each genomic position.

The analysis of the intra-chromosomal interaction ratio in RS (haploid cells) allowed for a predicted classification of fusions into homozygous or heterozygous by quantifying the interactions of fused chromosomes when compared to chromosomes not involved in Rb fusions. In this manner, we established that fusions 3.8, 6.10 and 5.15 were present in heterozygosis (ratio between 1.2 and 1.6) while fusions 9.11, 12.13 and 4.14 were in homozygosis (ratio around 2).

Chromosomal specific inter-chromosomal interactions analysis.   Inter-chromosomal regions were statistically defined as bins with a standard deviation (Z-score) higher than 2.58. Bin regions were also intersected with BEDTools intersect (version 2.25) against promoter regions (-2kbp to TSS) from the mouse GENCODE annotation vM14 to obtain the overlapping genes. These genes then underwent a Gene Ontology Enrichment Analysis (GOEA) as previously described (Vara et al. 2019b). We used CIRCOS (version 0.69-8) (Connors et al. 2009) to plot the genomic positions of the inter-chromosomal interactions across the mouse genome.

For the analysis of repetitive sequences within inter-chromosomal regions, repeatMasker (Smit et al. 2015) annotation on the *Mus musculus* genome version mm10 was downloaded and parsed to become a BED file. BEDtools (version 2.26) was used to intersect equal-sized bins considered as inter-chromosomal interaction regions against the RepeatMasker annotation with option "-wo", thus counting the number of overlapping bases. All equal-sized bins of the genome were also intersected, and the number of overlapping bases was also counted and considered as "background" repeat profile of the mm10 genome. Mann-Whitney test was applied to test for significant differences.

A/B compartments and TADs calling.   Analyses were conducted at the genome-wide level as previously described (Vara et al. 2019b). For A/B compartment calling columns with a low number of counts were filtered out using TADbit, setting the parameter min_count to 10. Since TADbit fits the column count distribution into a polynomial distribution, columns with a number of counts smaller than the first antimode of the distribution, which cannot be smaller than the min_count parameter, are filtered out. Then, the genome-wide matrices were normalized by the expected interactions at a given distance and by visibility by means of one-iteration of the ICE method. The correlation analysis was also performed with TADbit. In-house scripts computed A/B compartments from the first eigenvector, using 0 as threshold to differentiate both compartments and the gene density to label them.

TADs were identified using an in-house script that imported the "Chromosome" module of TADbit and added the raw and the ICE-normalized matrices of each chromosome separately. Filtered bins, due to low counts, were included to mask them when calling TADs. TAD insulation scores were obtained by first normalizing the different matrices for read depth in order for the scores to be comparable. Each matrix was then scaled to have 100M reads. Afterwards, TAD insulation scores were obtained from the output given by the "hicFindTADs" program from HiCExplorer.

Compartment switching. BED files with a resolution of 50 Kbp were available from the compartments definition step. Each genomic bin of 50 Kbp had its corresponding compartment attributed. Pair-wise comparisons between cell types -genome-wide and per-chromosome- were performed; the ratio of compartment switching was calculated as the number of genomic bins with a compartment change (A>B or B>A) divided by the total number of bins. From these files, a matrix file was created with 50 kbp-binned genomic coordinates as rows and cell types as columns, filled by the corresponding compartment labelling in each bin and cell type. Cell-specific A compartments were defined as those bins being compartment A in a cell type and compartment B in the remaining cell types.

Quantification and statistical analysis. The statistical analyses were performed using R. Statistical parameters and tests are reported in the Figures and Figure Legends.

Availability of data and materials. Hi-C dataset from standard mice are available in the NCBI GEO repository, accession number GSE132054. Raw and processed Hi-C data from Rb mice can be found in accession number GSE145978.

Supplementary Information

The Impact of Chromosomal Fusions on 3D Genome Folding and Recombination in the Germ Line

Supplementary Figure 1: (A) Correlation between the number of MLH1 foci per cell and (i) diploid number (2n) and (ii) the number of Rb fusions in all sampled mice (Spearman, p<0.0001). **(B)** Distribution of COs in Standard (St, red) and Rb mice (blue) showing greater co-variance in Rb mice (CV=10.14) respect their St counterparts (CV=9.14). **(C)** Mean numbers of RAD51 foci detected in laboratory mice (BL6) and wild-caught St and Rb mice. P-values (Dunn's test, **p<0.001; ***p<0.0001) are indicated. **(D)** Correlation between the number of RAD51 and MLH1 foci per cell in Standard (St, red) and Rb mice (blue) (Spearman, p<0.0001). **(E)** MLH1/RAD51 ratio in Standard (St, red) and Rb mice (blue) (Mann-Whitney test, p=0.001). **(F)** Proportion of MLH1 foci per arm type showing the proportion of #0, #1, #2 and #3 foci considering different chromosomal arm configurations: Acr BL6, all acrocentric chromosomes of BL6 mice; Acr St, all acrocentric chromosomes of standard mice; Acr Rb, acrocentric chromosomes of Rb mice; Met Het, Rb chromosomes in heterozygous state of Rb mice; Met Hom, Rb chromosomes in homozygous state of Rb mice. Each panel represent different mice from Viladecans populations with different karyotypes: specimens with four Rb fusions in homozygous state (HM), specimens with three Rb fusions in homozygous state plus two fusions in heterozygous state (HT), and specimens with four Rb fusions in homozygous state and one in heterozygous state. The p-arms of homozygous chromosomes are the ones presenting greater levels of achiasmy (Chi square, p<0.0001). Related to Figure 1.
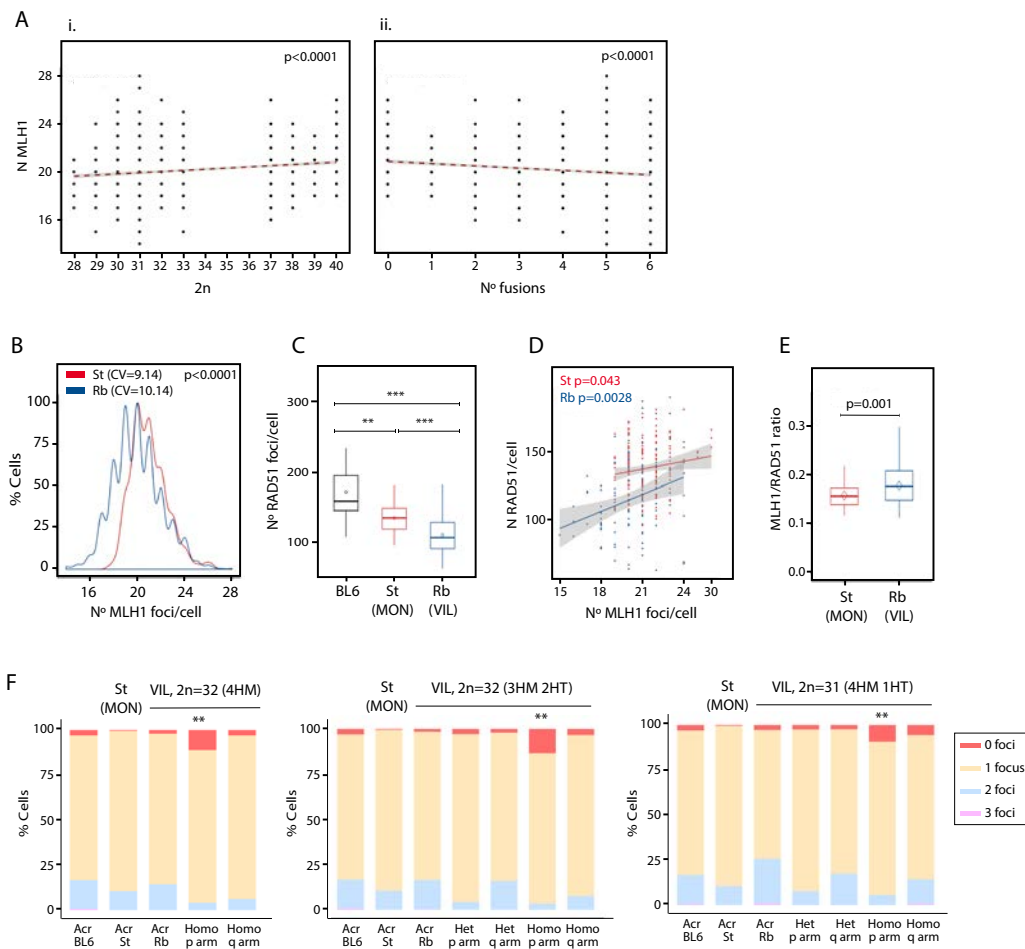
Supplementary Figure 2: (A) Principal component analysis in a subset of 34 mice from two standard populations (Castellfollit del Boix – BOI and Olost – OL) and two Rb populations (Sant Sadurní d'Anoia – SS and Castelldefels – CAS) (Table S1). **(B)** Plots showing the proportion of inferred ancestry for K = 2 to K = 4 in mice from panel A. **(C)** Recombination estimates expressed as $4N_er$/kbp between standard and Rb populations derived from the analysis of SNPs data using LDhelmet. (i) Box plots depict genome-wide recombination rates for standard (red) and Rb mice (blue). Diamonds indicate mean values (Mann-Whitney test ***p<2.2 e-16). (ii) Differences in recombination estimates between standards and Rb mice considering chromosomes involved in Rb fusions (3.8; 4.14; 5.15; 6.10; 9.11 and 12.13) and those that are not fused (1, 2, 7, 10, 16, 17, 18, 19 and X). **(D)** Representation of recombination rates (expressed as $4N_er$/kbp) derived from the analysis of SNPs data using LDhelmet for all autosomal chromosomes. Chromosomal distributions of recombination rates are represented for standard populations in red and for Rb populations in blue. Related to Figure 1.

A PCA

B

C i. ii.

D Chromosomal-specific recombination maps from SNPs data

genomic distance from centromere (kbp)
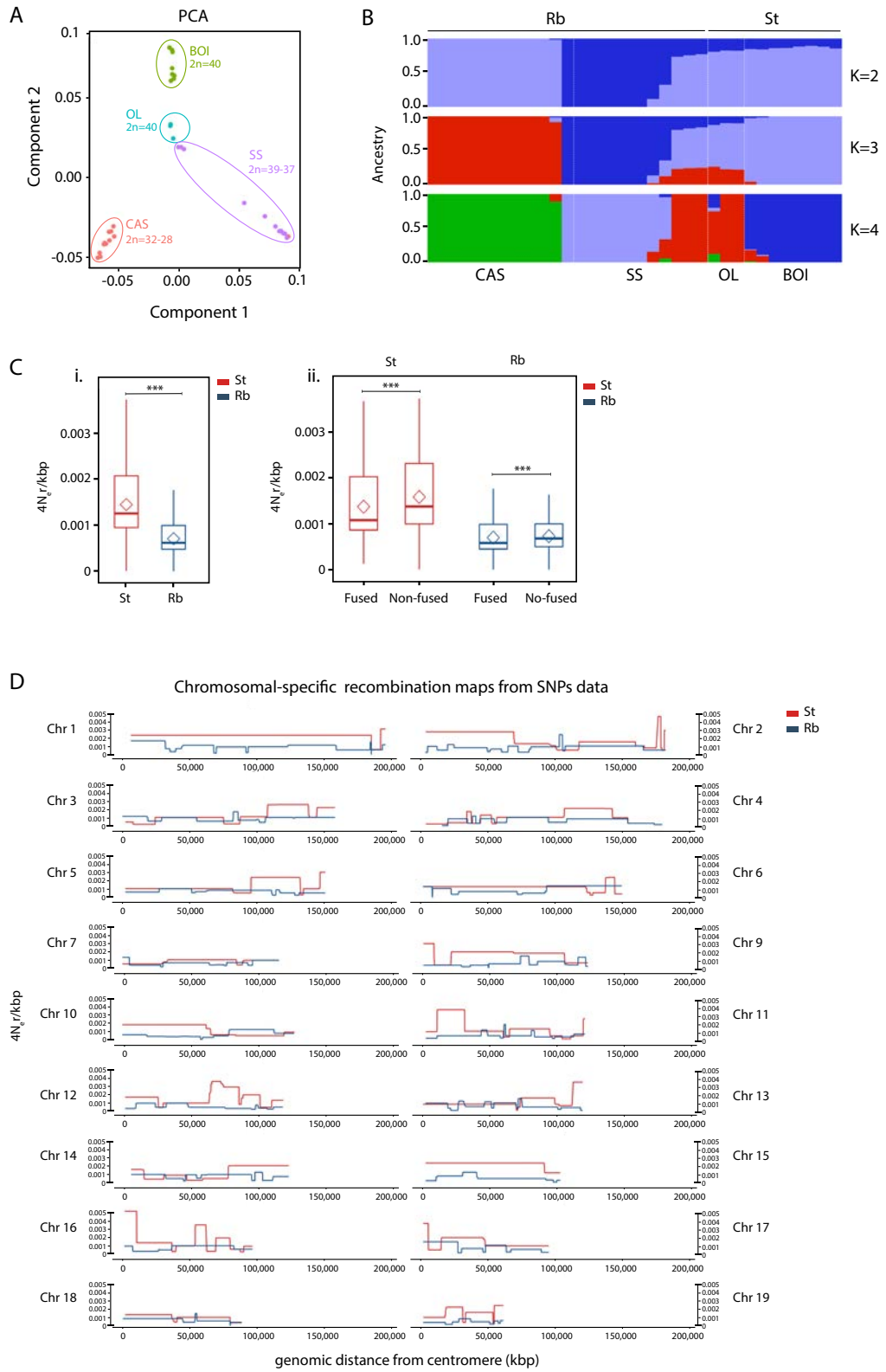
Supplementary Figure 3: **(A)** Axis length analysis in standard (St) and Robertsonian (Rb) mice from BRBS (see text and Table S1 for further details). Mann-Whitney test, p=0.86; NS: non-significant. **(B)** Correlation between axis length (expressed as microns) and the number of MLH1 foci (#0, #1, #2 and #3 foci) per chromosomal arm for each chromosome type (Spearman test, p<0.0001). **(C)** Distribution of MLH1 foci along individual chromosomal arms in heterozygous metacentrics. The X-axis represents the positions on the chromosomal axes from the centromeric end (black dot) to the distal telomere. The Y-axis indicates the frequency of MLH1 foci in each 10% interval of chromosomal length. **(D)** Proportion of asynapsed heterozygous Rb fusions per cell (left panel) and proportion of homozygous Rb fusions with double centromeric signals (CEN) in each Rb mice analyzed (Table S1, number of cells = 573). Related to Figure 2.
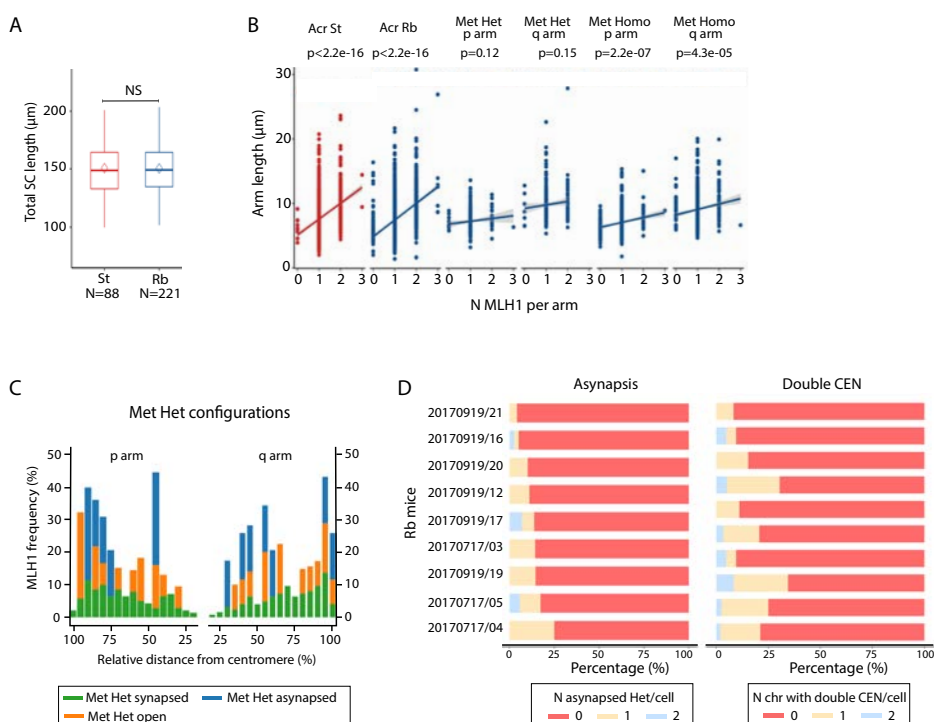
Supplementary Figure 4: **A)** Representation of the workflow for the Hi-C experiments. **(B)** FACS profile of Rb mice testis where gates for pachytene/diplotene (P/D) and round spermatids (RS) were placed and sorted accordingly. Insets show immunofluorescence examples of the cell type sorted in the P/D and RS population (scale bar=10 µm). In the case of P/D, DAPI is represented in blue, SYCP3 in green and gamma-H2AX in red. In round spermatids DAPI is represented in grey. (iii) Enrichment values obtained for P/D and RS populations after FACS, both above 90%. **(C)** Heatmap showing the correlation values among replicates based on the pairwise similarity score calculated using HiCRep (see Methods). Legend – P/D: Pachytene/Diplotene, RS: Round Spermatids, Rep#: replicates. Related to Figure 3 and 5.
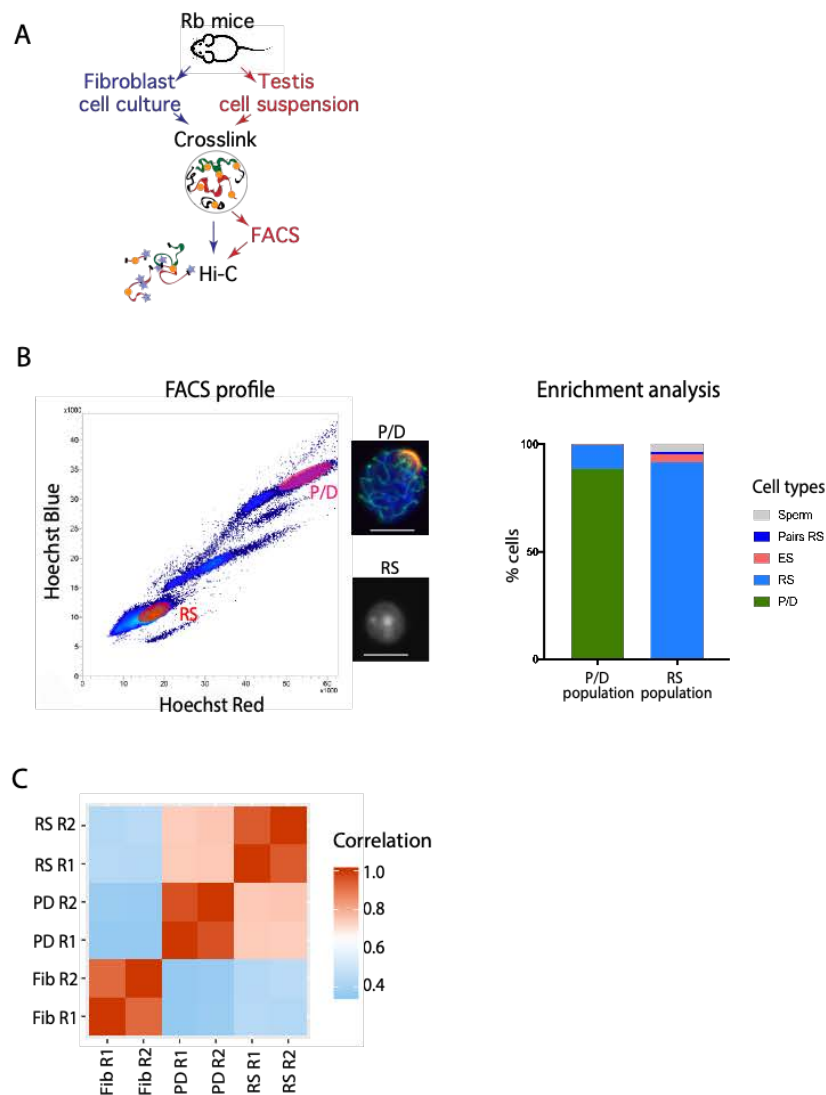
Supplementary Figure 5: TAD border alignments along chromosome 18 between standard and Rb mice in all cell types included in the study (fibroblasts, pachynema/diplonema – P/D and round spermatids – RS). Dark grey arches represent TADs with higher (light grey lower) than expected intra-TAD interactions. TAD border robustness (from 1 to 10) is represented by a colour gradient. Related to Figure 3 and 5.
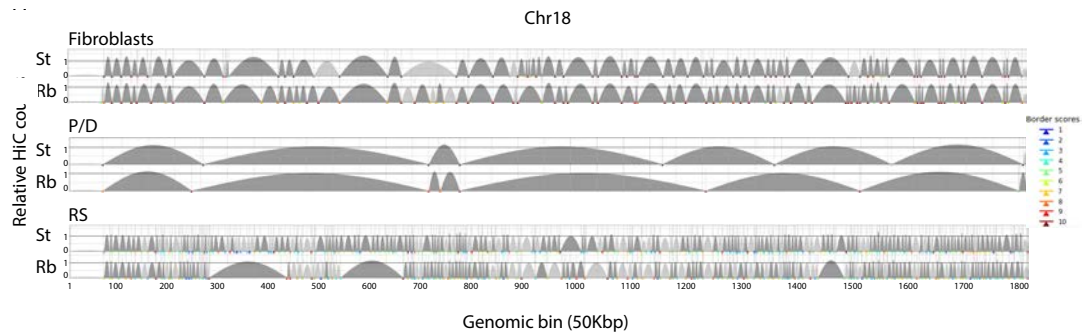
Supplementary Figure 6: **(A)** Representation of repetitive sequences represented in the mouse genome ('genome-wide') when compared to inter-chromosomal interactions (Wilcoxon test, ****p≤ 0.0001). **(B)** Detailed analysis of the different types of repeats present in sensory perception bins when compared to the rest of the mouse genome (Wilcoxon test, ****p≤.0001, ***p≤0.001, **p≤0.01, *p≤0.05). Related to Figure 6.
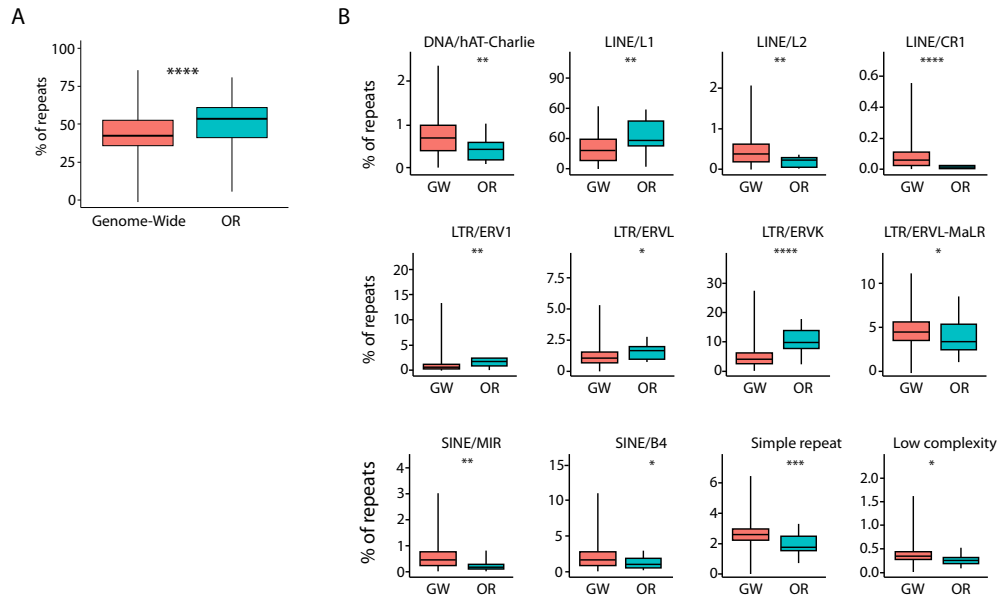
Table S1: List of all wild-caught mice included in the study. Related to Figures 4.10, 4.11, 4.12, 4.13, 4.14 and 4.15. Populations sampled, sex, individual ID, diploid number (2n) and Rb fusions for each mouse from the Barcelona Rb system. The analysis performed in each individual are detailed. POP-Population, F-Female, M-Male, len-length, R-replicate, n.a.-not analyzed, CO-crossover, SC-synaptonemal complex, CEN-centromere.

POP	SEX	ID	2n	FUSIONS	N MLH1 per cell	Mean MLH1 per cell	N RAD51 per cell	Mean RAD51 per cell	SC len. and CO/arm	Double CEN signal	Sperm	SNPs	Hi-C
Caldes de Montbuí	M	01_ST_-20160316_01	40		25	20.84	25	123.32	**yes**	n.a.	n.a.		
Caldes de Montbuí	M	01_ST_-20160316_02	40		25	20.52	25	140.68	**yes**	n.a.	n.a.		
Caldes de Montbuí	M	01_ST_-20160316_03	40		25	20.72	25	130.64	**yes**	n.a.	n.a.		
Caldes de Montbuí	M	01_ST_-20160316_04	40		25	21.32	25	128.40	**yes**	n.a.	n.a.		
Caldes de Montbuí	M	01_ST_-20160316_05	40		25	20.84	25	127.60	**yes**	n.a.	n.a.		
Caldes de Montbuí	M	01_ST_-20160316_07	40		25	22.12	25	143.40	n.a.	n.a.	n.a.		
Caldes de Montbuí	M	01_ST_-20160316_08	40		25	21.20	25	150.28	**yes**	n.a.	n.a.		
Castellfollit del Boix	M	02_ST_1000	40		25	20.68	n.a.	n.a.	n.a.	n.a.	**yes**		

Continued on next page

Table S1 – *Continued from previous page*

POP	SEX	ID	2n	FUSIONS	N MLH1 per cell	Mean MLH1 per cell	N RAD51 per cell	Mean RAD51 per cell	SC len. and CO/arm	Double CEN signal	Sperm	SNPs	Hi-C
Castellfollit del Boix	F	02_ST_1001	40		n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Castellfollit del Boix	M	02_ST_1019	40		n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Castellfollit del Boix	M	02_ST_1020	40		20	20.90	n.a.	n.a.	n.a.	n.a.	**yes**	**yes**	
Castellfollit del Boix	M	02_ST_1021	40		21	20.81	n.a.	n.a.	n.a.	n.a.	**yes**		
Castellfollit del Boix	M	02_ST_1024	40		25	20.16	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Castellfollit del Boix	M	02_ST_979	40		26	20.27	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Castellfollit del Boix	M	02_ST_980	40		22	20.68	n.a.	n.a.	n.a.	n.a.	**yes**		
Castellfollit del Boix	F	02_ST_984	40		n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Castellfollit del Boix	M	02_ST_998	40		21	20.57	n.a.	n.a.	n.a.	n.a.	**yes**	**yes**	
Castellfollit del Boix	M	02_ST_999	40		14	20.57	n.a.	n.a.	n.a.	n.a.	**yes**	**yes**	

Continued on next page

POP	SEX	ID	2n	FUSIONS	N MLH1 per cell	Mean MLH1 per cell	N RAD51 per cell	Mean RAD51 per cell	SC len. and CO/arm	Double CEN signal	Sperm	SNPs	Hi-C
Olost	M	03_ST_10L	40		n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Olost	F	03_ST_11L	40		n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Olost	M	03_ST_14L	40		n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS12	36	1 HM, 2 HT (4.14, 9.11, 12.13)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS13	38	1 HM (12.13)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS18	37	3 HT (4.14, 9.11, 12.13)	47	20.15	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS19	37	1 HM, 1 HT (4.14, 9.11)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS20	39	1 HT (12.13)	27	20.26	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS22	39	1 HT (4.14)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS2	35	1 HT, 2 HM (4.14,9.11, 12.13)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	

Table S1 – *Continued from previous page*

POP	SEX	ID	2n	FUSIONS	N MLH1 per cell	Mean MLH1 per cell	N RAD51 per cell	Mean RAD51 per cell	SC len. and CO/arm	Double CEN signal	Sperm	SNPs	Hi-C
Sant Sadurni d'Anoia	F	04_Rb_SS4	38	2 HT (4.14, 12.13)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS5	37	1 HM, 1 HT (4.14, 12.13)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS6	37	1 HT, 1 HM (4.14,9.11)	40	19.80	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS7	38	2 HT (4.14, 9.11)	46	19.72	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Sant Sadurni d'Anoia	M	04_Rb_SS8	37	3 HT (4.14, 9.11,12.13)	46	20.57	n.a.	n.a.	n.a.	n.a.	n.a.		
Viladecans	M	05_Rb_-20170717_02	33	3 HM, 1HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	25	20.68	20	142.45	**n.a.**	n.a.	n.a.		**R #1**
Viladecans	M	05_Rb_-20170717_03	32	3 HM, 2HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	32	20.38	20	134.7	**yes**	**yes**	n.a.		**R #1**
Viladecans	M	05_Rb_-20170717_04	31	4 HM, 1HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	25	20.00	n.a.	n.a.	n.a.	**yes**	n.a.		**R #1**
Viladecans	M	05_Rb_-20170717_05	32	3 HM, 2HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	33	20.52	n.a.	n.a.	n.a.	**yes**	n.a.		**R #1**

Continued on next page

POP	SEX	ID	2n	FUSIONS	N MLH1 per cell	Mean MLH1 per cell	N RAD51 per cell	Mean RAD51 per cell	SC len. and CO/arm	Double CEN signal	Sperm	SNPs	Hi-C
Viladecans	M	05_Rb_-20170721_01	32	4 HM (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	25	20.89	n.a.	n.a.	**yes**	n.a.	n.a.		
Viladecans	M	05_Rb_-20170727_01	33	3 HM, 1HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	41	20.12	n.a.	n.a.	**n.a.**	**n.a.**	n.a.		
Viladecans	M	05_Rb_-20170906_03	32	4 HM (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	25	19.16	n.a.	n.a.	**yes**	n.a.	n.a.		
Viladecans	M	05_Rb_-20170919_12	31	4 HM, 1HT (3.8, 4.14,5.15, 6.10, 9.11, 12.13)	15	19.47	21	90.29	n.a.	**yes**	n.a.		**R #2**
Viladecans	M	05_Rb_-20170919_16	31	4 HM, 1HT (3.8, 4.14,5.15, 6.10, 9.11, 12.13)	32	19.09	21	91.90	n.a.	**yes**	n.a.		**R #2**
Viladecans	M	05_Rb_-20170919_17	32	3 HM, 2HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	34	20.12	n.a.	n.a.	**yes**	**yes**	n.a.		
Viladecans	M	05_Rb_-20170919_19	31	4 HM, 1HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	74	21.25	n.a.	n.a.	**yes**	**yes**	n.a.		**R #2**
Viladecans	M	05_Rb_-20170919_20	31	4 HM, 1HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	48	19.97	24	97.33	**yes**	**yes**	n.a.		**R #2**

Table S1 – *Continued from previous page*

POP	SEX	ID	2n	FUSIONS	N MLH1 per cell	Mean MLH1 per cell	N RAD51 per cell	Mean RAD51 per cell	SC len. and CO/arm	Double CEN signal	Sperm	SNPs	Hi-C
Viladecans	M	05_Rb_-20170919_21	32	3HM, 2HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	25	18.96	16	113.94	**yes**	n.a.	n.a.		
Viladecans	M	05_Rb_946	31	3 HM, 3 HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	22	18.86	n.a.	n.a.	n.a.	n.a.	**yes**		
Viladecans	M	05_Rb_947	32	3 HM, 2HT (3.8, 4.14, 6.10, 9.11, 12.13)	55	19.45	n.a.	n.a.	n.a.	n.a.	n.a.		
Castelldefells	M	06_Rb_922	29	4 HM, 1 HT (4.14, 5.15, 6.10, 9.11,12.13)	22	18.95	n.a.	n.a.	n.a.	n.a.	**yes**		
Castelldefells	M	06_Rb_923	28	6 HM (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	18	19.39	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	
Castelldefells	M	06_Rb_954	30	4 HM, 2 HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	39	21.82	n.a.	n.a.	n.a.	n.a.	**yes**	**yes**	**Fibros Hi-C**
Castelldefells	M	06_Rb_955	32	3 HM, 2 HT (4.14, 5.15, 6.10, 9.11, 12.13)	40	19.93	n.a.	n.a.	n.a.	n.a.	**yes**	**yes**	
Castelldefells	M	06_Rb_956	30	4 HM, 2 HT (3.8, 4·14, 5.15, 6.10, 9.11, 12.13)	37	19.62	n.a.	n.a.	n.a.	n.a.	**yes**		
Castelldefells	F	06_Rb_960	28	5 HM (4.14, 5.15, 6.10, 9.11, 12.13)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**yes**	

POP	SEX	ID	2n	FUSIONS	N MLH1 per cell	Mean MLH1 per cell	N RAD51 per cell	Mean RAD51 per cell	SC len. and CO/arm	Double CEN signal	Sperm	SNPs	Hi-C
Castelldefells	F	06_Rb_963	29	4 HM, 3 HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13, 7.17)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	yes	yes	
Castelldefells	M	06_Rb_967	29	5 HM, 1 HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	44	19.95	n.a.	n.a.	n.a.	n.a.	yes	yes	
Castelldefells	M	06_Rb_968	30	4 HM, 2 HT (3.8, 4·14, 5.15, 6.10, 9.11, 12.13)	27	19.85	n.a.	n.a.	n.a.	n.a.	yes	yes	
Castelldefells	M	06_Rb_970	30	4 HM, 2 HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	19	19.00	n.a.	n.a.	n.a.	n.a.	n.a.	yes	
Castelldefells	M	06_Rb_971	30	4 HM, 2 HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	34	18.74	n.a.	n.a.	n.a.	n.a.	yes		
Castelldefells	M	06_Rb_972	31	3 HM, 3 HT (3.8, 4.14,5.15, 6.10, 9.11, 12.13)	16	18.13	n.a.	n.a.	n.a.	n.a.	n.a.	yes	
Castelldefells	M	06_Rb_CS10	31	4 HM, 1 HT (3.8, 4.14, 6.10, 9.11, 12.13)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	yes	
Castelldefells	F	06_Rb_CS13	30	4 HM, 2 HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	yes	
Castelldefells	F	06_Rb_CS14	30	4 HM, 2 HT (3.8, 4.14, 5.15, 6.10, 9.11, 12.13)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	yes	

Table S2: Genetic diversity. Representation of number of individuals (n), number of alleles (Na), allelic richness (Ar), observed heterozygosity (Ho), expected heterozygosity (Hs), inbreeding coefficient (Fis), and nucleotide diversity (pi).

Group	n	Na	Ar	Ho	He	Fis	pi
Standard	11	95006	95006	0.1576	0.1680	0.1048	0.1760
Sant Sadurní	11	91738	91738	0.1563	0.1483	-0.0055	0.1554
Castelldefels	12	93161	92862	0.1561	0.1536	0.0268	0.1602

Table S3: F_{ST} values. Related to Supplementary Figure 2. F_{ST} values between different populations. N.A.: Not analyzed.

	CASTELLFOLLIT	OLOST	CASTELLDEFELS
OLOST	0.1384	N.A.	N.A.
CASTELLDEFELS	0.1764	0.1656	N.A.
SANT SADURNI	0.1787	0.1736	0.1673

	CASTELLFOLLIT	OLOST	CASTELLDEFELS
OLOST	0.1384	N.A.	N.A.
CASTELLDEFELS	0.1764	0.1656	N.A.
SANT SADURNI	0.1787	0.1736	0.1673

Table S4: F_{ST} values between different populations according to the presence of Rb fusions. Related to Supplementary Figure 2. In bold, F_{ST} values when including chromosomes involved in Rb fusions (3.8; 4.14; 5.15; 6.10; 9.11 and 12.13). Not-bolded F_{ST} values correspond to comparisons including chromosomes not involved in Rb fusions (1, 2, 7, 10, 16, 17, 18, 19 and X). N.A.: Not analyzed.

	CASTELLFOLLIT	OLOST	CASTELLDEFELS	SANT SADURNI
CASTELLFOLLIT	N.A.	**0.1382**	**0.1781**	**0.1783**
OLOST	0.1362	N.A.	**0.1661**	**0.1709**
CASTELLDEFELS	0.1708	0.1621	N.A.	**0.1689**
SANT SADURNI	0.1780	0.1785	0.1630	N.A.

Table S5: Hi-C quality metrics per cell type. Related to Figures 4.12 and 4.14.

INFO PER CELL TYPE	Fib Rb	PD Rb	RS Rb
Raw (pairs)	499788826	451041073	450256473
Trimmed q20 (pairs)	473512735	424878135	423115492
Mapped uniquely (pairs)	338880234	296839992	297248047
Self-circle (% relative uniquely mapped)	0.16	0.24	0.32
Dangling-end (% relative uniquely mapped)	3.29	3.22	0.23
Error (% relative uniquely mapped)	2.87	2.72	0.11
Extra dangling-end (% relative uniquely mapped)	1.86	3.75	6.89
Too short (% relative uniquely mapped)	4.43	7.81	10.81
Too large (% relative uniquely mapped)	0.01	0.00	0.00
Duplicated (% relative uniquely mapped)	15.30	16.51	10.90
Random breaks (% relative uniquely mapped)	0.42	0.51	0.04
Total valid	255930322	253785807	255264346
Total valid (% relative to Raw)	51.21	56.27	56.69
Total valid (% relative to Trimmed)	54.05	59.73	60.33
Total valid (% relative to Mapped uniquely)	75.52	85.50	85.88

Table S6: Hi-C quality metrics per individual libraries. Related to Figures 4.12 and 4.14.

Information per library	HH3YTBBXX_6_7 (95401)	HH52MBBXX_2_7 (95401)	HHVVNBBXX_3_14 (95402)	HHW7MBBXX_3_14 (95402)	CBLPLANXX_4_12_1 (P7Rb)	CBLPLANXX_5_14_1 (P11Rb)
Raw (pairs)	115128742	143421403	177632905	63605776	209791352	225385285
Trimmed q20 (pairs)	110564127	136600866	166180747	60166995	203153522	217914691
Ratio Trimmed/Raw	0.96	0.95	0.94	0.95	0.97	0.97
Mapped uniquely (pairs)	81919920	101227127	114239111	41494076	141603450	150746989
Mapping efficiency (%)	74.09	74.10	68.74	68.96	69.70	69.18
Self-circle	80914	101630	245031	88014	273789	319209
Dangling-end	1460724	1832200	5620100	2057441	234781	204949
Error	956811	1759395	4880873	1735209	84316	114882
Extra dangling-end	860499	1078099	3060511	1085005	5581191	7090260
Too short	2047426	2511069	7419022	2705306	9953681	10435096
Too large	8801	10886	5275	1880	3865	4404
Duplicated	7507404	9414380	23861232	6585527	12592497	9848530
Random breaks	167942	208250	789805	366376	39701	38357
Total valid	69409690	85069608	73182181	28268843	114744489	124472314
Total valid (% relative to Raw)	60.29	59.31	41.2	44.44	54.69	55.23
Total valid (% relative to Trimmed)	62.78	62.28	44.04	46.98	56.48	57.12
Total valid (% relative to Mapped uniquely)	84.73	84.04	64.06	68.13	81.03	82.57

Information per library	CBMVPANXX_4_12_1 (P7Rb)	CCYYKANXX_3_14_1 (P7Rb)	CCYYKANXX_3_15_1 (P11Rb)	CD1RWANXX_3_15_1 (P11Rb)	CD1RWANXX_4_14_1 (P7Rb)
Raw (pairs)	30401344	67073167	71546194	153324994	143775210
Trimmed q20 (pairs)	28895159	61246874	65036146	140164655	131582580
Ratio Trimmed/Raw	0.95	0.91	0.91	0.91	0.92
Mapped uniquely (pairs)	20070529	42988025	46502151	99998907	92177988
Mapping efficiency (%)	69.46	70.19	71.50	71.34	70.05
Self-circle	38453	62978	68807	148676	133637
Dangling-end	33509	51368	30927	65999	110901
Error	11440	23840	17109	36111	50165
Extra dangling-end	788408	1387958	1973359	4424762	2631305
Too short	1415346	2283560	2410001	5205212	4920692
Too large	531	1376	1545	3305	2880
Duplicated	269464	462407	351747	1553969	2075878
Random breaks	5671	10288	7292	14256	29066
Total valid	17611261	38841899	41803930	88988102	82588158
Total valid (% relative to Raw)	57.93	57.91	58.43	58.04	57.44
Total valid (% relative to Trimmed)	60.95	63.42	64.28	63.49	62.77
Total valid (% relative to Mapped uniquely)	87.75	90.36	89.9	88.99	89.6

4.4 Characterization of PRDM9 variability in natural populations of house mice

PRDM9 Diversity at Fine Geographical Scale Reveals Contrasting Evolutionary Patterns and Functional Constraints in Natural Populations of House Mice

Covadonga Vara,[1,2] Laia Capilla,[1] Luca Ferretti,[3] Alice Ledda,[4] Rosa A. Sánchez-Guillén,[1,5] Sofia I. Gabriel,[6] Guillermo Albert-Lizandra,[1,2] Beatriu Florit-Sabater,[1,2] Judith Bello-Rodríguez,[1,2] Jacint Ventura,[7] Jeremy B. Searle,[8] Maria L. Mathias,[6] and Aurora Ruiz-Herrera*[1,2]

[1]Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain

[2]Departament de Biologia Cel.lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Barcelona, Spain

[3]Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

[4]Department for Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

[5]Instituto de Ecología AC (INECOL), Red de Biología Evolutiva, Xalapa, Veracruz, Mexico

[6]CESAM – Centre for Environmental and Marine Studies, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

[7]Departament de Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona, Barcelona, Spain

[8]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY

*Corresponding author: E-mail: aurora.ruizherrera@uab.cat.

Associate editor: Belinda Chang

Abstract

One of the major challenges in evolutionary biology is the identification of the genetic basis of postzygotic reproductive isolation. Given its pivotal role in this process, here we explore the drivers that may account for the evolutionary dynamics of the PRDM9 gene between continental and island systems of chromosomal variation in house mice. Using a data set of nearly 400 wild-caught mice of Robertsonian systems, we identify the extent of PRDM9 diversity in natural house mouse populations, determine the phylogeography of PRDM9 at a local and global scale based on a new measure of pairwise genetic divergence, and analyze selective constraints. We find 57 newly described PRDM9 variants, this diversity being especially high on Madeira Island, a result that is contrary to the expectations of reduced variation for island populations. Our analysis suggest that the PRDM9 allelic variability observed in Madeira mice might be influenced by the presence of distinct chromosomal fusions resulting from a complex pattern of introgression or multiple colonization events onto the island. Importantly, we detect a significant reduction in the proportion of PRDM9 heterozygotes in Robertsonian mice, which showed a high degree of similarity in the amino acids responsible for protein–DNA binding. Our results suggest that despite the rapid evolution of PRDM9 and the variability detected in natural populations, functional constraints could facilitate the accumulation of allelic combinations that maintain recombination hotspot symmetry. We anticipate that our study will provide the basis for examining the role of different PRDM9 genetic backgrounds in reproductive isolation in natural populations.

Key words: PRDM9, Mus musculus domesticus, Robertsonian fusion, postzygotic reproductive isolation, selection, recombination.

Introduction

Understanding the genetic basis of speciation is a long-standing quest in biology. This entails the investigation of mechanisms responsible for postzygotic reproductive isolation, which can be attributable to genic (e.g., speciation genes) or chromosomal (e.g., inversions and fusions) factors (Coyne and Orr 1998; Orr and Turelli 2001; Faria and Navarro 2010; Farré et al. 2013; Capilla et al. 2014, 2016). Whether both types of mechanisms can influence speciation processes independently or in combination remains largely unexplored.

Genes that cause hybrid sterility have been described mostly in *Drosophila* and include the *Odysseus-site Homeobox* (*OdsH*) gene (Ting et al. 1998), the *JYAlpha* gene (Masly et al. 2006), *Hmr* (Barbash et al. 2003), nucleoporin *Nup96* (Presgraves et al. 2003), and the *Overdrive* (*Ovd*) gene (Phadnis and Orr 2009). In mice, the PR domain zinc finger 9 (*Prdm9*) gene contributes to hybrid sterility in *Mus musculus* subspecies (*Mus musculus domesticus* × *Mus musculus musculus*) (Mihola et al. 2009). The *Prdm9* gene is located within the proximal centromeric regions of mouse chromosome 17

and it is implicated in the genomic distribution of recombination hotspots. It codes for a meiotic-specific histone (H3) methyltransferase with a C-terminal tandem repeat zinc finger (ZnF) domain that adds H3K4me3 marks at nucleosomes close to double-strand breaks (DSBs) in early meiosis; it does so through its recognition of a species-specific and highly polymorphic repetitive DNA motif (Mihola et al. 2009; Baudat et al. 2010). The high amino acid variation that characterizes PRDM9 in mice and the repetitive nature of the DNA motif that it recognizes have suggested the existence of a positive molecular feedback implicated in the generation of recombination hotspots (Oliver et al. 2009; Jeffreys et al. 2013; Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014; Schwartz et al. 2014; Baker et al. 2017). That is, the repair of DSBs in early meiosis can introduce mutations in preferred PRDM9 binding motifs, thus influencing the rapid evolution of novel binding domains (ZnFs) in the PRDM9 protein (Boulton et al. 1997; Coop and Myers 2007; Davies et al. 2016). Such evolutionary turnover has important implications for the establishment of reproductive postzygotic barriers since differences in recombination landscapes in hybrids can account for failure in chromosomal synapsis during meiosis (Qiao et al. 2012; Bhattacharyya et al. 2013; Smagulova et al. 2016). In this scenario, the presence of evolutionarily distinct heterozygous combinations of PRDM9 can result in asymmetric DSBs (i.e., differences in genomic distribution and number of DSBs) between homologous chromosomes, which depending on the interaction between different X-linked and autosomal loci (Balcova et al. 2016) can result in subfertile and even sterile phenotypes (Davies et al. 2016). Since interallelic PRDM9 incompatibilities can result in hybrid sterility due to failure in recognition of DNA-binding sites (Flachs et al. 2012), understanding PRDM9 variability in natural populations can reveal potential drivers behind the evolutionary dynamics of this gene.

With regards to the involvement of chromosomal rearrangements in evolution, house mice represent one of the most extraordinary models of chromosomal variation in mammals. The variation that exists in mice may promote speciation due to meiotic impairment and infertility associated with chromosomal heterozygosity (Capilla et al. 2014; Pavlova and Searle 2018). The standard karyotype of *M. m. domesticus* consists of 40 all-acrocentric chromosomes. However, numerous populations in Western Europe and North Africa show high variability in diploid numbers and karyotypes resulting from Robertsonian (Rb) fusions of non-homologous acrocentric chromosomes and/or whole arm reciprocal translocations, giving rise to new metacentric chromosomes (Piálek et al. 2005). This includes mice from the northeast of the Iberian Peninsula, the so-called "Barcelona Rb system" (Adolph and Klein 1981; Gündüz, López-Fuster, et al. 2001; Medarde et al. 2012) and on the island of Madeira (Britton-Davidian et al. 2000), both Rb systems are characterized by distinct patterns of chromosomal variation. On the one hand, the Barcelona Rb system extends over an area of 5,000 km^2 within the provinces of Barcelona, Tarragona, and Lleida (Spain) and includes individuals with diploid numbers (2n) ranging from 2n = 27 to 2n = 40 (Medarde et al. 2012,

and references therein). This Rb system is characterized by the presence of seven different metacentric chromosomes (Rb[3.8], Rb[4.14], Rb[5.15], Rb[6.10], Rb[7.17], Rb[9.11], and Rb[12.13]), distributed in nongeographically coincident (staggered) clines leading to a progressive reduction in diploid numbers towards the center of the range, about 30-km west of the city of Barcelona (Gündüz, López-Fuster, et al. 2001; Medarde et al. 2012). The Barcelona Rb system is also noteworthy for the high levels of chromosomal polymorphism, which is consistent with the absence of a metacentric race (Medarde et al. 2012; Capilla et al. 2014; Sánchez-Guillén et al. 2015). On the other hand, mice from the Madeira archipelago (including the islands of Madeira and Porto Santo) show an extensive chromosomal radiation that includes mice with the standard 2n = 40 (i.e., the island of Porto Santo) and mice with a highly structured chromosomal race organization on the island of Madeira itself (Britton-Davidian et al. 2000). Madeira, an island with an extreme topography, has six well-established chromosomal races that have been described within a geographical range of only 742 km^2. Diploid numbers vary from 2n = 22 to 2n = 28, with up to nine metacentric chromosomes accumulated within a maximum of 1,200 years (Förster et al. 2009). Importantly, most metacentric populations are geographically isolated and do not co-occur with others except for some cases of marginal overlap of the chromosomal races. The chromosomal differences between some of the Madeira races are so pronounced that hybrids are at very low frequency, presumably due to reduced F_1 fitness associated with meiotic impairment (Britton-Davidian et al. 2000).

The contrasting nature of the chromosomal variation in the Madeira and Barcelona Rb systems provides a unique opportunity to investigate the mechanisms underlying PRDM9 variation. It permits the examination of PRDM9 evolutionary dynamics between a continental mouse population (with a distribution of metacentric chromosomes in a polymorphic state; i.e., the Rb Barcelona system) and those of an island's chromosomal system (subdivided into entirely or partially reproductively isolated units with metacentric chromosomes in a homozygous state; i.e., the Madeira system). Here, using a phylogenetic approach on a data set of ∼400 wild-caught house mice from the Madeira archipelago (Madeira and Porto Santo Islands) and the Barcelona system (continental) (fig. 1) we 1) identify the extent of PRDM9 diversity in natural house mouse populations, 2) determine the phylogeography of PRDM9 at a local and global scale using a new method for the computation of genetic distances between complex repeats, and 3) analyze how selection might account for the PRDM9 natural diversity observed. In this context, our study provides the grounds for understanding the complexity of the mechanisms that can drive evolutionary dynamics of PRDM9 in natural populations.

Results

PRDM9 Diversity at a Local Scale

We successfully obtained polymerase chain reaction (PCR) products from the 395 wild-caught mice included in the

MBE
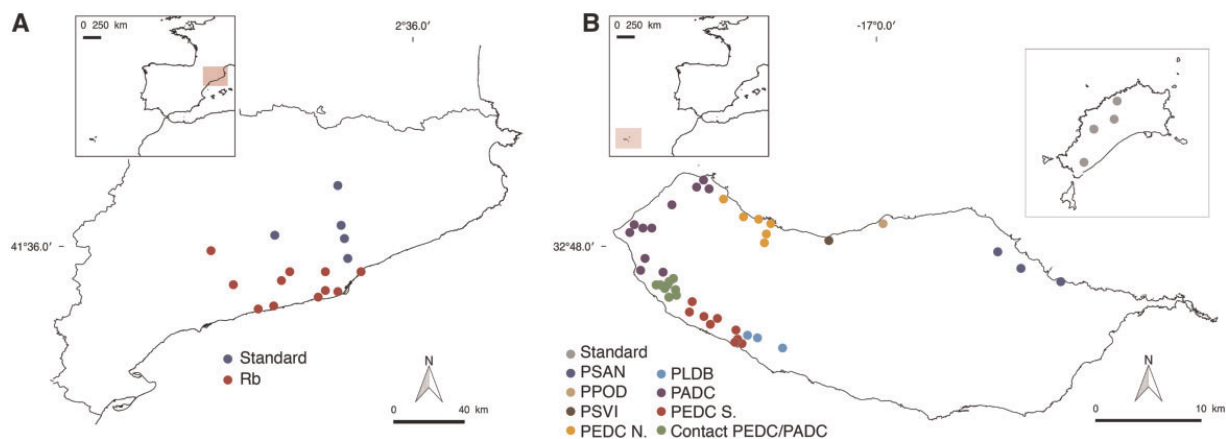
FIG. 1. Geographical distribution and chromosomal characteristics of the *Mus musculus domesticus* populations analyzed. (*A*) Localities sampled in the Barcelona Rb system. Standard populations (blue dots) correspond to mice with 40 acrocentric chromosomes (standard karyotype) whereas Rb populations (red dots) had diploid numbers ranging from 2*n* = 39 to 2*n* = 28 (see supplementary table 1, Supplementary Material online, for further details of chromosomes involved in Rb fusions). (*B*) Localities sampled in the Madeira archipelago. Porto Santo Island is displayed in the inset; it contains standard mice with 40 acrocentric chromosomes (standard karyotype). Six metacentric races have been sampled on the Madeira Island: PEDC, PADC, PLDB, PPOD, PSAN, and PSVI (see supplementary tables 2 and 3, Supplementary Material online, for further details on the chromosomal composition of the chromosomal races). Following previous studies (Britton-Davidian et al. 2000), the chromosomal race PEDC is distributed on both on the northern (PEDC N.) and southern (PEDC S.) coasts of the Madeira Island.

study, detecting considerable variation in the number of ZnF repeats, depending on the region sampled (table 1 and supplementary tables 1–4 and figs. 1–4, Supplementary Material online). In the Barcelona Rb system, alleles with 10 ZnF repeats were the most frequently detected (73%), followed by alleles with 12 ZnF repeats (13.2%), 11 ZnF repeats (10%), 8 ZnF repeats (2%), and 13 ZnF repeats (1.8%). In the case of Madeira however, a higher diversity was detected when considering the number of ZnF repeats. PRDM9 alleles with 11 ZnF repeats were the most frequently detected (47%), followed by alleles with 10 ZnF repeats (25%), 13 ZnF repeats (17%), 12 ZnF repeats (6%), 15 ZnF repeats (2%), and 14 and 16 ZnF repeats (1% in each).

In order to investigate the source of contrasting variability patterns observed in the Madeira and Barcelona Rb systems, we sequenced the ZnF array in a total of 292 mice. This represented 74% of the sampled mice, a proportion that is consistent with previous studies on wild specimens (e.g., Kono et al. 2014). This resulted in the identification of 25 distinct ZnFs, four of which (#22, #26, #27, and #29) were newly identified for *M. m. domesticus* in this study (supplementary table 5, Supplementary Material online).

In the Barcelona Rb system, we obtained PRDM9 sequences from 132 mice (101 homozygous and 31 heterozygous) that represented 13 different PRDM9 alleles (figs. 2 and 3). These were classified according to both the sequence and number of ZnF repeats: one allele with eight ZnF repeats (8A), three alleles with 10 repeats (10A, 10B, and 10C), three alleles with 11 repeats (11B, 11C, and 11D), and six alleles with 12 repeats (12B, 12C, 12D, 12E, 12F, and 12G) (fig. 2). Eight of the 13 different PRDM9 alleles found in the Barcelona Rb system (8A, 10C, 11C, 11D, 12D, 12E, 12F, and 12G) were newly identified in this study. The diversity of PRDM9 alleles was widely variable among localities, particularly when

Table 1. PRDM9 Variability Found in Wild *Mus musculus domesticus* including the Present Survey (Barcelona Rb system and Madeira archipelago) and Previous Studies in Eurasia (Buard et al. 2014; Kono et al. 2014).

	Barcelona	Madeira Archipelago		Eurasia[a]
		Madeira	Porto Santo	
N	185	199	11	76
PRDM9 alleles	13	53	7	27
Min ZnF	8	10	11	8
Max ZnF	13	16	12	17

NOTE.—*N*, number of wild mice included in the study; PRDM9 alleles, number of distinct PRDM9 alleles found based on the amino acid sequence; Min ZnF, minimum number of ZnF repeats; and Max ZnF, maximum number of ZnF repeats.
[a]Data obtained from Kono et al. (2014) and Buard et al. (2014).

comparing standard versus Rb populations, with the 10A allele the most commonly detected variant followed by 11B and 12B (fig. 3*A*). The highest variability was found in populations characterized by the standard karyotype (2*n* = 40; Castellfollit del Boix, Olost, and Santa Perpètua de Mogoda) where the majority of PRDM9 variants were represented. In addition, standard populations presented private alleles (8A, 10B, 11B, 10C, 12F, and 12G) that were undetected in Rb populations. In fact, only five alleles (10A, 11C, 12B, 12E, and 12D) were found in Rb mice. Among these, 10A was the most frequently distributed allele, with frequencies ranging from 23% to 100% in the localities Les Pobles, L'Ametlla de Segarra, Sant Sadurní d'Anoia, and Castelldefels (fig. 3*A*).

In the case of the Madeira archipelago, 160 mice were successfully sequenced (110 homozygous and 50 heterozygous), 152 of which were from the Madeira Island Rb system and 8 mice from Porto Santo Island (fig. 4). This represented 54 different alleles: six alleles with 10 ZnF repeats, 20 alleles with 11 ZnF repeats, 7 alleles with 12 ZnF repeats, 1 allele with

MBE
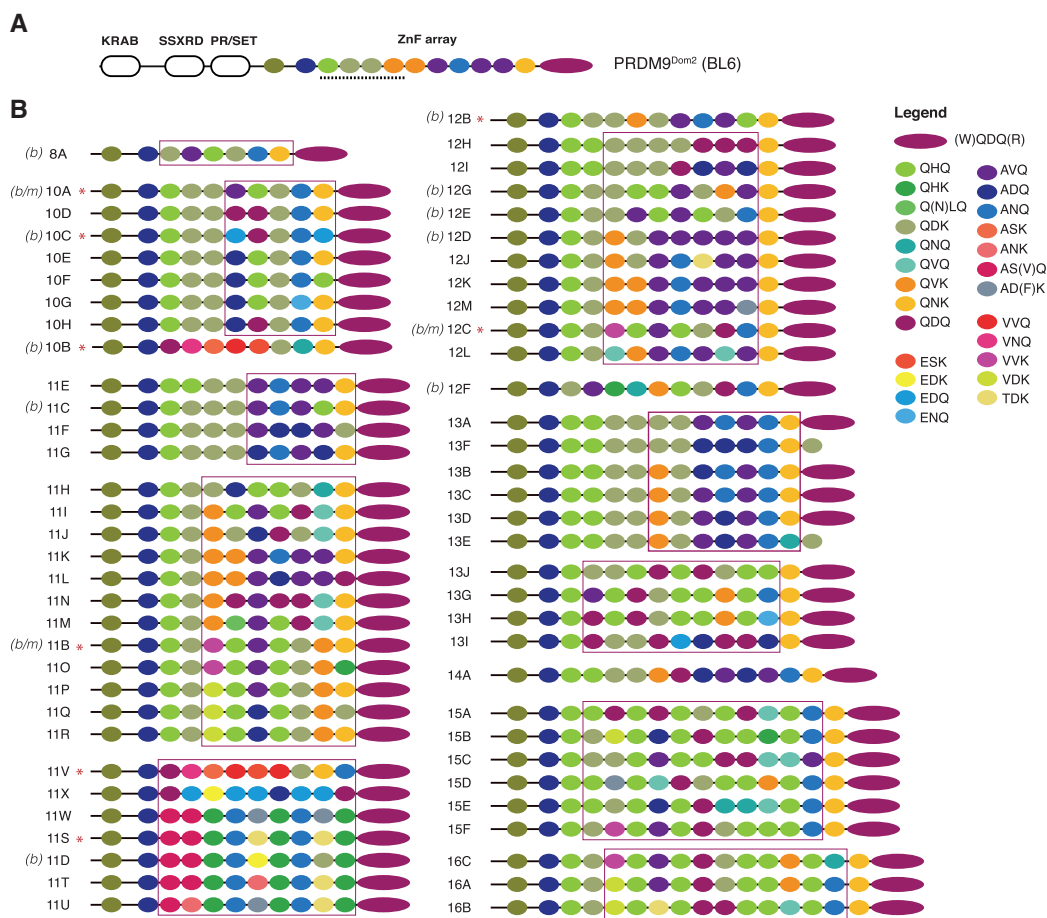
FIG. 2. Natural PRDM9 allelic diversity found in the Barcelona Rb system and Madeira archipelago *Mus musculus domesticus* populations. (A) Representation of the mouse PRDM9^{Dom2} protein (reference genome, C57/BL6 strain). It consists of four domains: KRAB-like, SSXRD, PR/SET, and ZnF array. The underlined ZnFs represent repeats from position ZnF3 to ZnF6 that bind to the DNA in the mouse (Baker et al. 2015; Paigen and Petkov 2018). (B) Representation of the ZnF alignments for all PRDM9 alleles found in the present study. Each ZnF is color coded based on amino acid sequence affinity found at the most variable sites (−1, +3, +6) responsible for DNA binding. Additional information on ZnF amino acid sequences is provided in supplementary table 5, Supplementary Material online. Purple boxes encompass the more variable ZnF among individuals. Red asterisks indicate previously described PRDM9 alleles (Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014). Geographical labels: (*b*) PRDM9 alleles found only in the Barcelona Rb system, (*b/m*) PRDM9 alleles found in both the Barcelona Rb system and the Madeira archipelago. PRDM9 alleles without geographical label correspond to alleles found only in the Madeira archipelago.

14 ZnF repeats, 6 alleles with 15 ZnF repeats, and 3 alleles with 16 ZnF repeats (figs. 2 and 4). Only five of them (10A, 11B, 11S, 11V, and 12C) were previously described in Eurasian populations and the Barcelona system (Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014). The allele 10A corresponds to the alleles 48 and Ce3 described by Kono et al. (2014) and Buard et al. (2014) in *M. m. domesticus* and *Mus musculus castaneus*, respectively. The allele 11B is homologous to the previously reported alleles 54, 55, and Db1 in *M. m. domesticus* (Buard et al. 2014; Kono et al. 2014), whereas 11S is the PRDM9 allele present on the t-haplotype (Kono et al. 2014). The 11V and 12C allele corresponded to allele 3 and 16, respectively (Buard et al. 2014).

Among the 49 alleles newly identified in the Madeira archipelago, the most commonly represented on Madeira Island was 11K, followed by 10A, 10E, 11S, and 13C (fig. 4A).

Moreover, some alleles were detected in all Rb races (supplementary table 2, Supplementary Material online) but absent in standard populations (40 all-acrocentric chromosomes) from Porto Santo (i.e., 10A and 13C). We found the greatest allele heterogeneity among the chromosomal races PADC ($2n = 24–28$) and PEDC ($2n = 23–26$), occurring in the western part of the island. In contrast, populations situated in eastern regions (PLDB, PPOD, PSAN, and PSVI) had a more homogeneous allele distribution (fig. 4A). Regarding standard populations ($2n = 40$) on the island of Porto Santo, the allele variability was lower than other populations from eastern Madeira. Among the alleles reported on Porto Santo, three of them (11X, 11V, and 11W) were private alleles that were undetected in Rb populations and two more (12C and 11F) were present in very low frequencies in the Madeira Island.
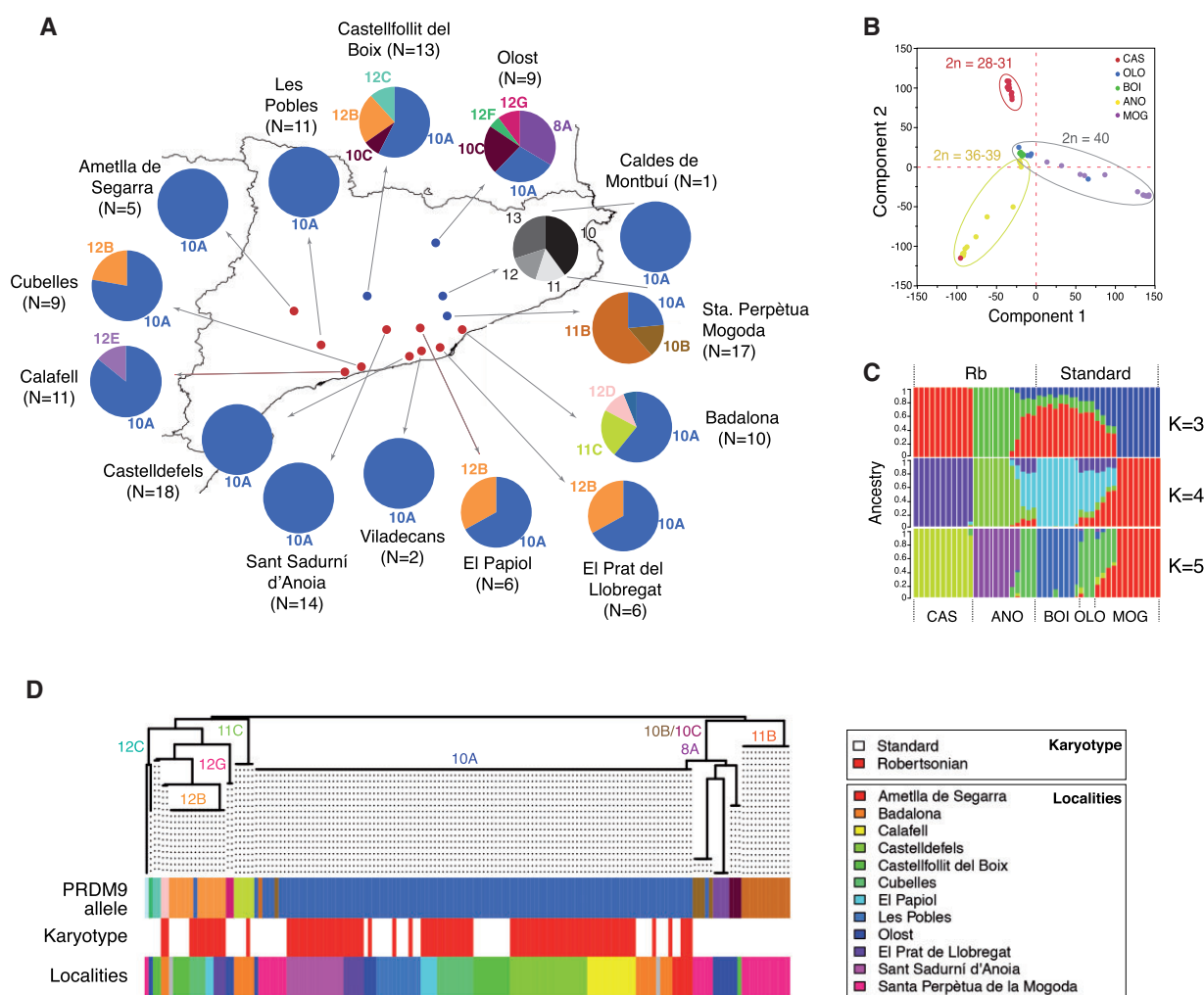
FIG. 3. Allelic frequencies, population structure and phylogeny of the ZnFs found in the Barcelona Rb system. (*A*) Representation of the PRDM9 allelic frequencies in each locality (see table 1 and supplementary table 1, Supplementary Material online, for further details on the chromosomal composition of each locality). Legend—*N*, number of specimens sequenced per locality. (*B*) PCA in a subset of 50 mice from 5 localities of the Barcelona Rb system (CAS, Castelldefels; OLO, Olost; BOI, Castellfollit del Boix; ANO, Sant Sadurní d'Anoia; and MOG, Sta. Perpètua de Mogoda). (*C*) Plots showing the proportion of inferred ancestry for *K* = 3 to *K* = 5 in a subset of 50 mice from 5 localities of the Barcelona system (CAS, Castelldefels; OLO, Olost; BOI, Castellfollit del Boix; ANO, Sant Sadurní d'Anoia; and MOG, Sta. Perpètua de Mogoda). (*D*) Phylogenetic tree including all mice sequenced from the Barcelona Rb system (see text for further details). PRDM9 alleles are color coded following the pattern used in panel (*A*). For each individual, both the karyotype (white: standard; red: Rb fusions) and the locality (see inserted color legend) are indicated.

Genetic Diversity and Signal of Positive Selection on the ZnF Array Domain

Once the PRDM9 diversity was characterized at a local scale, we analyzed genetic diversity and the signal of positive selection on the ZnF array domain at a global scale, including previously described Eurasian populations (Buard et al. 2014; Kono et al. 2014). When comparing the overall genetic (nucleotide) variability of the ZnF array domain found in all three regions included in the study (Barcelona, Madeira, and Eurasia), we observed that the genetic diversity in Madeira was greater than in the Barcelona Rb system and comparable to the diversity across the whole of Eurasia (table 2). Moreover, the genetic differentiation between Madeira and the Barcelona Rb system ($F_{ST} = 0.0295$, $P < 0.0001$) or even between Madeira and Eurasia was low ($F_{ST} = 0.0084$, $P < 0.0001$).

Understanding the evolutionary consequences of this ZnF array variability is important since polymorphisms in both number and sequence of the arrays might affect DNA-binding specificity resulting in changes in the recombination landscape of each individual (Davies et al. 2016). Thus, we analyzed whether selection played a role in shaping this variability by inferring dN/dS estimates averaged across all pairwise PRDM9 sequence alignments considering either hypervariable codons only (−1, +3, +6) or excluding them (see Material and Methods). We observed that the amino acids that recognize the specific DNA motif at positions −1, +3, and +6 are under strong positive selection, as shown by their extremely high dN/dS values (table 3). Position +6 showed less variability involving two of the residues described (Q and K) compared with positions −1 (A, E, Q, T, and V
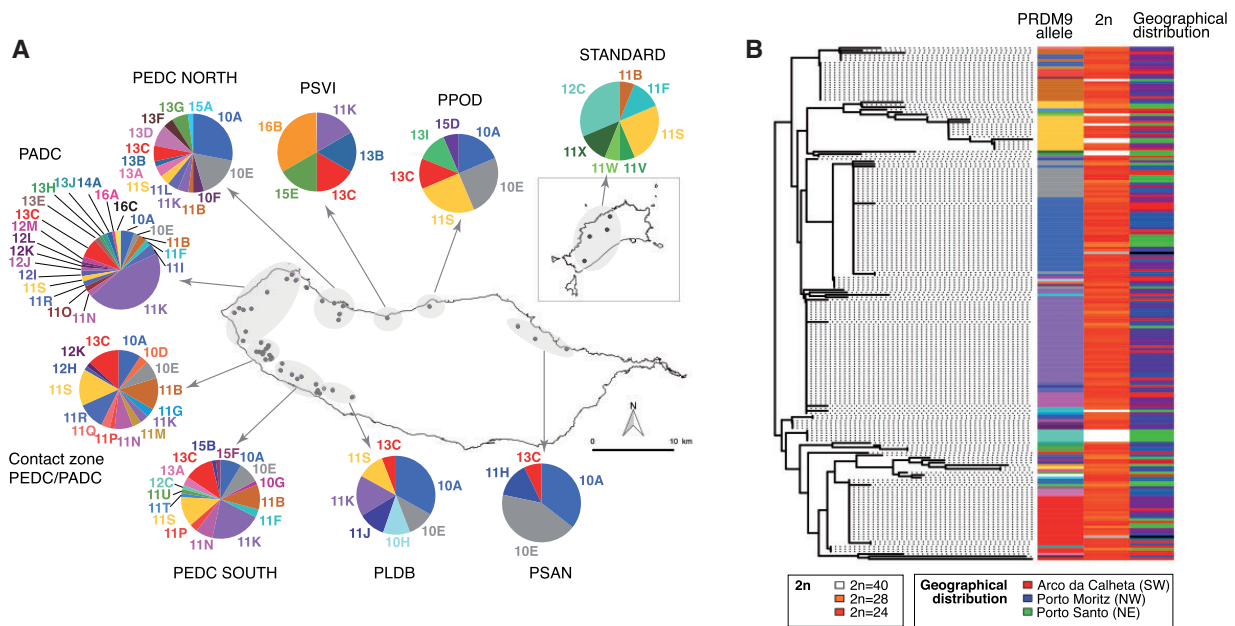
FIG. 4. Allelic frequencies and phylogeny of the ZnFs found in the Madeira archipelago. (*A*) Representation of the PRDM9 allelic frequencies found in each metacentric race from the Madeira Island and Porto Santo Island (see supplementary tables 2 and 3, Supplementary Material online, for further details on chromosomal composition of metacentric races). Number of specimens sequenced per locality on Madeira Island: Achadas da Cruz (*N* = 7), Arco da Calheta (*N* = 11), Chão da Ribeira (*N* = 16), Estreito da Calheta (*N* = 3), Faial (*N* = 2), Fajã da Ovelha (*N* = 2), Levada Grande (*N* = 2), Lombada dos Cedros (*N* = 1), Lombada dos Marinheiros (*N* = 4), Lombo da Velha (*N* = 3), Lombo das Laranjeiras (*N* = 3), Lombo do Doutor (*N* = 3), Lugar da Raposeira (*N* = 2), Madalena do Mar (*N* = 2), Maloeira (*N* = 18), Moledos (*N* = 2), Ponta Delgada (*N* = 8), Ponta do Pargo (*N* = 18), Ponta do Sol (*N* = 3), Porto da Cruz (*N* = 1), Porto Moniz (*N* = 3), Prazeres (*N* = 2), Ribeira da Janela (*N* = 6), Ribeira da Laje (*N* = 3), Ribeira da Vaca (*N* = 3), Ribeira Funda (*N* = 3), Santa (*N* = 1), Santana (*N* = 4), São Vicente (*N* = 3), Seixal (*N* = 1), Sítio da Fajã (*N* = 3), Socorro (*N* = 7), Solar da Maloeira (*N* = 1), Solar dos Prazeres (*N* = 1), and Porto Santo Island (*N* = 8). (*B*) Phylogenetic tree including all mice sequenced from the Madeira archipelago (see text for further details). PRDM9 alleles are color coded following the pattern used in panel (*A*). For each individual, both the diploid number (2*n*) and the geographical distribution are indicated (see inserted color legend).

substitutions) and +3 (D, H, N, V, and S substitutions) (fig. 2). Interestingly, the rest of the PRDM9 sequence (the so-called nonhypervariable sites) was similarly subjected to positive selection in both the Madeira the Barcelona Rb systems, whereas it was under weak negative selection across Eurasia (table 3). Moreover, we detected differences in dN/dS values between areas of study, with the effects of selection being pronounced in the Madeira archipelago, intermediate in the Barcelona Rb system and weaker across Eurasia (table 3). To test whether these contrasting patterns in positive selection across regions was explained by the presence of Rb fusions, we calculated dN/dS values considering mice with and without Rb fusions in the Madeira and Barcelona Rb systems. Remarkably, individuals with Rb fusions showed higher dN/dS values than standard ones, both in Madeira and the Barcelona Rb system (table 3). This was consistent for both hypervariable sites and the rest of the ZnF array domain (table 3).

Differences in Heterozygous PRDM9 Allele Combinations

Even though the PRDM9 protein can use all its ZnFs to bind to the DNA in a sequence-specific manner, it has been suggested that only a subset of ZnFs within the array contributes to the binding of the specific DNA motif

Table 2. Genetic Diversity Estimates of the ZnF Array Found in the Studied Areas (Madeira archipelago, Barcelona Rb system, and Eurasia).

	Madeira	Barcelona	Eurasia
Average genetic distance	5.406	2.307	6.123
Average pairwise nucleotide diversity excluding hypervariable sites (−1, +3, +6)	0.00456	0.002856	0.00416
Average pairwise nucleotide diversity (nucleotide differences per base)	0.02351	0.018626	0.02694

(Billings et al. 2013; Baker et al. 2015; Paigen and Petkov 2018). In the case of the house mouse, this hypothesis implies that the ZnF repeats located at positions ZnF3 to ZnF6 within the array would make primary contact with the DNA strand before recruiting the DSBs-induction machinery for the formation of recombination hotspots (review by Paigen and Petkov [2018]). Since the presence of PRDM9 heterozygous allelic combinations might result in subfertile phenotypes caused by asymmetric DSBs between homologous chromosomes, we sought to understand the functional constraints that might affect PRDM9 diversity in Barcelona and Madeira populations by analyzing the degree of conservation of positions ZnF3–ZnF6 within the array in PRDM9 heterozygous mice.

Table 3. Average dN, dS, and dN/dS Values between Aligned ZnF Arrays in the Studied Areas (Madeira archipelago, Barcelona Rb system, and Eurasia).

	Madeira				Barcelona				Eurasia	All
	All	St	Rb	St versus Rb	All	St	Rb	St versus Rb	All	
dN(hs)	0.181 (0.15–0.21)	0.232 (0.19–0.28)	0.175 (0.15–0.20)	0.225 (0.19–0.26)	0.089 (0.07–0.11)	0.128 (0.10–0.15)	0.0516 (0.03–0.07)	0.0951 (0.07–0.12)	0.218 (0.19–0.24)	0.169 (0.15–0.19)
dS(hs)	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)	0.00407 (0.001–0.008)	0.00086 (0.0001–0.0019)
dN(nhs)	0.0053 (0.004–0.007)	0.00929 (0.006–0.013)	0.00488 (0.003–0.006)	0.00817 (0.005–0.011)	0.00226 (0.001–0.004)	0.00298 (0.002–0.005)	0.00144 (0.0002–0.003)	0.00237 (0.001–0.004)	0.00362 (0.002–0.005)	0.00413 (0.003–0.006)
dS(nhs)	0.0006 (0.0003–0.001)	0.00433 (0.001–0.008)	0.00053 (0.0002–0.0009)	0.00158 (0.0006–0.003)	0.00028 (2.4^{-05}–0.0007)	0.00067 (9.1^{-05}–0.001)	0 (0–0)	0.00032 (1.1^{-05}–0.0008)	0.00289 (0.002–0.004)	0.00086 (0.0006–0.001)
dN(hs)/dS(hs)	∞ (∞–∞)	∞ (∞–∞)	∞ (∞–∞)	∞ (∞–∞)	∞ (∞–∞)	∞ (∞–∞)	∞ (∞–∞)	∞ (∞–∞)	53.6 (24.8–236)	197 (83–1110)
dN(hs)/dS(nhs)	278 (159–566)	53.7 (27.7–179)	330 (189–708)	142 (74.7–367)	313 (113–3180)	187 (78.5–1290)	∞ (∞–∞)	293 (100–8920)	75.3 (51.5–115)	196 (136–321)
dN(nhs)/dS(nhs)	8.12 (3.99–17.9)	2.15 (0.94–7.62)	9.2 (4.65–20.7)	5.16 (2.25–14.8)	7.95 (2.45–83.7)	4.34 (1.51–36.7)	∞ (∞–∞)	7.3 (2.2–182)	1.25 (0.75–1.97)	4.79 (2.82–8.13)

NOTE.—Ninety-five percent confidence intervals for all estimates are indicated in parentheses and italics. hs, hypervariable sites; nhs, nonhypervariable sites; Rb, individuals with Rb fusions; St, individuals with standard karyotype (i.e., absence of Rb fusions); and ∞, infinite values.

In the case of the Madeira system, we detected a clear deficit of heterozygotes with respect to Hardy–Weinberg equilibrium within Rb races, with PRDM9 heterozygotes reduced to about half of the expected (supplementary table 6, Supplementary Material online). These deviations from Hardy–Weinberg equilibrium were not found in standard mice in Porto Santo, suggesting a relationship between the presence of Rb fusions and the selection against PRDM9 heterozygotes on Madeira. A contrasting pattern was found in the Barcelona Rb system where both standard and Rb populations show a reduction in the number of heterozygote individuals for PRDM9 (supplementary table 6, Supplementary Material online).

As a way to explain this reduction in heterozygosity, we observed that not all heterozygous PRDM9 combinations were equally represented among populations (fig. 5 and supplementary table 7 and fig. 4, Supplementary Material online). The most common PRDM9 allele combinations in the Barcelona Rb system were 10A/12B, 10B/11B, and 10A/11C, whereas in the Madeira archipelago, most frequent alleles were 11K/13C, 10A/11K, and 10A/13C. Although in the Madeira archipelago no heterozygous combinations were shared between standard and Rb mice, only one allele combination (10A/12B) was present in both standard and Rb mice in the Barcelona Rb system (fig. 5A). In both Rb systems (Barcelona and Madeira), differences in population allelic frequency were significant between standard and Rb mice (Pearson test, $P \leq 0.05$). When analyzing the conservation in amino acid sequence of the ZnF repeats located at positions ZnF3–ZnF6 along the ZnF array in the different heterozygous combinations, we detected higher sequence conservation in Rb mice when compared with standard mice in both the Madeira and Barcelona Rb systems (fig. 5). In the case of the Barcelona Rb system, an average conservation of 94.4% was found in Rb mice versus 75% in standard mice, this difference being statistically significant (Wilcoxon test, $P = 0.0075$). These values were lower in the Madeiran archipelago, where the average sequence conservation of repeats from ZnF3–ZnF6 in Rb mice was 70.5% versus 44.4% in mice with no Rb fusions, though the differences in conservation were statistically significant (Wilcoxon test, $P = 0.0252$) (fig. 5 and supplementary fig. 4, Supplementary Material online).

Phylogeography of PRDM9

In order to investigate whether the observed variability of PRDM9 was reflective of population structure, we studied the phylogeography of the different alleles detected in the Rb systems individually (Barcelona and Madeira) and in relation to previously described alleles in Eurasian populations (Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014). Although previous allozyme and microsatellite analysis (Britton-Davidian et al. 2007; Förster et al. 2013) provided information regarding the origin of the chromosomal radiation of Madeiran mice, the population structure of the Barcelona system is not entirely known at this stage. Thus, we first determined genetic diversity and population differentiation in this system based on the single nucleotide

MBE
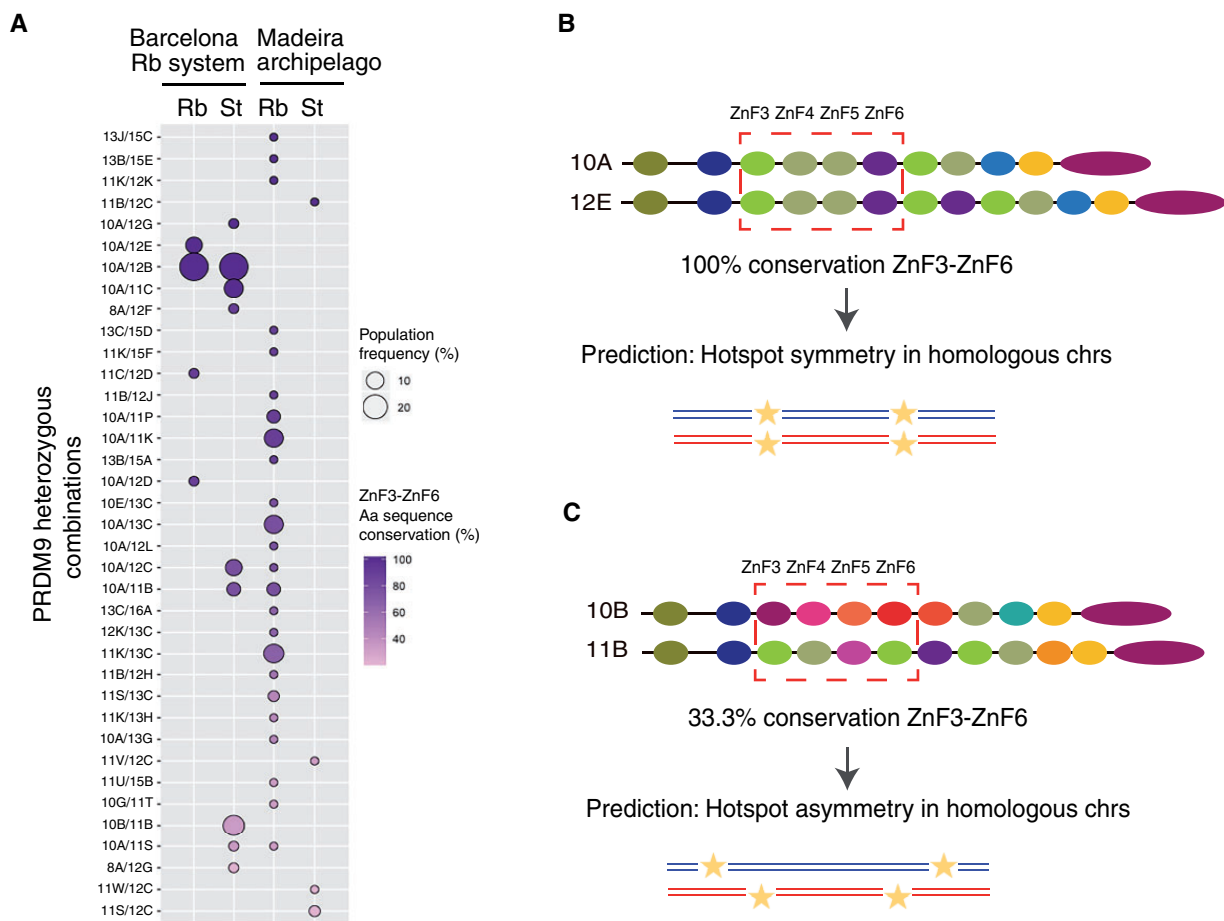
FIG. 5. Conservation in amino acid sequence of the ZnF repeats located in positions 3–6 (ZnF3–ZnF6) along the ZnF array in PRDM9 heterozygous combinations found in the study. (*A*) Bubble chart representing population frequency (size of the bubble) and the percentage of amino acid (Aa) sequence conservation of the highly variable positions (−1, +3, and +6) of the initial repeats along the array (from ZnF3 to ZnF6) (purple intensity). Data are shown for each heterozygous combination found in Standard and Rb mice in the Barcelona Rb system and the Madeira archipelago (see supplementary table 7, Supplementary Material online, for further details). (*B, C*) Examples of PRDM9 alleles in heterozygous combination and the predicted distribution of recombination hotspots. Yellow starts represent the location of recombination hotspots across homologous chromosomes (red and blue lines). (*B*) The ZnF repeats located in positions 3–6 (ZnF3–ZnF6) in the 10A/12E combination present 100% of Aa sequence conservation. This can result in symmetric distribution of recombination hotspot in homologous chromosomes. (*C*) The ZnF repeats located in positions 3–6 (ZnF3–ZnF6) in the 10B/11B combination present 33.3% of Aa sequence conservation. This can result in asymmetric distribution of recombination hotspot in homologous chromosomes.

polymorphism (SNP) genotyping of a subset of 50 mice using the Mega Mouse Universal Genotyping Array. We applied a maximum likelihood method to infer the genetic ancestry of each individual, where the individuals are assumed to have originated from K ancestral clusters. The results of the principal component analysis (PCA) and the plots for *K* = 3 to *K* = 5 are shown in figure 3*B and C*. We observed clustering for the two Rb populations (Castelldefels and Sant Sadurní d'Anoia) when compared with the standard populations (Olost, Castellfollit del Boix, and Santa Perpètua de Mogoda). In fact, the PCA revealed that Olost and Castellfollit del Boix populations, both located at the northern area of the Rb system distribution, clustered together. Compared with these, both Rb populations (Castelldefels and Sant Sadurní d'Anoia) showed progressive genetic

differentiation, being more pronounced in Castelldefels (fig. 3*B*), a population with the lowest diploid number (2*n* = 28–31) and with the presence of a high number of heterozygous Rb fusions (supplementary table 1, Supplementary Material online).

Likewise, estimation of population structure revealed heterogeneity of genetic clusters according to populations (fig. 3*C*). The genetic structure based on the ADMIXTURE analysis (Alexander et al. 2009) revealed that mice from Castelldefels (one of the populations with the highest number of Rb fusions detected) showed high levels of ancestry (0.99%), forming a cohesive cluster (fig. 3*C*). An exception to this pattern was a single individual (specimen 955), which shared ancestry with mice from Sant Sadurní d'Anoia. This individual was characterized by the highest diploid number in

Castelldefels (2n = 32) and by the absence of Rb fusion 3.18, characteristic of the remaining mice from this population (supplementary table 1, Supplementary Material online). Within the Sant Sadurní d'Anoia population (2n ranging from 35 to 39), mice were distributed into two different groups, with three individuals sharing ancestry with the standard populations Olost and Castellfollit, both populations highly homogenous according to the PCA. The Santa Perpètua Mogoda population was very unusual, despite its standard condition (i.e., the absence of Rb fusions), it was highly differentiated from Olost and Castellfollit del Boix in both the PCA and the estimates of individual ancestry.

We then constructed phylogenetic trees for both the Madeira and Barcelona Rb systems by the Neighbor-Joining approach using a new measure of pairwise genetic divergence (see Material and Methods). The key feature of our approach is that it considers point mutations, insertions, and deletions of whole repeat units, as well as duplications of single repeat units as a consequence of nonhomologous recombination, slippage, or related biological processes. This is of relevance since single-unit duplications and indels have different weights in their contribution to the genetic distance. This approach allowed us to effectively estimate genetic distances among the polymorphic repeats of PRDM9 detected in the Madeira and the Barcelona Rb systems. The distribution of haplotypes obtained showed greater diversification on the island of Madeira than the Barcelona system (figs. 3D and 4B). Although the phylogenetic analysis grouped PRDM9 alleles into three major phylogroups in the Barcelona system (those representing the alleles 10A, 11B, and 12B; fig. 3D), the allelic diversification was much higher in the Madeira Island with alleles 10A, 11B, 11S, 11K, and 13C representing the most common haplotypes (fig. 4B). Differences between both Rb systems were also exemplified by higher genetic diversity estimates of the ZnF array on Madeira (5.406) than in the Barcelona Rb system (2.307), the former being comparable to the diversity across the whole Eurasia (6.123) (table 2). When plotting the chromosomal configuration in the phylogenetic trees we detected that in both Rb systems (Madeira and Barcelona), standard mice showed a tendency to carry private alleles. This was exemplified by alleles 8A, 10B, 10C, 11B, 12C, 12G, and 12F in Olost, St. Perpètua de Mogoda, and Castellfollit del Boix (fig. 3D), and 11W and 11X in Porto Santo (fig. 4A).

Taking advantage of previous surveys of wild captured mice in Europe and Asia, we further compared the allele variation found in the Madeira and Barcelona Rb systems with those described in Eurasia (fig. 6). When the geographical distribution was plotted onto the global phylogenetic tree, Madeira haplotypes showed mixed origins. The phylogenetic reconstruction of all three sampled regions (and informed by evidence that the genetic differentiation between Madeira and the Barcelona Rb system or even Eurasia is small, table 2) suggests that Madeira was colonized by a highly different population, most likely as a result of multiple colonization/introgression events from multiple parts of Eurasia.

Discussion

Distinctive Phylogeographic Patterns of ZnF Variability and the Evolution of PRDM9 in the House Mouse across Eurasia

Our study includes a wide survey of nearly 400 wild-caught mice representing two highly distinct house mouse Rb systems: the Madeira archipelago (Madeira and Porto Santo Islands) and the Barcelona system (continental). We identified 49 newly described alleles on the Madeira archipelago, and 8 new alleles in the continental (Barcelona) Rb system, revealing that intraspecific PRDM9 diversity in *M. m. domesticus* is far greater than previously reported. This adds substantially to previously described allelic diversity in natural populations of *Mus musculus* subspecies (28 distinct PRDM9 alleles in *M. m. domesticus*, 34 in *Mus musculus musculus*, and 37 in *M. m. castaneus*) (Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014).

Most importantly, we detected contrasting evolutionary patterns in the continental versus island chromosomal races of the house mouse. This was reflected by unprecedented high levels of PRDM9 diversity in Madeiran house mice, an insular Rb system characterized by distinct metacentric races, when compared with the Barcelona Rb system (where Rb fusions are not yet fixed within populations and are present in polymorphic state). Our analysis of genetic diversity and positive selection, together with the phylogeographic reconstruction of PRDM9, suggests that the variability observed could be the result of, at least, two possible (not mutually exclusive) scenarios in the Madeira system: 1) the current populations of Rb mice could reflect a complex pattern of introgression or multiple colonization events into the island and 2) meiotic impairment on hybridization of different populations of Rb mice allowed PRDM9 alleles to diverge among metacentric populations.

Initial studies based on mtDNA lineages (Bonhomme et al. 2011) indicated a recent common ancestry of all extant house mice populations, as well as a complex history owing to founder events, genetic drift and secondary admixture. This was suggestive of two phases of mouse colonization westward across the Mediterranean basin; an initial event during the early progression of Neolithic human expansion (starting 12,000 years ago) followed by a second, more recent (some hundred years ago), related to maritime trade. This progression can explain the presence of Mediterranean alleles in the Barcelona system and subsequent in situ diversification. Madeira, on the other hand, has a more recent history of mouse colonization originated first from Danish Vikings in the 9th century followed by a second incursion by 15th century Portuguese settlers (Gündüz, Auffray, et al. 2001; Britton-Davidian et al. 2007; Förster et al. 2009, 2013). Our analyses of genetic diversity reveal a complex colonization pattern in Madeira, much more so than in the Barcelona Rb system and leading to a PRDM9 diversity comparable to the diversity across the whole of Eurasia. Additionally, the phylogenetic reconstruction of PRDM9 in Madeira indicated that the same group of alleles appeared in different clusters independently. Altogether, our results suggest multiple colonization/

MBE
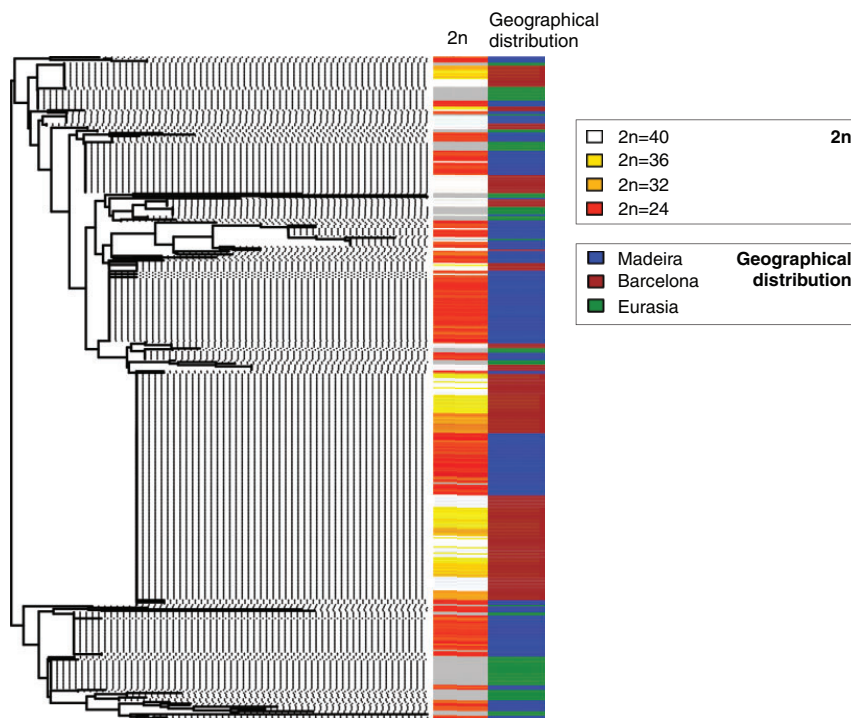
Fig. 6. Phylogeographic depiction of PRDM9 at a global scale. Phylogenetic reconstruction including mice sequenced from the Madeira archipelago, Barcelona Rb system and Eurasia (data extracted from Kono et al. [2014] and Buard et al. [2014]). For each individual, both the diploid number (2n) and the geographical distribution (Madeira, Barcelona, and Eurasia) are indicated (see inserted legends). The diploid number is not known for Eurasian samples so they are represented in gray.

introgression events into Madeira from Eurasia (including the Mediterranean basin).

Under this scenario of multiple colonization events and together with the extreme topography of Madeira (i.e., chromosomal races separated by mountain barriers), the presence of Rb fusions might have resulted in meiotic impairment in hybrids that subsequently contributed to the high rate of diversification of PRDM9. Our analysis of genetic diversity and positive selection supports this view. Since chromosomal fusions can act as postzygotic barriers and can restrict gene flow among chromosomal races (Franchini et al. 2010; Giménez et al. 2013; Capilla et al. 2014), this can result in high genetic differentiation between populations, thus facilitating the appearance of new PRDM9 variants. In fact, in the case of Madeira, each chromosomal race is geographically localized and homogeneous with respect to the constitution of Rb fusions, and chromosomal hybrids are rare (Britton-Davidian et al. 2000; Nunes et al. 2005). On the other hand, Madeira is extremely variable regarding its genetic structure, as can be appreciated both from the PRDM9 phylogenetic tree and from the high genetic differentiation between chromosomal races. Moreover, the relative deficit of heterozygous for PRDM9 (despite the strong genetic mixing) points to strong selection for karyotype compatibility. In this context, differences in positive selection can be explained by the presence of Rb fusions since dN/dS values were

higher in Rb mice than standard mice, both in Madeira and the Barcelona Rb systems. This was true both for hypervariable codons and for the rest of PRDM9 sequence.

An association between Rb fusions and PRDM9 diversity was also observed in the Barcelona Rb system, although the effects were mild, probably related to a differential demographic history. The Barcelona system represented approximately 10% of the novel PRDM9 alleles found in our study, with the highest diversity observed in standard mice located in the northern-eastern distribution of the sampled populations. The allelic distribution together with the pattern of genetic diversity detected by the PCA and ancestry analyses was suggestive of a larger connectivity among populations in the Barcelona Rb system than in the case of Madeira. This was exemplified by the 10A allele, which was present in all populations especially in Rb mice, representing the greatest allele homogeneity (nearly 95% of mice from Rb populations carried 10A). Considering the factors that might be responsible for the widespread distribution of this allele in the Barcelona system, it is possible that the PRDM9 allele frequency pattern observed in this system could result from a population bottleneck that facilitated the formation and subsequent expansion of Rb fusions. According to the PCA and ancestry analysis, standard mouse populations within the Barcelona system, such as Olost and Castellfollit del Boix, may represent the origin of the genetic diversity detected out of which Rb populations such as Gavà and Sant Sadurní

MBE

d'Anoia subsequently appeared as differentiated populations. Consequently, Rb mice may have diverged into different, genetically distinct populations resulting in a homogeneous distribution of PRDM9 alleles with a predominance of the 10A allele. This is consistent with suggestions by Gündüz, López-Fuster, et al. (2001) for the origin of the Barcelona system by primary intergradation. This Rb system may therefore represent an example of a radiation that could eventually lead to the formation of a metacentric race without geographic isolation (Sans-Fuentes et al. 2010).

Functional Implications of Maintaining Conserved DNA-Binding Repeats

Despite the ZnF variability observed in both Rb systems, we detected significant deviations from Hardy–Weinberg equilibrium (i.e., a deficit of heterozygotes) where the PRDM9 heterozygous combinations most frequently encountered maintained amino acid sequence similarity from position ZnF3 to ZnF6 in the ZnF array. These results suggest that observed PRDM9 variability (either as a result of phylogeographic dynamics and/or population isolation triggered by meiotic impairment of hybrids), can be subjected to functional constraints that facilitate the accumulation of allelic combinations that maintain recombination hotspot symmetry.

Under this scenario, it is important to take into account the particular features of the PRDM9 protein (Davies et al. 2016; Tiemann-Boege et al. 2017; Paigen and Petkov 2018), which include 1) its DNA-binding properties, 2) the rapid coevolution of the ZnF domain and the DNA motif that it recognizes, 3) asymmetric DSBs due to preferential binding of PRDM9 to new target sequences, and 4) hybrid sterility triggered by asymmetric DSBs and interallelic PRDM9 combinations. Previous work on house mouse inbred strains found that male sterility is a complex process that can be influenced by the heterozygous allelic combination of PRDM9 (Dzur-Gejdosova et al. 2012; Gregorová et al. 2018) and their interaction with different X-linked and autosomal loci (Balcova et al. 2016). Likewise, the presence of highly divergent PRDM9 alleles in a heterozygous state can be detrimental due to elevated levels of asymmetric DSBs between homologous chromosomes (Davies et al. 2016). This is especially important since allelic combinations of PRDM9 with conserved DNA-binding repeats were maintained in metacentric populations, suggestive of the presence of functional constraints.

In the light of our findings, the rapid formation of new PRDM9 alleles in the presence of Rb fusions, can act synergistically with the formation of new binding motifs distributed across the genome in different metacentric populations. This could result in functional constraints that might facilitate the accumulation of PRDM9 heterozygous combinations that maintain a high degree of similarity in the amino acids responsible for protein–DNA binding, thus resulting in recombination hotspot symmetry in homologous chromosomes. Mounting evidence in humans and mice has shown that variation in both the sequence and number of ZnF repeats strongly influence the distribution of meiotic DSBs (Berg et al. 2010; Grey et al. 2011; Brick et al. 2012; Baker et al. 2015). Small

variations in the amino acid sequence of ZnF repeats can modulate PRDM9 binding affinities to specific DNA motifs (Berg et al. 2010). In fact, differences in a single ZnF repeat can affect the specificity of over 70% of the meiotic DSBs, leading to the redistribution of recombination sites in a single generation (Smagulova et al. 2016). The rapid evolution of ZnF sequences would therefore lead to rapid changes in the distribution of recombination sites across the genome by creating new DNA motifs that would be recognized with stronger affinity by the new PRDM9 allelic variants. This will, in turn, compensate the loss of the primary DNA motifs by gene conversion during the repairing of crossovers during the early stages of meiosis (Smagulova et al. 2016).

Although little is known at this stage about the dynamics and the mechanistic constraints that affect recombination landscapes at a finer scale in metacentric wild populations, our results nonetheless suggest that Rb fusions might play a role in bringing new selective pressures on PRDM9. These outcomes provide impetus to examine the effect of genetic backgrounds on PRDM9 and the development of reproductive isolation between natural populations.

Materials and Methods

Sampling and DNA Extraction

Our data set comprised a total of 395 wild-caught house mice, 185 mice from the Barcelona Rb system (Medarde et al. 2012), 199 mice from Madeira Island, and 11 from Porto Santo Island (table 1, fig. 1, and supplementary tables 1 and 2, Supplementary Material online). Mice were karyotyped previously (Medarde et al. 2012; Chmátal et al. 2014, supplementary fig. 1, Supplementary Material online). Animals were housed and treated in strict accordance with ethical guidelines approved by the Universitat Autònoma de Barcelona (Spain) and University of Lisbon (Portugal). Genomic DNA was extracted from tissue biopsies preserved in absolute ethanol or fresh tissue using proteinase K digestion (Sambrook et al. 1989).

The Barcelona Rb system included mice from 15 localities selected for their geographical distribution and distinctive chromosomal configurations (fig. 1 and supplementary table 1, Supplementary Material online). This covered the full extent of the chromosomal polymorphism previously described for the Barcelona Rb system (Medarde et al. 2012). This included localities with high frequencies of almost all metacentric chromosomes ($2n = 28–39$), localities in the vicinity of Rb areas but without chromosomal fusions ($2n = 40$), and localities geographically located between these population groups containing intermediate diploid numbers ($2n = 39–40$) (supplementary table 1, Supplementary Material online). Four of the localities considered in the study had mice with standard karyotypes ($2n = 40$, Castellfollit del Boix, Santa Perpètua de Mogoda, Olost, and Caldes de Montbuí) and defined here as standard populations (supplementary table 1, Supplementary Material online). The remaining localities ($N = 11$) included mice with Rb fusions, defined here as Rb populations ($2n = 28–39$, Badalona Les Pobles, L'Ametlla de Segarra, El Papiol, Calafell, Sant Sadurní d'Anoia, Cubelles, La Granada, El Prat de

MBE

Llobregat, Viladecans, and Castelldefels). Mice from Rb populations were characterized by having between one to seven Rb fusions involving 14 different chromosomes (Rb[3.8], Rb[4.14], Rb[5.15], Rb[6.10], Rb[7.17], Rb[9.11], and Rb[12.13]), either in heterozygous or homozygous states (supplementary table 1, Supplementary Material online). The Rb fusions were present as chromosomal polymorphisms in the Barcelona Rb system, with varying frequencies (supplementary table 1, Supplementary Material online).

The Madeira system was represented by mice from 34 localities that included all six chromosomal races occurring on the island of Madeira (as originally defined by Britton-Davidian et al. [2000]) (supplementary table 2, Supplementary Material online): PSAN ($2n = 22$), PADC ($2n = 24–28$), PEDC ($2n = 23–26$), PLDB ($2n = 24$), PPOD ($2n = 27–28$), and PSVI ($2n = 26–27$). A total of 20 different metacentrics in homozygous state are described in Madeira, 9 of which have not been detected elsewhere (Rb[2.19], Rb[4.5], Rb[4.16], Rb[5.18], Rb[9.18], Rb[11.12], Rb[11.19], Rb[13.17], and Rb[15.18]), highlighting the uniqueness of this system (see supplementary table 3, Supplementary Material online, for details on the chromosomal composition of chromosomal races). Mice with the standard karyotype ($2n = 40$) from the neighboring island of Porto Santo (four localities) were also included in our investigation (fig. 1 and supplementary table 2, Supplementary Material online).

Prdm9 Sequencing

The entire ZnF array is encoded by the last exon of the *Prdm9* gene that extends from the second ZnF repeat towards the C-terminal domain. We amplified this region in all mice by PCR using the primers ZFA-L_F (forward) and ZFA-L_R (reverse) described by Kono et al. (2014) (supplementary table 4, Supplementary Material online) using the ExTaq (TaKaRa) following the protocol in Capilla et al. (2014). Briefly, the PCR conditions were 95 °C (3 min), 30 cycles of 95 °C (30 s), 64 °C (30 s), and followed by 72 °C (90 s). The ZnF array exhibits not only length polymorphisms (variation of the number of ZnF repeats) but also amino acid variation (alleles with the same number of ZnF repeats but different amino acid composition). As a consequence, it is necessary in each case to identify the length of the allele and subsequently the amino acidic composition of the sequence. To this end, PCR products were separated on 1% agarose gels. This allowed both the ZnF array length to be discerned (homozygous and heterozygous) and optimal DNA recovery from bands for subsequent sequencing (supplementary fig. 2, Supplementary Material online).

To determine the sequences of ZnF repeats, PCR products of the homozygous mice were subjected to bidirectional Sanger sequencing with the same primers used for amplification: primers ZFA-L_F and ZFA-L_R. Only sequences reproducible in at least three independent amplification reactions were included in the data analyses. Bands of heterozygous mice were purified from the gels using the Nucleospin Gel and PCR Clean-up kit (Macherey-Nagel) and subsequently sequenced. Sequences from heterozygous mice were analyzed using the Bioedit 7.2.5 package. Two different software

packages were used to distinguish allelic variants and to define haplotypes of heterozygous mice: Phase (http://stephenslab.uchicago.edu/software.html#phase) and Champuru v1.0 (http://www.mnhn.fr/jfflot/champuru/). This allowed inferences on whether a sample was homozygous or heterozygous for its amino acid composition (Capilla et al. 2014). All allele sequences were translated using ExPasy: SIB bioinformatics resource portal (Artimo et al. 2012).

Additionally, we included 118 published PRDM9 sequences of wild-derived *M. m. domesticus* and *M. m. castaneus* specimens drawn from various locations in Eurasia (Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014) for the phylogeographic analysis of PRDM9 at a global scale. Overall, the phylogenetic comparative study included three *M. m. domesticus* sampling regions: the Barcelona Rb system, the Madeira archipelago (including both Madeira and Porto Santo Islands) and Eurasia (previously published data; Buard et al. 2014; Kono et al. 2014) (table 1).

PRDM9 Allele Classification

Following the nomenclature used in previous studies (Buard et al. 2014; Capilla et al. 2014; Kono et al. 2014), PRDM9 alleles were classified using the number of ZnF repeats and the extent of amino acid variation in the highly variable positions -1, $+3$, and $+6$ of each ZnF repeat (table 1 and supplementary table 5, Supplementary Material online). Each ZnF repeat was identified by a number (from #3 to #34) based on its amino acid sequence (supplementary table 5, Supplementary Material online). Subsequently, the combination of different of ZnF repeats was classified as distinct PRDM9 alleles (fig. 2). Moreover, for each PRDM9 allele found in heterozygous state, the amino acid conservation of the ZnF repeats located in positions ZnF3–ZnF6 of the array was calculated based on the hypervariable sites (-1, $+3$, and $+6$; Kono et al. 2014) (supplementary table 6, Supplementary Material online).

SNP Genotyping

SNP genotyping data included 50 mice from 5 localities of the Barcelona Rb system: 26 mice from three standard populations (Castellfollit del Boix, $N = 8$; Santa Perpètua de Mogoda, $N = 13$; and Olost, $N = 5$) and 24 mice from two Rb populations (Sant Sadurní d'Anoia, $N = 12$ and Castelldefels, $N = 12$). Genomic DNA was extracted using a standard protocol with proteinase K (Sambrook et al. 1989). Subsequently, mice were genotyped using the Mega Mouse Universal Genotyping Array (Morgan et al. 2015), which consist of 77,808 evenly distributed SNP markers built on the Illumina Infinium platform. SNPs were filtered to remove markers with missing values $>5\%$ threshold (i.e., markers that do not fit Hardy–Weinberg expectations, $P < 10^{-5}$) using PLINK version 1.9 (Purcell et al. 2007). The proportion of missing data and heterozygosity per locus and per sample were also calculated to evaluate possible bias. This resulted in a final data set of 63,344 informative SNPs distributed across all chromosomes, with the exception of the Y. Using this data set, the ADMIXTURE software (Alexander et al. 2009) was used to estimate individual ancestry and admixture proportions assuming K populations based on a maximum likelihood

method. Analyses were run only for SNPs with a greater than 95% genotype call. The numbers of clusters (K) evaluated here ranged from 1 to 10. To further evaluate the final K value, an Evanno's ΔK was applied (Evanno et al. 2005). Subsequently, three different K values ($K = 3, 4, 5$) showed the lowest likelihood values (supplementary fig. 3, Supplementary Material online). ADMIXTURE analyses were plotted using an R framework. In addition, PCA was implemented using a module of PLINK 1.9.

Phylogenetic Analysis of PRDM9 Alleles

Phylogenetic trees for the nucleotide sequences of PRDM9 alleles were built via a Neighbor-Joining approach informed by a new measure of pairwise genetic divergence between ZnF repeats. Briefly, divergence between pairs of ZnF repeats was computed by first masking codons corresponding to the hypervariable amino acid positions ($-1, +3, +6$). All repeat units share a high sequence similarity and have the same length, it was therefore straightforward to align these units and compute the pairwise genetic (Hamming) distances $\mu(\upsilon, \upsilon')$ between any two units υ and υ'. Then, the evolution of the ZnF repeats was modeled as a Markov process involving nucleotide mutation, insertion and deletions of a single repeat unit as well as single-unit duplication/slippage. Under this framework, we defined the genetic distance between repeats as an edit distance, that is, as the minimum cost to change a repeat to another through a series of elementary operations. The edit distance is minimized by one or possibly several alignments between repeats $r = (r_1; r_2; r_3; \ldots)$ and $r' = (r'_1; r'_2; r'_3; \ldots)$. The cost was defined as the weighted sum of all contributing elements: 1) point mutations, small indels, and other within-unit processes (cost $\mu[r_j, r'_{j'}]$ for an alignment between the jth and the jth units of the two repeats, weight w_m); 2) insertions and deletions of whole units (cost 1, weight w_i); and 3) single-unit duplication/slippage (cost 1, weight w_s; plus an extra cost $\mu[r_j, r_{j-1}]$ or $\mu[r'_{j'}, r'_{j'-1}]$ for the point mutations and small indels between neighboring duplicated units, which was weighted by w_m). An R implementation of the algorithms described here is provided at https://github.com/lucaferretti/RepeatDistance.

Subsequently, each pair of ZnF repeats was aligned using a modified Needleman–Wunsch algorithm corresponding to the model considered, with different costs for each type of mutations (w_m for point mutations, w_i for insertions/deletions, and w_s for slippage). The genetic distance between two sequences was defined as the weighted sum of the cost of each mutation between the sequences. Our selection of costs was as follows: 1) Mismatch between units r_{ju} and $r'_{j'}$: $w_m \mu(r_j, r'_{j'})$ and 2) insertion of a whole unit r_j in repeat r: $\min[w_i, w_s + w_m \mu(r_j, r_{j-1}), w_s + w_m \mu(r'_{j'}, r'_{j'-1})]$. The best alignment according to this choice of costs corresponds precisely to the alignment that minimizes the sum of costs of mutations between sequences; hence the minimum cost obtained from the Needleman–Wunsch algorithm provides the distance between repeats defined above.

Weight parameters were chosen among the ones that maximized the agreement between mean genetic distances in the subset of 50 mice from the Barcelona system included

in the SNP genotyping (computed from the 20 Mb around the PRDM9 nucleotide sequence) and the mean ZnF genetic distance for the same individuals. Agreement between the genetic distances was defined in terms of Pearson correlation of the pairwise distances of all individuals for which both genotype and PRDM9 sequence data were available. The correlation was computed numerically for a grid of weight values with a step of 0.25. In case of ties, the combination with the lowest ratios w_i/w_m and w_s/w_m was chosen. The final choice of weights maximizing the agreement between genetic distances is $w_m = 1$ for nucleotide point mutations, $w_i = 3.5$ for unit insertions/deletions, and $w_s = 1.75$ for single-unit duplications. All pairs of sequences were pairwise aligned with our approach, using the genetic distance corresponding to these weights.

The PRDM9 phylogenetic trees were reconstructed from the molecular distance discussed above, using the Neighbor-Joining method implemented in the R library APE (Paradis et al. 2004). Trees were rooted at midpoint using the R package phangorn (Schliep 2011).

Genetic Diversity and Positive Selection

The PRDM9 alleles were tested for deviations from Hardy–Weinberg proportions in all populations with more than ten sequenced individuals. The significance of the deficit of heterozygous individuals for PRDM9 alleles with respect to Hardy–Weinberg predictions was assessed by an exact permutation test based on 1,000,000 permutations.

Pairwise nucleotide diversity was estimated as the mean number of nucleotide differences per base averaged across all pairs of aligned nucleotide sequences, ignoring gaps in the pairwise sequence alignment. F_{ST} values between multiple population/chromosomal races were computed using the formula described by Hudson et al. (1992). This computed the within-population diversity as the average of the within-population diversities of all population (in order to weight the contribution of each population equally) and permitted the removal of hypervariable codons from the sequence.

Selection was inferred from dN/dS estimates averaged across all pairwise sequence alignments as follows. First, we computed all the pairwise alignments using the Needleman–Wunsch algorithm described before. For each pair, we estimated dN as the number of nonsynonymous mutations per nonsynonymous site, considering either hypervariable codons only ($-1, +3, +6$) or excluding them, and ignoring gaps in the pairwise alignment. The fraction of nonsynonymous sites was estimated by randomization conditional on the nucleotide frequencies of the sequence. Then, dS was similarly estimated as the number of synonymous mutations per synonymous site (since there were often no synonymous mutations in hypervariable codons, we used synonymous mutations in the other codons as an alternative estimate of neutral rates). Finally, we averaged dN and dS across all sequence pairs and estimated dN/dS as the ratio of these averages. This corresponds to the approximate Maximum Likelihood estimate of dN/dS for Poisson-distributed mutations. Neutral evolution would result in an expected value of dN/dS around 1. The 95% confidence intervals were

MBE

computed by bootstrapping (sampling 500 random sets of codons with replacement). We computed dN/dS both across all sequences and within/between specific subsets (such as sequences from individuals with standard karyotype or with Rb fusions). All analyses of genetic diversity were implemented via custom R scripts.

Data Accessibility

Data underlying this article are available on GeneBank (from MK848086 to MK848149).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Author Contributions

A.R.-H. conceived and devised the study. C.V., L.C., L.F., A.L., R.A.S.-G., S.I.G., M.L.M., J.B.S., and A.R.-H. contributed to the design of the methodological approaches. C.V., L.C., R.A.S.-G., B.F.-S., J.B.-R., and A.R.-H. contributed to PRDM9 sequencing experiments and allele classification. C.V., G.A.-L., L.F., and A.R.-H. contributed to SNP genotyping and phylogenetic analysis of PRDM9 alleles. L.F. and A.L. developed the approach for the computation of genetic distances between complex repeats. S.I.G., J.V., J.B.S., M.L.M., and A.R.-H. contributed to reagents and data collection. C.V., L.C., and A.R.-H. wrote the first draft of the manuscript with contributions of all authors. All authors read and approved the final version of the manuscript.

References

Adolph S, Klein J. 1981. Robertsonian variation in *Mus musculus* from Central Europe, Spain, and Scotland. *J Hered.* 72(3):219–221.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.

Artimo P, Jonnalagedda M, Arnold K, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, et al. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40:597–603.

Baker CL, Kajita S, Walker M, Saxl RL, Raghupathy N, Choi K, Petkov PM, Paigen K. 2015. PRDM9 drives evolutionary erosion of hotspots in *Mus musculus* through haplotype-specific initiation of meiotic recombination. *PLoS Genet.* 11(1):e1004916.

Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, Przeworski M. 2017. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife* 6:e24133.

Balcova M, Faltusova B, Gergelits V, Bhattacharyya T, Mihola O, Trachtulec Z, Knopf C, Fotopulosova V, Chvatalova I, Gregorova S, et al. 2016. Hybrid sterility locus on chromosome X controls meiotic recombination rate in mouse. *PLoS Genet.* 12(4):e1005906.

Barbash DA, Siino DF, Tarone AM, Roote J. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc Natl Acad Sci U S A.* 100(9):5302–5307.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2010. Prdm9 is a major determinant of meiotic recombination hotspots in human and mice. *Science* 21:836–840.

Berg IL, Neumann R, Lam K-W, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ. 2010. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat Genet.* 42(10):859–863.

Bhattacharyya T, Gregorova S, Mihola O, Anger M, Sebestova J, Denny P. 2013. Mechanistic basis of infertility of mouse intersubspecific hybrids. *Proc Nat Acad Sci U S A.* 110:468–477.

Billings T, Parvanov ED, Baker CL, Walker M, Paigen K, Petkov PM. 2013. DNA binding specificities of the long zinc-finger recombination protein PRDM9. *Genome Biol.* 14(4):R35.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffray J-C, Britton-Davidian J. 2011. Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. *Proc Biol Sci.* 278(1708):1034–1043.

Boulton A, Myers RS, Redfield RJ. 1997. The hotspot conversion paradox and the evolution of meiotic recombination. *Proc Natl Acad Sci U S A.* 94(15):8058–8063.

Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485(7400):642–645.

Britton-Davidian J, Catalan J, Lopez J, Ganem G, Nunes AC, Ramalhinho MG, Auffray JC, Searle JB, Mathias ML. 2007. Patterns of genic diversity and structure in a species undergoing rapid chromosomal radiation: an allozyme analysis of house mice from the Madeira archipelago. *Heredity* 99(4):432–442.

Britton-Davidian J, Catalan J, Ramalhinho MG, Ganem G, Auffray J-C, Capela R, Biscoito M, Searle JB, Mathias ML. 2000. Rapid chromosomal evolution in island mice. *Nature* 403(6766):158.

Buard J, Rivals E, Dunoyer de Segonzac D, Garres C, Caminade P, de Massy B, Boursot P. 2014. Diversity of Prdm9 zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite. *PLoS One* 9(1):e85021.

Capilla L, Garcia Caldés M, Ruiz-Herrera A. 2016. Mammalian meiotic recombination: a toolbox for genome evolution. *Cytogenet Genome Res.* 150(1):1–16.

Capilla L, Medarde N, Alemany-Schmidt A, Oliver-Bonet M, Ventura J, Ruiz-Herrera A. 2014. Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and Prdm9 allelic background. *Proc R Soc B* 281:1–18.

Chmátal L, Gabriel SI, Mitsainas GP, Martínez-Vargas J, Ventura J, Searle JB, Schultz RM, Lampson MA. 2014. Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr Biol.* 24(19):2295–2300.

Coop G, Myers SR. 2007. Live hot, die young: transmission distortion in recombination hotspots. *PLoS Genet.* 3:377–386.

MBE

Coyne JA, Orr HA. 1998. The evolutionary genetics of speciation. *Philos Trans R Soc Lond B Biol Sci.* 353(1366):287–305.

Davies AB, Hatton E, Altemose N, Hussin JG, Pratto F, Zhang G, Hinch AG, Moralli D, Biggs D, Camerini-Otero RD, et al. 2016. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* 530(7589):171–176.

Dzur-Gejdosova M, Simecek P, Gregorova S, Bhattacharyya T, Forejt J. 2012. Dissecting the genetic architecture of F_1 hybrid sterility in house mice. *Evolution* 66(11):3321–3335.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14(8):2611–2620.

Faria R, Navarro A. 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol Evol (Amst).* 25(11):660–669.

Farré M, Micheletti D, Ruiz-Herrera A. 2013. Recombination rates and genomic shuffling in human and chimpanzee—a new twist in the chromosomal speciation theory. *Mol Biol Evol.* 30(4):853–864.

Flachs P, Mihola O, Šimeček P, Gregorová S, Schimenti JC, Matsui Y, Baudat F, de Massy B, Piálek J, Forejt J, et al. 2012. Interallelic and intergenic incompatibilities of the Prdm9 (Hst1) gene in mouse hybrid sterility. *PLoS Genet.* 8(11):e1003044.

Förster DW, Günduz İ, Nunes AC, Gabriel S, Ramalhinho MG, Mathias ML, Britton-Davidian J, Searle JB. 2009. Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Mol Ecol.* 18(21):4477–4494.

Förster DW, Mathias ML, Britton-Davidian J, Searle JB. 2013. Origin of the chromosomal radiation of Madeiran house mice: a microsatellite analysis of metacentric chromosomes. *Heredity (Edinb).* 110(4):380–388.

Franchini P, Colangelo P, Solano E, Capanna E, Verheyen E, Castiglia R. 2010. Reduced gene flow at pericentromeric loci in a hybrid zone involving chromosomal races of the house mouse *Mus musculus domesticus*. *Evolution* 64:2020–2032.

Giménez MD, White TA, Hauffe HC, Panithanarak T, Searle JB. 2013. Understanding the basis of diminished gene flow between hybridizing chromosome races of the house mouse. *Evolution* 67(5):1446–1462.

Gregorová S, Gergelits V, Chvatalova I, Bhattacharyya T, Valiskova B, Fotopulosova V, Jansa P, Wiatrowska D, Forejt J. 2018. Modulation of Prdm9-controlled meiotic chromosome asynapsis overrides hybrid sterility in mice. *eLife* 7:e34282.

Grey C, Barthès P, Chauveau-Le Friec G, Langa F, Baudat F, de Massy B. 2011. Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol.* 9(10):e1001176.

Gündüz İ, Auffray J-C, Britton-Davidian J, Catalan J, Ganem G, Ramalhinho MG, Mathias ML, Searle JB. 2001. Molecular studies on the colonization of the Madeiran archipelago by house mice. *Mol Ecol.* 10(8):2023–2029.

Gündüz I, López-Fuster MJ, Ventura J, Searle JB. 2001. Clinal analysis of a chromosomal hybrid zone in the house mouse. *Genet Res.* 77(1):41–51.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132(2):583–589.

Jeffreys AJ, Cotton VE, Neumann R, Lam KG. 2013. Recombination regulator PRDM9 influences the instability of its own coding sequence in humans. *Proc Natl Acad Sci U S A.* 110(2):600–605.

Kono H, Tamura M, Osada N, Suzuki H, Abe K, Moriwaki K, Ohta K, Shiroishi T. 2014. Prdm9 polymorphism unveils mouse evolutionary tracks. *DNA Res.* 21(3):315–326.

Masly JP, Jones CD, Noor MAF, Locke J, Orr HA. 2006. Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science* 313(5792):1448–1450.

Medarde N, López-Fuster MJ, Muñoz-Muñoz F, Ventura J. 2012. Spatiotemporal variation in the structure of a chromosomal polymorphism zone in the house mouse. *Heredity (Edinb).* 109(2):78–89.

Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323(5912):373–375.

Morgan AP, Fu C-P, Kao C-Y, Welsh CE, Didion JP, Yadgary L, Hyacinth L, Ferris MT, Bell TA, Miller DR. 2015. The mouse universal genotyping array: from substrains to subspecies. *G3* 6:263–279.

Nunes AC, Britton-Davidian J, Catalan J, Ramalhinho MG, Capela R, Mathias ML, Ganem G. 2005. Influence of physical environmental characteristics and anthropogenic factors on the position and structure of a contact zone between two chromosomal races of the house mouse on the island of Madeira (North Atlantic, Portugal). *J Biogeogr.* 32(12):2123–2134.

Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP. 2009. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS Genet.* 5(12):e1000753.

Orr HA, Turelli M. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky–Muller incompatibilities. *Evolution* 55(6):1085–1094.

Paigen K, Petkov PM. 2018. PRDM9 and its role in genetic recombination. *Trends Genet.* 34(4):291–300.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290.

Pavlova SV, Searle JB. 2018. Chromosomes and speciation in mammals. In: Zachos FE, Asher RJ, editors. Mammalian evolution, diversity and systematics. Berlin (Germany): De Gruyter. p. 17–38.

Phadnis N, Orr HA. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323(5912):376–379.

Piálek J, Hauffe HC, Searle JB. 2005. Chromosomal variation in the house mouse. *Biol J Linn Soc.* 84(3):535–563.

Presgraves DC, Balagopalan L, Abmayr SM, Orr HA. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* 423(6941):715–719.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.

Qiao H, Chen JK, Reynolds A, Höög C, Paddy M, Hunter N. 2012. Interplay between synaptonemal complex, homologous recombination, and centromeres during mammalian meiosis. *PLoS Genet.* 8(6):e1002790.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press.

Sánchez-Guillén RA, Capilla L, Reig-Viader R, Martínez-Plana M, Pardo-Camacho C, Andrés-Nieto M, Ventura J, Ruiz-Herrera A. 2015. On the origin of Robertsonian fusions in nature: evidence of telomere shortening in wild house mice. *J Evol Biol.* 28(1):241–249.

Sans-Fuentes MA, García-Valero J, Ventura J, López-Fuster MJ. 2010. Spermatogenesis in house mouse in a Robertsonian polymorphism zone. *Reproduction* 140(4):569–581.

Schliep KP. 2011. Phangorn: phylogenetic analysis in R. *Bioinformatics* 27(4):592–593.

Schwartz JJ, Roach DJ, Thomas JH, Shendure J. 2014. Primate evolution of the recombination regulator PRDM9. *Nat Commun.* 5:4370.

Smagulova F, Brick K, Pu Y, Camerini-Otero RD, Petukhova GV. 2016. The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes Dev.* 30(3):266–280.

Tiemann-Boege I, Schwarz T, Striedner Y, Heissl A. 2017. The consequences of sequence erosion in the evolution of recombination hotspots. *Philos Trans R Soc B* 372(1736):20160462.

Ting CT, Tsaur SC, Wu ML, Wu CI. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282(5393):1501–1504.

Supplementary Information

PRDM9 diversity at fine geographical scale reveals contrasting evolutionary patterns and functional constraints in natural populations of house mice

Supplementary Table 1: Localities sampled, sample size (N), diploid number (2n) and Rb fusion frequencies of mice from the Barcelona Rb system. See Figure 1 for details on the geographical distribution of localities.

Localities	N	2n	Frequency Rb fusions						
			(3.8)	**(4.14)**	**(5.15)**	**(6.10)**	**(7.17)**	**(9.11)**	**(12.13)**
Castellfollit del Boix	19	40	-	-	-	-	-	-	-
Sta. Perpètua de Mogoda	22	40	-	-	-	-	-	-	-
Olost	17	40	-	-	-	-	-	-	-
Caldes de Montbui	10	40	-	-	-	-	-	-	-
Badalona	11	39-40	-	-	-	-	-	-	0.21
Les Pobles	16	39-40	-	-	0.08	-	-	-	-
L'Ametlla de Segarra	7	39-40	-	0.29	-	-	-	-	-
El Papiol	7	37-39	-	0.14	0.14	-	-	0.64	0.14
Calafell	14	37-39	0.09	0.29	-	0.03	-	0.15	0.56
St. Sadurní d'Anoia	15	35-39	-	0.58	-	-	-	0.31	0.39
Cubelles	11	32-39	0.04	0.46	0.17	0.58	-	0.71	0.63
La Granada	3	32-37	0.46	0.87	0.81	0.44	-	0.81	0.81
El Prat de Llobregat	11	31-35	-	0.73	0.50	0.86	-	0.91	0.77
Viladecans	2	31-32	1	1	0.50	0.50	-	1	0.50
Castelldefels	23	28-33	0.50	0.98	0.59	0.88	0.03	0.81	0.98

Supplementary Table 2: Localities sampled on the island of Madeira, including locality names, sample size (N), diploid number (2n) and chromosomal race. See Figure 1 for details on the geographical distribution of localities and chromosomal races. Madeiran chromosomal race abbreviations where first described by Britton-Davidian et al. (2000) and follows the nomenclature originally defined by Piálek et al. (2005), where the first one or two letters indicate country in which metacentric population was described or is distributed, followed by the first letters of localities harboring the metacentric population. Abbreviations: PSAN – PortugalSANtana, PADC – PortugalAchadasDaCruz, PEDC – PortugalEstreitoDaCalheta, PLDB – PortugalLugarDeBaixo, PPOD – PortugalPOntaDelgada.

Island	Localities	N	2n	Chromosomal race
Porto Santo Island	Camacha	3	40	Standard
	Farrobo	1	40	Standard
	Ponta	3	40	Standard
	Quinta das Palmeiras	4	40	Standard
Madeira Island	Porto da Cruz	3	22	PSAN
	Santana	4	22	PSAN
	Achadas da Cruz	8	24-27	PADC
	Fajã da Ovelha	3	25-26	PADC
	Levada Grande	2	27	PADC
	Lombada dos Cedros	1	26	PADC
	Lombada dos Marinheiros	5	25-26	PADC
	Ponta do Pargo	22	24-25	PADC
	Porto Moniz	3	28	PADC
	Ribeira da Vaca	4	24	PADC
	Santa	4	24-26	PPADC
	Ribeira da Janela	7	26-27	PADC and PEDC
	Lombo da Velha	3	24-26	Contact zone PEDC/PADC
	Lugar da Raposeira	2	25-26	Contact zone PEDC/PADC
	Maloeira	2	24-26	Contact zone PEDC/PADC
	Prazeres	2	24-25	Contact zone PEDC/PADC
	Solar da Maloeira	2	25-26	Contact zone PEDC/PADC
	Solar dos Prazeres	1	26	Contact zone PEDC/PADC
	Madalena do Mar	5	24	PEDC
	Chão da Ribeira	20	23-25	PEDC
	Ribeira da Laje	3	25-26	PEDC
	Ribeira Funda	4	25-26	PEDC
	Seixal	1	24	PEDC
	Arco da Calheta	17	23-24	PEDC
	Estreito da Calheta	9	24-26	PEDC
	Lombo das Laranjeiras	3	24-25	PEDC
	Lombo do Doutor	5	24-25	PEDC
	Moledos	2	24	PEDC
	Sítio da Fajã	3	24	Contact zone PEDC/PLDB
	Ponta do Sol	4	24	PLDB
	Socorro	9	24	PLDB and PEDC/PLDB
	Ponta Delgada	10	27-28	PPOD
	São Vicente	4	26-27	PSVI

Supplementary Table 3: Karyotypes of chromosomal races in Madeira. For each chromosomal race, the chromosomes involved in Rb fusions are indicated.

Chromosomal race	Rb fusions
PSAN	Rb (2.19) (3.8) (4.16) (5.14) (6.7) (9.10) (11.12) (13.17) (15.18)
PADC	Rb (2.4) (3.14) (5.18) (7.15) (8.18) (9.12) (10.16) (13.17)
PEDC	Rb (2.4) (3.14) (6.7) (8.11) (9.12) (10.16) (13.17)
PLDB	Rb (2.4) (3.14) (6.7) (8.11) (9.12) (10.16) (13.17) (15.18)
PPOD	Rb (3.8) (4.5) (9.18) (10.16) (11.12) (14.17)
PSVI	Rb (2.4) (3.8) (5.18) (6.7) (9.12) (10.16) (13.17)

Supplementary Table 4: Primers used both to generate *Prdm9* amplicons and for Sanger sequencing as described by Kono et al. (2014).

Primer	Sequence
Prdm9 ZFA-L_F	GAAAGTAAGAGAACTGTGGAAGAGCTCAGAA
Prdm9 ZFA-L_R	GAGATGTGGTTTTATTGCTGTTGGCTTTCTC

Supplementary Table 5: Amino acid (Aa) variation in the ZnF domain found in the Barcelona Rb system and the Madeira archipelago. Aa sequence alignments are presented for the first, internal and the last ZnF repeat of the array. Each ZnF repeat is assigned to a numeric ID based on the Aa variation of the highly variable positions -1, +3 and +6 (highlighted in green). Additional Aa variation found outside the highly variable positions are highlighted in yellow. Asterisk: new ZnF repeats identified in the present study.

	ID	Aa variation													-1			+3			+6									
First repeat		DNE	S	S	I	E	R	Q	C	G	Q	Y	F	S	D	K	S	N	V	N	E	H	Q	K	T	H	T	G	E	K
	3	Q(N)HQ	P	Y	V	C	R	E	C	G	R	G	F	T	Q	N	S	H	L	I	Q	H	Q	R	T	H	T	G	E	K
		QHQ	P	Y	V	C	R	E	C	G	R	G	F	T	Q	K	S	H	L	I	Q	H	Q	R	T	H	T	G	E	K
	4	QDK	P	Y	V	C	R	E	C	G	R	G	F	T	Q	K	S	D	L	I	K	H	Q	R	T	H	T	G	E	K
		(FW)QD(F)K(P	P	Y	V	C	R	E	F	G	W	G	F	T	Q	K	S	D	F	I	K	T	Q	R	T	H	P	R	E	K	
	6	QVK	P	Y	V	C	R	E	C	G	R	G	F	T	Q	K	S	V	L	I	K	H	Q	R	T	H	T	G	E	K
	8	AVQ	P	Y	V	C	R	E	C	G	R	G	F	T	A	K	S	V	L	I	Q	H	Q	R	T	H	T	G	E	K
	9	ANQ	P	Y	V	C	R	E	C	G	R	G	F	T	A	K	S	N	L	I	Q	H	Q	R	T	H	T	G	E	K
		(I)ANQ	P	Y	V	C	R	E	C	G	R	G	F	I	A	K	S	N	L	I	Q	H	Q	R	T	H	T	G	E	K
	12	QNK	P	Y	V	C	R	E	C	G	R	G	F	T	Q	K	S	N	L	I	K	H	Q	R	T	H	T	G	E	K
	14	VVK	P	Y	V	C	R	E	C	G	R	G	F	T	V	K	S	V	L	I	K	H	Q	R	T	H	T	G	E	K
	13	QDQ	P	Y	V	C	R	E	C	G	R	G	F	T	Q	K	S	D	L	I	Q	H	Q	R	T	H	T	G	E	K
		(W)QDQ(P)	P	Y	V	C	R	E	C	G	W	G	F	T	Q	K	S	D	L	I	Q	H	Q	R	T	P	T	G	E	K	
Internal repeat	16	VNQ	P	Y	V	C	R	E	C	G	R	G	F	T	V	K	S	N	L	I	Q	H	Q	R	T	H	T	G	E	K
	17	ASK	P	Y	V	C	R	E	C	G	R	G	F	T	A	K	S	S	L	I	K	H	Q	R	T	H	T	G	E	K
	18	VVQ	P	Y	V	C	R	E	C	G	R	G	F	T	V	K	S	V	L	I	Q	H	Q	R	T	H	T	G	E	K
	19	ESK	P	Y	V	C	R	E	C	G	R	G	F	T	E	K	S	S	L	I	K	H	Q	R	T	H	T	G	E	K
	22*	EDQ	P	Y	V	C	R	E	C	G	R	G	F	T	E	K	S	D	L	I	Q	H	Q	R	T	H	T	G	E	K
		EDQ(N)	P	Y	V	C	R	E	C	G	R	G	F	T	E	K	S	D	L	I	Q	N	Q	R	T	H	T	G	E	K
	23	AS(V)Q	P	Y	V	C	R	E	C	G	R	G	F	T	A	K	S	S	L	V	Q	H	Q	R	T	H	T	G	E	K
	25	QVQ	P	Y	V	C	R	E	C	G	R	G	F	T	Q	K	S	V	L	I	Q	H	Q	R	T	H	T	G	E	K
	24	QHK	P	Y	V	C	R	E	C	G	R	G	F	T	Q	K	S	H	L	I	K	H	Q	R	T	H	T	G	E	K
		Q(N)HK	P	Y	V	C	R	E	C	G	R	G	F	T	Q	N	S	H	L	I	K	H	Q	R	T	H	T	G	E	K
	26*	Q(N)LQ	P	Y	V	C	R	E	C	G	R	G	F	T	Q	N	S	L	L	I	Q	H	Q	R	T	H	T	G	E	K
	27*	ADQ	P	Y	V	C	R	E	C	G	R	G	F	T	A	K	S	D	L	I	Q	H	Q	R	T	H	T	G	E	K
	28	AD(F)K	P	Y	V	C	R	E	C	G	R	G	F	T	A	K	S	D	F	I	K	H	Q	R	T	H	T	G	E	K
	29*	VDK	P	Y	V	C	R	E	C	G	R	G	F	T	V	K	S	D	L	I	K	H	Q	R	T	H	T	G	E	K
	30	TDK	P	Y	V	C	R	E	C	G	R	G	F	T	T	K	S	D	L	I	K	H	Q	R	T	H	T	G	E	K
	31	EDK	P	Y	V	C	R	E	C	G	R	G	F	T	E	K	S	D	L	I	K	H	Q	R	T	H	T	G	E	K
	20	QNQ	P	Y	V	C	R	E	C	G	R	G	F	T	Q	K	S	N	L	I	Q	H	Q	R	T	H	T	G	E	K
	32	ENQ	P	Y	V	C	R	E	C	G	R	G	F	T	E	K	S	N	L	I	Q	H	Q	R	T	H	T	G	E	K
	34	ANK	P	Y	V	C	R	E	C	G	R	G	F	T	A	K	S	N	L	I	K	H	Q	R	T	H	T	G	E	K
Last repeat	13	(W)QDQ (R)	P	Y	V	C	R	E	C	G	W	G	F	T	Q	K	S	D	L	I	Q	H	Q	R	T	H	T	R	E	K
		(W)QDQ	P	Y	V	C	R	E	C	G	W	G	F	T	Q	K	S	D	L	I	Q	H	Q	R	T	P	T	G	E	K
	4	(W)QDK	P	Y	V	C	R	E	C	G	W	G	F	T	Q	K	S	D	L	I	K	H	Q	R	T	H	T	G	E	K

Supplementary Table 6: Hardy-Weinberg analysis in the Madeira archipelago and the Barcelona Rb system. Ho = observed heterozygosity. He = expected heterozygosity. N.A.: not analyzed (too few PRDM9 alleles for Hardy-Weinberg analysis, see main text for further details).

Area of study	Rb fusions	Population/Rb race	Ho/He	P-value
Barcelona Rb system	Standard	Castelfollit del Boix	0.7462	0.1328
		Sta.Perpètua Mogoda	0.5512	0.0060
		Olost	0.3877	0.0011
		Caldes de Montbui	NA	NA
	Rb	Badalona	1.1769	0.9087
		Les Pobles	NA	NA
		L'Ametlla de Segarra	NA	NA
		El Papiol	0.6875	0.5158
		Calafell	1.0952	0.1972
		Sant Sadurni d'Anoia	NA	NA
		Cubelles	0.4375	0.1082
		La Granada	NA	NA
		Prat de Llobregat	0.6875	0.5153
		Viladecans	NA	NA
		Casteldefells	NA	NA
Madeira archipelago	Standard	Porto Santo	0.7353	0.0883
	Rb	PSAN	0.2000	0.0017
		PADC	0.3485	<0.000001
		contact zone PEDC/PADC	0.4018	<0.000001
		PEDC_SOUTH	0.3538	<0.000001
		PEDC_NORTH	0.2964	<0.000001
		PPOD	0.2913	0.000167
		PLDB	0.2595	0.000016
		PSVI	NA	NA
		contact zone PEDC/PLDB	NA	NA

Supplementary Table 7: Degree of conservation in amino acid (Aa) sequence of the ZnF repeats located in positions 3 to 6 (ZnF3-6) along the ZnF array in the different PRDM9 allelic heterozygous (HT) combinations found in the study. For each PRDM9 HT combination population frequencies are shown for the Barcelona Rb system and the Madeiran archipelago, distinguishing between standard (all acrocentric chromosomes) and Rb mice (metacentric chromosomes).

PRDM9 HT combination	Conservation in Aa sequence (%)	Population frequency (%)			
		Barcelona Rb system		Madeira archipelago	
		Standard	Rb	Standard	Rb
8A/12F	88.9	4.76	0	0	0
8A/12G	22.2	4.76	0	0	0
10A/11B	77.78	9.52	0	0	6.67
10A/11C	100	19.05	0	0	0
10A/11K	88.9	0	0	0	13.33
10A/11P	88.8	0	0	0	6.67
10A/11S	33.3	4.76	0	0	2.22
10A/12B	100	14.29	54.50	0	0
10A/12C	77.8	14.29	0	0	2.22
10A/12D	88.8	0	9.10	0	0
10A/12E	100	0	27.30	0	0
10A/12G	100	4.76	0	0	0
10A/12L	77.8	0	0	0	2.22
10A/13C	77.8	0	0	0	13.33
10A/13G	44.4	0	0	0	2.22
10B/11B	33.3	23.81	0	0	0
10E/13C	77.8	0	0	0	2.22
10G/11T	33.3	0	0	0	2.22
11B/12C	100	0	0	20.00	0
11B/12H	55.6	0	0	0	2.22
11B/12J	88.9	0	0	0	2.22
11C/12D	88.9	0	9.10	0	0
11K/12K	100	0	0	0	2.22
11K/13C	66.7	0	0	0	15.56
11K/13H	44.4	0	0	0	2.22
11K/15F	88.9	0	0	0	2.22
11S/12C	22.2	0	0	40.00	0
11S/13C	44.4	0	0	0	4.44
11U/15B	33.3	0	0	0	2.22
11V/12C	33.3	0	0	20.00	0
11W/12C	22.2	0	0	20.00	0
12K/13C	66.7	0	0	0	2.22
13B/15A	88.8	0	0	0	2.22
13B/15E	100	0	0	0	2.22
13C/15D	88.9	0	0	0	2.22
13C/16A	66.7	0	0	0	2.22
13J/15C	100	0	0	0	2.22

Supplementary Figure 1: Representative metaphase plates from Rb mice in Madeira. (A) 2n=22, (B) 2n=28, (C) 2n=24. Scale bar = 10 μm.
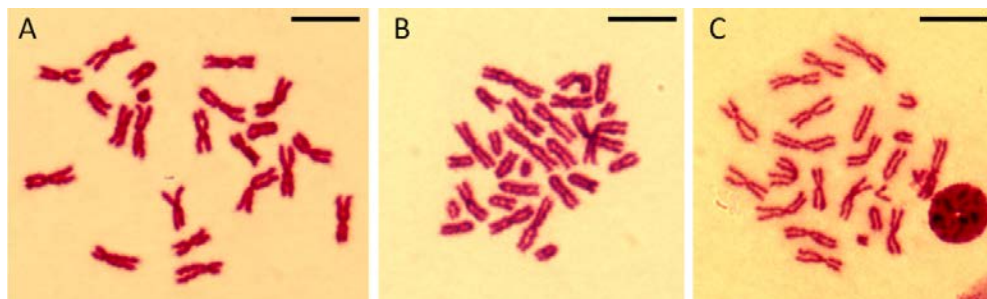
Supplementary Figure 2: Representative agarose gel (1%) showing amplification of heterozygous samples with different allelic lengths. (A) 11-13 ZnF repeats. (B) 11-12 ZnF repeats. (C) 10-11 ZnF repeats. DNA marker: GeneRuler DNA ladder Mix (range 100–10,000 bp).
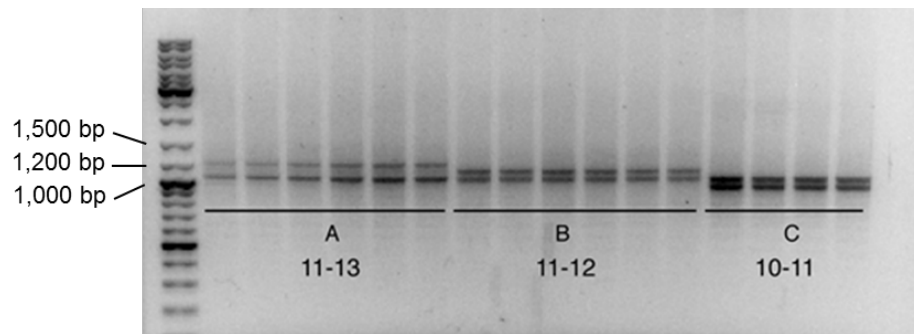
Supplementary Figure 3: ΔK-values for different K (from K=2 to K=10). K-values with the lowest likelihood are encircled (K =3, K = 4 and K = 5).

Supplementary Figure 4: Heat map plot representing the conservation in amino acid sequence of the ZnF repeats located in positions 3 to 6 (ZnF3-6) along the ZnF array in the different PRDM9 allelic heterozygous combinations found in the Madeiran archipelago. Standard and the different chromosomal races are represented. Colour intensity represents the percentage of amino acid sequence conservation (see Supplementary Table 7 for details).

GENERAL DISCUSSION

Recent advances in the study of chromatin 3D folding have provided new insights into our understanding of genome structure and function (Dekker et al. 2013; Fudenberg et al. 2016; Lieberman-Aiden et al. 2009; McCord et al. 2020; Naumova et al. 2013; Rao et al. 2014; Rowley and Corces 2018; Sexton et al. 2012b). It is now clear that the hierarchical organization of the genome comprises chromosomal territories in where chromatin is organized into compartments (open/closed), which consist of topologically associated domains and DNA loops, an organization that is widely conserved across mammalian species (Concia et al. 2020; George et al. 2020; Szabo et al. 2019). It is also generally accepted that this genome structure is highly dynamic during the cell cycle, subjected to changes in chromatin accessibility and histone modifications, ultimately modulating transcriptional activity (Delaneau et al. 2019; Despang et al. 2019; Grubert et al. 2015; Lupiáñez et al. 2015; Nurick et al. 2018; Symmons et al. 2014). However, when this work started, little was known about how genome organization is established and maintained during the formation of germ cells.

In this thesis we identified and characterized the structural and functional genomic changes that take place during mammalian spermatogenesis using the house mouse as model species. To attain this, we first developed a reproducible flow cytometry protocol (**section 4.1**) to obtain enriched mouse male germ cell populations required for the study of the dynamics of the 3D genome folding at fine scale (**section 4.2**). We then studied the germline of wild Rb mice populations to gather a broader understanding of how chromosomal rearrangements impact the 3D folding of the genome and meiotic recombination (**section 4.3**). Moreover, we conducted a comprehensive study of PRDM9 variability in natural populations of house mice to elucidate the influence of this essential protein in the recombination patterns observed in wild mice (**section 4.4**). Overall, we provide a comprehensive view of the plasticity of genome architecture during mammalian spermatogenesis and its implications for evolution and fertility.

5.1 Dynamics of the 3D structure of the genome during spermatogenesis and its functional implications

In order to study the plasticity of genome architecture during spermatogenesis, we first developed a flow cytometry method to efficiently isolate germ cell populations from mouse testis. Methodological approaches available when we initiated this project (i.e., sedimentation and sorting protocols, Bryant et al. 2013; Fallahi et al. 2010; Gassei et al. 2009; Gaysinskaya and Bortvin 2015; Getun et al. 2011) lacked reproducibility and presented some caveats regarding cell purity that prevented to conduct high-throughput studies at the fine scale of resolution. In order to circumvent these limitations, we developed a standardized FACS protocol that permitted the isolation of five germ cell fractions with high enrichment, including spermatogonia (91% average enrichment), primary spermatocytes at L/Z (88.7% average enrichment) and P/D (average 90% enrichment), RS (average 92% enrichment), and sperm (92.5% average enrichment). The application of this protocol in combination with the growing knowledge of cell-specific markers makes it a useful tool for the study of different germ cell types with a high level of isolation specificity. We foresee that this approach can be applied not only to spermatogenesis in rodents, but also to other mammalian taxa and even vertebrates.

5.1.1 Spermatogenesis undergoes structural changes correlated with transcription and cohesin occupancy

The application of the Hi-C methodology on FACS-isolated pre-meiotic, meiotic and post-meiotic cells, in combination with the study of meiotic cohesins and the CTCF insulator occupancy together with RNA-seq has allowed us to generate a structural and functional atlas of the 3D organization of the mouse genome

during spermatogenesis. This has permitted the study of different stages of spermatogenesis from a structural and functional point of view.

5.1.1.1 Spermatogonia: structural features for the somatic-to-meiotic transition

Spermatogonia represent an heterogeneous population of diploid cells, including self-renewing stem cells that maintain the spermatogonial pool and differentiated meiotic-committed cells, which will undergo spermatogenesis (de Rooij 2017; Phillips et al. 2010; Wistuba et al. 2007). Contact probability maps generated in the present study suggested a somatic-like chromatin organization for spermatogonia with interaction profiles similar to fibroblasts (our somatic control). This included the presence of clear A/B compartments and TADs.

As the inter-/intra-chromosomal interaction ratio can be considered a proxy for the presence of CTs (Cremer and Cremer 2010; Fritz et al. 2019), the pattern found in spermatogonia suggests that territories are still detectable in this cell type. However, CTs seem to be transitioning to the disposition subsequently found in meiotic stages, given that the overall ratio values decreased for all chromosomes when compared to fibroblasts but were still higher than in meiotic cells. Such changes in chromosomal disposition could be related to the histone marks remodeling (i.e., H3K9me2) that characterizes the transition from self-renewing to meiotic-committed cells (Shirakawa et al. 2013; Tseng et al. 2015) and the chromosomal movements necessary to initiate pairing of homologous chromosomes in pre-meiotic cells (Handel and Schimenti 2010).

In addition, we detected that active transcription correlated to open chromatin states in spermatogonia (A compartments), mirroring previous observations in somatic cells (Lieberman-Aiden et al. 2009; Nurick et al. 2018; Rao et al. 2014; van Steensel and Furlong 2019). Interestingly, transcripts found in spermatogonia were mostly involved in cell communication, chemotaxis and regulation of molecular functions, consistent with the spermatogonial process of cell differentiation. For instance, we found high transcription of *Cmtm2a*, which recent studies have shown that has an essential role for mouse fertility (Fujihara et al. 2018). Interestingly, *Cmtm2a* is involved in the synthesis process of testosterone and chemotaxis, evidencing a pivotal role in spermatogonial regulation and in initiating meiosis. Overall, our results suggest that commitment to enter meiosis is accompanied by changes in chromosome occupancy and transcription in spermatogonia.

5.1.1.2 Meiotic chromosome organization is highly dynamic during prophase I and accommodates transcription

It is during prophase I when major chromatin remodeling takes place, including chromatin condensation, chromosomal movements (i.e., telomere tethering to the nuclear envelope, pairing and synapsis) and the formation and repair of DSBs (Capilla et al. 2016; Handel and Schimenti 2010). The structural pattern obtained in our Hi-C analyses suggest that DNA folding of meiotic chromosomes (early and late stages of prophase I) adopt a state related, albeit different, to what has been described for chromosomes in metaphase (Gibcus et al. 2018; Naumova et al. 2013). We detected that, in primary spermatocytes, genomic regions located at short distances (less than 10 Mb) interacted more frequently than those positions at longer distances (further than 10 Mb). This observation was consistent with a mitotic-like condensation status according to previous descriptions of mitotic chromosomes (Gibcus et al. 2018; Naumova et al. 2013).

The similarity between the P(s) curves described in this thesis when compared to the pattern found in yeast and rhesus monkey germ cells (Alavattam et al. 2019; Patel et al. 2019; Schalbetter et al. 2019; Wang et al. 2019b), evidences that meiotic chromosomes are organized in mitotic-like compressed arrays of chromatin loops along the synaptonemal complex (Gibcus et al. 2018; Handel and Schimenti 2010; Muller et al. 2018; Naumova et al. 2013; Schalbetter et al. 2019). Moreover, Hi-C maps obtained for primary spermatocytes suggested the lengthening of DNA loops emerging out of the chromosomal axes as prophase I progresses (Muller et al. 2018; Patel et al. 2019; Zickler and Kleckner 1999). In fact, P/D presented more interactions at longer distances when compared to L/Z Hi-C maps. These changes in chromatin folding observed during prophase I were further evidenced by the absence of compartment domains in primary spermatocytes, which

lacked the strong layout found in both spermatogonia and fibroblasts (Patel et al. 2019; Schalbetter et al. 2019; Spindler et al. 2019).

Importantly, we also observed that meiotic chromosome can accommodate transcriptional activity. In particular, transcripts with functions related to meiotic progression were detected in primary spermatocytes. But not only that, transcripts related with spermiogenesis, fertilization, and embryo development (such as from *Cdh2*, involved in neural development), were also observed, mirroring previous observations (da Cruz et al. 2016; Ernst et al. 2019). Given the dynamics of chromatin remodeling detected in prophase I, our results suggest that meiotic chromosomes balance chromatin condensation with active transcription. This is a pattern that contrasts with what has been observed in mitotic chromosomes, which are considered transcriptionally silent (Gottesfeld and Forbes 1997; Palozola et al. 2017).

Moreover, active transcription correlated with A compartments in primary spermatocytes (despite compartments showed a weak signal), an observation further confirmed by recent data in the germline of the rhesus monkey (Wang et al. 2019b). Our detection of weak compartmentalization also agreed with what was recently shown by Alavattam and collaborators (Alavattam et al. 2019), who computed compartments in mouse spermatocytes using a similar statistical approach. This contrasts, however, with the observations made by Patel and collaborators, who showed that transcription-correlated compartments are conserved during meiotic prophase (Patel et al. 2019). These contrasting interpretations put into relevance the complexity of the Hi-C data and the statistical methodological approaches used for compartment calling. Whereas Patel and colleagues argued the need for compartments to maintain transcription (Patel et al. 2019), transcription does not seem to be required for higher-level genome compartmentalization (van Steensel and Furlong 2019; Wang et al. 2019b). Moreover, the extended view of the correlation between stable enhancer-promoter contacts and active transcription has been recently challenged (Alexander et al. 2019; Leidescher et al. 2020), as transient contacts could also lead to transcription. In this context, our results, along with others (Alavattam et al. 2019; Luo et al. 2020; Wang et al. 2019b) support the loss of both A/B compartmentalization and TADs during prophase I. Absence of A/B compartmentalization will serve to accommodate major events that takes place during prophase I, including chromatin condensation, chromosomal movements and the formation and repair of DSBs.

The active transcription observed in primary spermatocytes could also be the result of transient chromatin contacts. As TADs are mainly associated to long-range gene regulation (Dekker and Heard 2015; Jin et al. 2013; Thurman et al. 2012; Zhan et al. 2017), the lack of TADs observed in primary spermatocytes suggest the presence of local transcription regulation during meiotic prophase I. Altogether, transcription would have a moderate effect on the higher-order domain organization of the genome, accommodating the idea that during prophase I, despite compartmental domains lose their strength, active transcription takes place.

5.1.1.3 Cohesin occupancy points to a dual structural and functional role during prophase I

It is well-known the role of CTCF and cohesins in establishing chromatin topological organization in somatic cells, as they create boundary domains by the loop extrusion process (Fudenberg et al. 2016; Krietenstein et al. 2020; Rao et al. 2017). However, their role in establishing and maintaining the 3D genome folding during spermatogenesis is less explored.

During meiosis, canonical mitotic cohesins (i.e., RAD21) are replaced by meiotic-specific cohesins (i.e., RAD21L and REC8), which have a pivotal role in preserving sister chromatid cohesion, homolog synapsis, and the formation and repair of DSBs (Agostinho et al. 2016; Herrán et al. 2011; Lee and Hirano 2011; Llano et al. 2012; Rong et al. 2016). Previous studies have shown that REC8 and RAD21L present an alternate distribution along the SC that has been linked with their loading timing during prophase I, with RAD21L appearing earlier than REC8 (Ishiguro et al. 2011, 2014; Lee and Hirano 2011; Lee et al. 2003). This pattern suggests a progressive and stepped loading of REC8 and RAD21L to the SC during early stages of meiosis. Our ChIP-seq analysis and high-resolution microscopy observations support the view of an alternate disposition of cohesins along chromosomal axes while displaying a substantial overlap (50%) between RAD21L and REC8 peaks.

Our results also suggested a previously unreported role of cohesins in transcription during meiosis. On the one hand, meiotic cohesins were found to be distributed every 220-265 kb along the chromosomal axes of autosomes. On the other hand, DNA loops anchored at the SC are estimated to be 1.5-2 Mb in size in primary spermatocytes (Patel et al. 2019; Ruiz-Herrera et al. 2017; Wang et al. 2019a). Therefore, we interpret the excess of cohesins peaks observed in relation to DNA loop size as cohesins associating to chromatin loops emerging out of the chromosomal axis. Related to this, cohesins were preferentially located at promoters and TSS in primary spermatocytes, being their preferential binding site an RNA polymerase II regulatory region (motif CCGCHGCC). This pattern has been previously reported for somatic cells (Dorsett and Merkenschlager 2013; Maya-Miles et al. 2019; Panigrahi et al. 2012). Considering our results, we can speculate that cohesins are associated to the transcriptional machinery in the DNA loops that protrude out of the SC in meiotic chromosomes. This can be related to local interactions (2.5-4.5 Mb, as observed in P/D Hi-C maps) that can be highly dynamic (Leidescher et al. 2020; van Steensel and Furlong 2019). Altogether, our results suggest a role of meiotic cohesin occupancy in transcription that needs further functional validation.

While cohesins were preferentially associated with TSS, CTCF binding sites were mainly located in intergenic regions and showed strong insulator properties in primary spermatocytes, as previously reported for somatic cells (Kim et al. 2017; Nora et al. 2017; Phillips-Cremins and Corces 2013; Sofueva et al. 2013). These results indicate that CTCF conserves its original insulator role in prophase I, thus delimiting topological domains.

In the light of our results, we can hypothesize that a subset of meiotic cohesins associated to intergenic regions would predominantly develop their structural role in chromatin folding during SC assembly in association to CTCF. Complementary, another subset of REC8 and RAD21L located at the vicinity of TSS would be involved in recruiting transcriptional factories in the DNA loops that protrude out of meiotic chromosomal axes. This suggests a multilayered role of cohesins in chromatin dynamics, accommodating the structural and functional needs of meiotic chromosomes to progress throughout prophase I.

5.1.1.4 Cohesin-modulated chromatin folding in round spermatids

Once meiosis is completed, spermatids undergo extensive chromatin remodeling, along with the morphological changes necessary for the transition from RS to elongating spermatids that will finally result in the formation of sperm (Balhorn 2011). One of the most relevant changes that takes place at the structural level is the replacement of histones by protamines (Balhorn 2007, 2011; Bao and Bedford 2016; Hao et al. 2019; Hud et al. 1993; Miller et al. 2010; Rathke et al. 2014). Accordingly, we found that genes related to the histone-to-protamine transition were highly expressed in early spermiogenesis (RS), mirroring previous studies (Blanco and Cocquet 2019; Chen et al. 2018; da Cruz et al. 2016; Trovero et al. 2020).

In this context, our results revealed a reprogramming of the genome compartmentalization in post-meiotic cells. On the one hand, A/B compartments were re-established in RS after meiosis, reflected by the recovery of a somatic-like plaid-pattern with an increase in the inter-/intra-chromosomal interactions ratio when compared to pre-meiotic and meiotic cells. On the other hand, TAD borders in RS did not present a strong insulator pattern but less defined borders, suggesting a different pattern of genomic folding at the sub-Mb scale. Thus, TADs are not established in RS in the same way as in other cell types, being more transient structures. This could be interpreted as an organizational preamble for chromatin remodeling during spermiogenesis, as flexibility and accessibility are needed for an efficient histone-to-protamine transition [(Blanco and Cocquet 2019) and references therein].

Given the re-establishment of structural domains in RS, the genomic distribution of cohesins and CTCF in post-meiotic was intriguing. The distribution of REC8 in RS resembled the pattern detected in primary spermatocytes, suggesting a preservation of REC8 in post-meiotic cells. By contrast, the presence of RAD21L peaks increased in RS when compared to primary spermatocytes. As both REC8 and RAD21L preserved their preferred locations in the vicinity of promoters, we speculate that RAD21L could play a more predominant role than REC8 in the transcriptional up-regulation that takes place in RS (da Cruz et al. 2016; Hammoud et al. 2014). Interestingly, CTCF was drastically reduced in RS (97.1% decrease) when compared to primary spermatocytes, probably related to the chromatin remodeling detailed above. The CTCF reduction in RS was surprising, as previous studies have shown more CTCF binding sites than we detected in these cells

associated to enhancers and promoters (Hernández-Hernández et al. 2016; Jung et al. 2017; Pugacheva et al. 2015; Rivero-Hinojosa et al. 2017). As the CTCF peaks detected in our study were mostly located in intergenic regions, we cannot rule out the possibility that our experimental set up detected a portion of the CTCF binding sites present in RS, the ones involved in the structural folding of the DNA (Wang et al. 2016b).

5.1.1.5 Chromatin condensation spatially constraints genome architecture in sperm

As for the final product of spermatogenesis, the sperm, our results revealed the presence of yet another level of chromatin folding, different from the patterns described in pre-meiotic cells, primary spermatocytes and RS. Mouse spermatozoa are characterized by a hook-shaped cell nucleus that contains highly condensed chromatin. Moreover, protamines adopt a three-faced structural role in the sperm nucleus: (i) coiling DNA into toroids, (ii) increasing tension within the DNA molecule and (iii) reducing DNA loops size (Balhorn 2011; Battulin et al. 2015; Brewer 2011; Cree et al. 2011; De Vries et al. 2012; Zalensky et al. 1995).

In this context, the analysis of the inter-/intra-chromosomal interactions ratio suggested that although highly compacted, chromosomes in sperm are likely arranged into CTs, consistent with early FISH studies (Brinkley et al. 1986; Haaf and Ward 1995; Zalensky et al. 1995). We also detected an increase in pericentromeric interactions when compared to meiotic cells, attributable to centromere clustering into the chromocenter (Haaf and Ward 1995; Hoyer-Fender et al. 2000; Meyer-Ficca et al. 1998; van Steensel and Furlong 2019; Zalensky et al. 1995). At the Mb scale, we detected the presence of clear A/B compartments. Furthermore, although sperm presented the lowest variance in TAD insulation scores among all cell types we studied, TAD borders presented stronger insulator scores when compared to RS.

The TAD score variance observed can be attributable to the tight toroid organization of the sperm chromatin (Brewer 2011; Cree et al. 2011) that would constrain the dynamics of TAD formation previously present in RS, thus yielding more stable TADs. Moreover, sperm TADs are larger in size than in fibroblasts, matching previous observations (Battulin et al. 2015). Altogether, our results support the view that sperm chromatin adopts a unique folding configuration organized in somatic-like chromosome territories, as shown in other studies (Jung et al. 2017; Ke et al. 2017), but modulated at the sub-Mb scale by the toroidal organization.

5.1.2 Implications of germline chromatin remodeling for embryonic development

Importantly, we detected the presence of gene transcripts relevant for fertilization and embryogenesis in post-meiotic cells (i.e., RS and sperm), mirroring previous studies (da Cruz et al. 2016; Hammoud et al. 2014; Sin et al. 2015). That was the case of *Nanog*, *Notch1*, *Kif3c*, *Amer3* and *Ccd88a*, whose expression in RS was correlated with meiotic cohesin occupancy. For instance, *Notch1* transcripts were found in RS [involved in oligodendrocyte differentiation and cardiac development, (Wang et al. 2019c)] with cohesins located near its TSS. As such, *Kif3c* [with a relevant function in the development of the nervous system (Navone et al. 2001)], Amer3 [involved in the regulation of cell growth and the *Wnt* pathway (Comai et al. 2010)] and *Ccd88a* (implicated in nervous system development), presented the same pattern (i.e., gene transcripts with cohesins associated to TSS).

In the light of these results, the activation of transcription during meiosis and spermiogenesis might play a role in the development of the future embryo, providing transcripts needed upon zygote genome activation (Flyamer et al. 2017; Ke et al. 2017; Vallot and Tachibana 2020). In fact, previous studies have described the preservation of an atypical promoter usage in postmeiotic cells (established during early spermatogenesis), resulting in a bivalent/poised chromatin state (Hammoud et al. 2014). Chromatin in poised state has both active and repressive histone modifications and can prevent DNA methylation, thus maintaining germ cell identity for developmental totipotency (Hammoud et al. 2014; Lesch et al. 2013; Lesch and Page 2014; Sin et al. 2015). In fact, the recovery of the somatic methylation program after fertilization has been linked to the presence of long-range interactions in sperm (Battulin et al. 2015; Jung et al. 2017; Ke et al. 2017), with a prominent role in gene expression regulation (Wang et al. 2016a). That was the case, for example, of *Nanog*, whose transcripts were detected in sperm, and it has been related to embryonic development by the modulation of long-range interactions (Denholtz et al. 2013). Growing evidence suggests that the

paternal chromatin of the zygote would directly inherit the sperm chromatin compartmentalization, setting the higher-order conformation basis for the zygote [(Vallot and Tachibana 2020) and references therein].

5.1.3 Modulation of chromatin structure and transcriptional silencing of the X chromosome

During prophase I, the X chromosome suffers transcriptional silencing (MSCI) (Margolin et al. 2014; Soumillon et al. 2013). This meiotic X inactivation is part of the meiotic checkpoint that detects the presence of partially or completely unsynapsed regions, which if triggered induces transcriptional repression called meiotic silencing of unsynapsed chromatin (MSUC) – an epigenetic silencing program that is conserved across eukaryotes (Ernst et al. 2019; Subramanian and Hochwagen 2014; Turner 2007, 2015). Transcriptional silencing is attained by the accumulation of histone marks such as H3K9me3/2, H2A ubiquitylation, HP1β (Namekawa et al. 2006; Turner et al. 2006) and absence of active RNA polymerase II, H3K27m1/3, H3K9ac and H4K16ac (Namekawa et al. 2006). As such, the X chromosome, along with the Y, forms a distinct (i.e., interacting less frequently with autosomes) repressive nuclear compartment (i.e., the sex body) during prophase I detectable by Hi-C maps. At the chromosomal level, MSCI was accompanied by the loss of A/B compartmentalization in the X chromosomes, along with fewer and larger TADs when compared with pre-meiotic cells. This lack of compartmentalization is consistent with the repressive environment that characterizes MSCI. The same pattern was observed in post-meiotic cells, consistent with the presence of PMSC (Namekawa et al. 2006).

As for insulator factors, the X chromosome was mostly depleted for CTCF in both primary spermatocytes and RS, concomitant with the lack of compartmentalization observed in this chromosome. Meiotic cohesins were also significantly reduced in the X chromosome of both P/D and RS when compared to autosomes. This decrease of cohesins can be related to their role in maintaining homolog pairing, as it is restricted to the PAR region in sex chromosomes (Rong et al. 2017). Importantly, the few remaining REC8 and RAD21L peaks detected in the X chromosome were mostly found in the vicinity of genes that are known to escape silencing during meiosis (da Cruz et al. 2016; Turner 2015), further evidencing the possible role of meiotic cohesins in transcription. Cohesin patterns along the X chromosome revealed by super-resolution microscopy were, however, striking. Whereas REC8 presented a lighter signal on the X chromosome when compared to autosomes, the RAD21L signal was stronger in the X when compared to autosomes. These results suggests that RAD21L might play a relevant role in transcription (as described above for autosomes), possibly by the recruitment of transcriptional hubs (e.g., by dimerization) (Maya-Miles et al. 2019). This would also explain the increased signal while their binding sites are significantly reduced on the X chromosome. Altogether, chromatin remodeling in the X chromosome was concomitant with the sex body transcriptional silencing during meiosis in both meiotic and post-meiotic cells.

5.2 Structural and functional impact of Rb fusions

Given the importance of the higher-order chromatin structure in stablishing gene-regulatory domains, disturbances in 3D folding can involve changes in gene regulation over evolutionary time. This is because the shifting of domain boundaries can expose coding regions to novel regulatory environments. In fact, previous studies have shown how genome reorganizations can result in changes in promoter-enhancer contacts (Hnisz et al. 2016b; Lupiáñez et al. 2015), inducing ectopic gene expression and forming asymmetrical chromatin interactions (Kraft et al. 2019).

To further understand the impact of chromosomal rearrangements on the overall genome architecture of germ cells, we analyzed the physical chromatin interactions in a unique natural house mice population, the BRbS, which is characterized by the presence of chromosomal fusions in polymorphic state (Adolph and Klein 1981; Gündüz et al. 2001b; Sánchez-Guillén et al. 2015; Sans-Fuentes et al. 2010). We detected disturbances in chromatin remodeling at different hierarchical levels in both somatic and meiotic cells of Rb mice. These changes ranged from an alteration of heterologous chromosome interactions to the reshaping of local chromatin configurations.

5.2.1 Rb fusions reshape heterologous interactions genome-wide

Initial microscopy studies suggested that Rb fusions can alter the disposition of chromosomes within the nucleus of mouse spermatocytes (Garagna et al. 2001). More recent descriptions extended these observations, as Rb fusions can introduce nuclear constraints, creating new spatial configurations and chromosome associations (Berríos 2017; Berríos et al. 2014, 2010). Our work represents a departure from those previously conducted studies as it relies on the characterization of the effect of Rb fusions on 3D folding genome-wide by studying chromatin interactions during different stages of spermatogenesis.

Our data revealed that the presence of Rb fusions alter the inter/intra-chromosomal interactions ratio genome-wide in germ cells (primary spermatocytes and RS) of mice of the BRbS. Particularly, we detected an increase of heterologous interactions in primary spermatocytes, which could be interpreted as a direct result of the reorganization of chromosome territories within the nucleus. As the presence of Rb fusions can disrupt the standard nuclear architecture of peripherally-attached telocentrics to center-located metacentrics (Berríos 2017; Berríos et al. 2014), this spatial chromosome reorganization can promote new interactions between domains that, eventually, can facilitate the formation of additional rearrangements (Ashley et al. 2006; Berríos et al. 2010; Branco and Pombo 2006; Matveevsky et al. 2020). In fact, it has been proposed that convergence of heterochromatic pericentric domains could promote the formation of Rb metacentrics (Berríos et al. 2017). In agreement with this view, heterologous associations between acrocentric chromosomes were more frequently found in the presence of Rb metacentrics than in all-acrocentric spermatocytes. As such, we detected that Rb fusions disrupt the chromosome spatial conformation, inducing the formation of new chromosomal associations that can trigger the occurrence of new chromosomal fusions.

Our results also suggest that chromosomal distribution within the nucleus is dependent on the synapsis state of Rb fusions during prophase I. In fact, we observed three different synapsis states in the primary spermatocytes of Rb mice: fully synapsed, open, and asynapsed. Based on the synapsis state of Rb metacentrics, the permissiveness in the occurrence of heterologous associations varied. Partially synapsed metacentrics prompted the disruption of inter-chromosome associations, including sex chromosomes. In fact, the formation of the sex body was also affected by the presence of Rb fusions, as fully asynapsed metacentrics were associated with the heterochromatinization of sex chromosomes. Given that H3K9me3 deposition is associated to 'closed' chromatin states (Margolin et al. 2014; Soumillon et al. 2013), the excess of H3K9me3 signal detected in the sex body of Rb mice spermatocytes could also be the result of additional transcriptional repression. However, the functional consequences of fully heterochromatinization of sex chromosomes are not known at this stage. Our results on sperm viability (along with previous observations of abnormal sperm configurations in BRbS, Medarde et al. 2013) suggest that the abnormal silencing due to asynapsis could be affecting the transcription of essential genes for proper spermiogenesis in primary spermatocytes. That can be the case, for instance, of X-linked genes (i.e., *Zfy1/2*) that normally escape transcriptional silencing and are involved in sperm morphogenesis (head morphology and sperm tail formation) (Royo et al. 2010; Vernet et al. 2012). According to this hypothesis, Rb fusions can have an impact on the modularity pattern of the sperm head in the BRbS (Medarde et al. 2013). Thus, excessive heterochromatinization (related to transcriptional repression) can be associated to the high levels of spermatocyte apoptosis and subfertility that characterize the Rb populations of the BRbS (Medarde et al. 2015; Sans-Fuentes et al. 2010).

In the case of post-meiotic cells (RS), however, the opposite pattern in chromosome interactions was observed. That is, the overall heterologous interactions were lower in Rb mice when compared to standard mice, suggesting the presence of a more spatially constrained genome conformation in post-meiotic cells to the presence of chromosomal fusions. It has to be taken into account that in RS, chromosomes are organized so that centromeres are found in the center of the cell (the so-called chromocenter), while telomeres are attached to the nuclear lamina (Brinkley et al. 1986; Hoyer-Fender et al. 2000). In this nuclear context, the presence of Rb fusions posits high mechanical constrains in chromosome conformation around centromeric regions, reducing chromosome interactions genome-wide.

5.2.2 Novel regulatory environments triggered by Rb fusions

Although the presence of Rb fusions did not involve substantial changes at the sub-Mb scale in primary spermatocytes (i.e., A/B compartments and TADs were detected), it affected the disposition of chromosomes within the nucleus. In fact, Rb fusions triggered the formation of new inter-chromosomal interactions during prophase I, with implications for the remodeling of the transcriptional landscape. Newly formed inter-chromosomal interactions detected in Rb mice were enriched for olfactory/vomeronasal receptor (OR/VR) genes. It is known that OR genes are expressed during meiosis, as they play an important role in sperm maturation and motility, being involved in sperm-egg interactions (Tatsura et al. 2001; Vanderhaeghen et al. 1997). Moreover, OR expression has been related to the formation of multi-chromosomal enhancer hubs in somatic cells (Bashkirova and Lomvardas 2019). In fact, OR-related interactions are regulated by LHX2 (LIM Homeobox protein 2) (Monahan et al. 2019), which is expressed during meiosis and associated with meiotic cohesins. Although we did not conduct functional analysis in Rb mice in this thesis (i.e., RNA-seq), we hypothesize that newly established transcriptional hubs can affect OR transcription by disrupting enhancer-promoter contacts of OR genes that are essential for spermiogenesis and fertilization. If this holds true, disruptions of OR transcription could be involved in the subfertility observed in Rb mice (this work; Sans-Fuentes et al. 2010; Wallace et al. 2002).

Altogether, we hypothesize that the remodeling of the nuclear architecture detected in Rb mice would expose chromosomal domains to novel regulatory environments, potentially affecting gene expression and/or regulation, as initially proposed by the integrative breakage model of genome architecture (Deakin et al. 2019; Farré et al. 2015). This can have implications in both fertility and evolution. As such, the presence of new chromosomal interactions may rewire or attenuate gene networks, providing new grounds for evolutionary novelty in the long round.

5.3 Chromosomal fusions as modifiers of meiotic recombination

As Rb fusions affect 3D folding, we sought to understand the impact in recombination. Meiotic chromosomes are organized into DNA loops anchored to the SC, whose lengths are correlated to the number of COs (Capilla et al. 2016; Qiao et al. 2012; Ruiz-Herrera et al. 2017; Wang et al. 2019a, 2015b; Zickler and Kleckner 2015). Therefore, the organization of meiotic chromosomes can be an important modulator of meiotic recombination. In fact, previous studies in Rb mice, not only in the BRbS but also in other Rb systems reported a reduction in the number of COs (Capilla et al. 2014; Dumas et al. 2015; Merico et al. 2003; Merico et al. 2013) and chiasmata (Castiglia and Capanna 2002; Dumas et al. 2015; Dumas and Britton-Davidian 2002) due the presence of chromosomal fusions.

Here we studied the impact of Rb fusions in recombination rates using an integrative approach that combines estimates of linkage disequilibrium based on SNP genotyping and cytological analysis that maps CO events (exemplified by MLH1 signals) directly on male germ cells. In particular, we aimed to elucidate the genome-wide effect of Rb fusion on recombination landscape, paying special attention to chromosomal configurations. Mirroring previous observations (Capilla et al. 2014), we detected low recombination rates in fused chromosomes when compared to their non-fused counterparts using both direct (immunolocalization of meiotic COs) and indirect (linkage disequilibrium and genomic divergence) measures of recombination. Remarkably, recombination reduction was stronger in homozygous than in heterozygous, concomitant with the presence of quadrivalents with misaligned centromeres. This was consistent with high values of genomic divergence (expressed as F_{ST}) as a result of the redistribution of COs along chromosomal arms.

5.3.1 Rb fusions introduce pairing mechanical disturbances and centromere misalignment

Importantly, the presence of Rb fusions reduced not only the overall number of COs, but also the number of RAD51 foci detected per cell. Since RAD51 foci can be considered a proxy for DSBs (Jasin and Rothstein

2013; Moens et al. 2002), our data suggested that in Rb mice, although the number of DSBs were lower when compared to standard mice a higher proportion of them was resolved as COs, eventually ensuring the obligatory CO per chromosome. Regarding the factors that could contribute to the overall reduction of DSBs in Rb mice, we can only speculate at this stage. However, in the light of our results in the 3D genome folding analysis (**section 4.3**), the altered chromosome spatial distributions observed in Rb mice could also affect DSB formation, as chromatin compartmentalization and accessibility can be affected. Further experimental analyses (i.e., DMC1 ChIP-seq) will help us to verify this hypothesis.

The stronger recombination suppression around the centromeric regions observed in homozygous metacentric chromosomes when compared to heterozygous was striking, as according to the 'suppressed recombination' model (Faria and Navarro 2010; Farré et al. 2013; Rieseberg 2001), a reduction in recombination is expected within reorganized regions in heterokaryotypes. Thus, we analyzed which mechanical factors might be responsible for the pattern observed in homozygous metacentric chromosomes. It is known that the number of COs is correlated to SC length (Kleckner et al. 2003; Ruiz-Herrera et al. 2017; Wang et al. 2019a, 2015b; Zickler and Kleckner 2015). In this context, variations in COs were associated with changes in both the SC length and DNA loop size in Rb mice (as inferred from cytological and Hi-C maps). We detected that chromosomal axes of heterozygous fused chromosomes were significantly longer in size than both homozygous and acrocentric arms. Considering that the presence of Rb can alter the nuclear disposition of chromosomes in prophase I (**section 4.3**), the synapsis of trivalents (e.g., heterozygous chromosomes) involves chromosomal movements from centric regions of the nucleus towards the nuclear periphery (Berríos et al. 2014; Garagna et al. 2014). This can lead to a delay in the completion of synapsis, resulting in the detection of elongated chromosomal axes. As chromosomal axes elongate, there is more substrate for the formation of COs. Conversely, as quadrivalents (e.g., homozygous chromosomes) did not show synapse disturbances, chromosomal axes length was not altered, being shorter in size when compared to heterozygous. But as homozygous and non-fused chromosomes did not differ in SC length in primary spermatocytes, additional factors are probably responsible for the reduction of COs observed in homozygous metacentric chromosomes.

In fact, the reduction in recombination detected in homozygous chromosomes was associated to the presence of quadrivalents with double centromere signals and with a distribution of COs towards distal regions of chromosomal arms. Previous cytogenetic studies have reported the existence of bivalents of fused chromosomes with double centromeric signal in shrews (Borodin et al. 2008), hamsters (Bikchurina et al. 2018), horses (Cappelletti et al. 2019) and the mole vole (Matveevsky et al. 2020), although their prevalence in other mammalian species is currently unknown. Here we provide evidence for the presence of double centromere signals in quadrivalents in prophase I in mice with Rb fusions in homozygous state, which could be the result of misaligned centromeres, thus affecting the formation of COs at proximal and middle regions of chromosomal arms. Since the centromere position has an influence in the reduction of the recombination rates (the so-called centromere interference, Lynn et al. 2004), the presence of double centromeres could magnify this effect, interfering with both the formation of COs and in chromatid segregation in metaphase I. Previous studies on BRbS have detected heterochromatinization disturbances at the centromeres of fused chromosomes (Capilla et al. 2014), and a higher frequency of apoptotic spermatogenetic cells in homozygous when compared to heterozygous mice (Medarde et al. 2015). In the light of our results, we hypothesize that the reduction in recombination can be triggered by different mechanistic disturbances depending on the Rb fusion state: (i) centromeric misalignment in quadrivalents, and (ii) "open" asynapsed pericentromeric regions in trivalents.

5.3.2 Recombination remodeling and genetic diversity

In the case of natural Rb populations, where the distribution of metacentrics is widespread, the presence of heterokaryotypes with a high number of Rb fusions is not expected to be common. In fact, the prevalence of Rb fusions in heterozygous state was moderate (from one to three trivalents) in the sampled BRbS populations included in the present study. This is probably because the higher the selection against the heterokaryotype (underdominance) is, the more extreme are the demographic conditions required for its fixation [(Dobigny et al. 2017) and references therein]. On the contrary, the lower the degree of underdominance associated with a particular Rb fusion, the higher the probability to be present as a polymorphic form. Thus, the presence

of Rb fusions in the BRbS can be related to a mild underdominance, given that previous studies in this system did not detect strong fertility impairment (Medarde et al. 2015; Sans-Fuentes et al. 2010). Despite a redistribution and reduction of COs in Rb mice, the frequency of chromosomal arms with absence of COs was not significantly altered in mice with Rb fusions in heterozygosis. This observation has important implications since it suggests that meiosis is not meaningfully compromised in Rb mice from the BRbS (negating any underdominant effect), given that the proper disjunction of chromosomes would not be compromised as at least one CO is present per bivalent (Hassold et al. 2000; Segura et al. 2013). This is consistent with the absence of a significant alteration in reproduction reported in other house mice Rb systems with one to three fusions (Castiglia and Capanna 2000; Wallace et al. 1992, 2002). Under this scenario, a decrease in meiotic recombination due to the presence of a relatively low number of Rb fusions present as chromosomal polymorphism does not necessarily affect fertility significantly but might lead to an increase in genetic divergence within affected regions.

In fact, this was the pattern observed when analyzing the genomic landscape of divergence and recombination using SNP genotyping. We detected higher genomic divergence in fused chromosomes than in non-fused chromosomes, consistent with the observation of low COs frequencies (MLH1 foci per cell) and low values of recombination rates (expressed as 4Ner/Kbp). Together with heterozygous unfitness (Giménez et al. 2013), the effect of Rb fusions could lead to the accumulation of genetic incompatibilities and possibly to genetic isolation between populations, as has been described in other systems (Faria and Navarro 2010; Noor et al. 2001; Rieseberg 2001). Our results suggest that the redistribution of experimentally detected COs in the BRbS might result in a tendency for higher levels of genetic differentiation due to CO suppression in the presence of Rb fusions.

5.4 High PRDM9 variability in wild populations of house mice

As recombination hotspots in mammals are mainly determined by PRDM9, we also investigated the allelic variability of PRDM9 in wild populations of mice and its link with genetic diversity. To that aim we surveyed two different Rb systems: the Madeira Rb system, an insular system, and the BRbS, a continental system. The contrasting nature of the chromosomal variation in the Madeira and Barcelona Rb systems provides a unique opportunity to investigate the mechanisms underlying PRDM9 variation in natural populations.

Our analysis unveiled a high PRDM9 variability considering both the number and sequence of the ZnF array, which was unprecedented to previous surveys (Buard et al. 2014; Kono et al. 2014). As such, we distinguished 25 different ZnF domains. We found alleles containing from 8 ZnF to 16 ZnF repeats, conforming a total of 57 alleles between the Madeiran and BRbS populations, from which 49 alleles were described for the first time. Moreover, the analysis at fine geographical scale revealed contrasting evolutionary patterns: Madeiran Rb mice presented greater allelic variability than their non-fused counterparts, while BRbS Rb mice showed less variability. Such differences could be attributed to the combination of different factors: (i) the evolutionary history of each Rb system, (ii) the prevalence of Rb fusions and (iii) to a lesser extent, meiotic functional constrains (i.e., recombination hotspot asymmetry).

In the case of the Madeira Rb system, the phylogeographic analysis presented in this thesis together with previous studies (Britton-Davidian et al. 2000; Gündüz et al. 2001a; Mathias and Mira 1992; Medarde et al. 2012) revealed that despite presenting common Eurasian origins, mice populations distributed across Madeira result from a complex colonization process when compared to the BRbS. Madeira, an island with an extreme topography, has six well-established chromosomal races that have been described within a geographical range of only 742 km^2 (Britton-Davidian et al. 2000). Diploid numbers vary from 2n=22 to 2n=28, with up to nine metacentric chromosomes accumulated within a maximum of 1,200 years (Förster et al. 2009). Importantly, all metacentric populations are geographically isolated and do not co-occur with others except in very rare situations, where some of the chromosomal races marginally overlap (e.g., PADC and PEDC). The chromosomal differences between Madeira races are so pronounced that hybrids are almost non-existent mainly due to meiotic impairment of the F1 (Britton-Davidian et al. 2000). It has been suggested that Rb mice populations would have arisen from multiple introgression events in Madeira (Britton-Davidian et al. 2007;

Förster et al. 2009, 2013; Gündüz et al. 2001b), mirroring previous reports in insular Rb mice systems, such as the Aeolian archipelago in Southern Italy (Franchini et al. 2020). Considering the geographical distribution of Rb mice in Madeira, we found that populations from the west coast (PEDC and PADC) and their contact zone (contact zone PEDC/PADC) presented more alleles (from 13 to 22 different alleles per population) than populations from other metacentric races (PSAN, PLDB, PPOD and PSVI, with 4 to 7 different alleles per population). This observation suggests that a hybridization process between metacentric races might be taking place, as the joint presence of *Prdm9* alleles from different races would prompt hybrid dysgenesis. Thus, new alleles would appear due to mutation processes fueled by the differential erosion of hotspots by different *Prdm9* alleles (Giménez et al. 2016; Hauffe et al. 2011; Hauffe et al. 2012; Tiemann-Boege et al. 2017).

The BRbS represents a different evolutionary scenario. The BRbS is characterized by the presence of seven different metacentrics, Rb(3.8), Rb(4.14), Rb(5.15), Rb(6.10), Rb(7.17), Rb(9.11) and Rb(12.13), distributed in non-geographically coincident (staggered) clines leading to a progressive reduction in diploid numbers towards the centre of the range, about 30km west of the city of Barcelona (Adolph and Klein 1981; Gündüz et al. 2001a; Medarde et al. 2012; Sánchez-Guillén et al. 2015; Sans-Fuentes et al. 2010). In this context, our analysis (i.e., PCA) are in agreement with a clinal evolution of these Rb populations, as we detected common ancestral populations for both Rb and non-fused populations that expanded radially from more interior populations towards coastal populations. This would explain the high prevalence of the 10A *Prdm9* allele, as it would have appeared in the initial population (likely being located in Olost or Castellfollit del Boix) and expanded accordingly in different populations. Thus, the presence of shared alleles in BRbS Rb mice suggests that they come from a recent expansion, probably arising from a bottleneck event.

Moreover, and given the pivotal role of PRDM9 in recombination hotspot determination, it was remarkable the overall deficit of heterozygous individuals detected for PRDM9, being specifically relevant in Madeira, where nearly half of the expected heterozygotes were observed. Interestingly, heterozygous individuals presented specific sequence conservation in those ZnF directly involved in DNA-binding (ZnF #3 to #6) (Baker et al. 2015; Billings et al. 2013; Paigen and Petkov 2018). This suggests that although presenting different PRDM9 combinations, heterozygous individuals tend to preserve hotspot symmetry. The preservation of hotspot symmetry is highly relevant to ensure that DSBs are formed and repaired correctly in the homolog context of the meiotic axis. Conversely, the asymmetrical disposition of hotspots represents a mechanical constraint that contributes to increased infertility (Brand and Presgraves 2016; Davies et al. 2016; Smagulova et al. 2011). Given these functional constrains, Rb mice have been probably selected for PRDM9 combinations that preserve symmetrical hotspots in the germ line.

Altogether, our data suggest that the PRDM9 allelic variability observed in Rb systems can be the result of the interplay between factors affecting population structure and the selective forces they are subjected to. These selective forces can be modulated upon the preservation of efficient recombination landscapes.

CONCLUSIONS

From the results obtained in this thesis, the following conclusions can be drawn:

1. Regarding the study of the dynamics of the 3D conformation of the genome during spermatogenesis in a laboratory house mouse strain:

   1.1 The development of a reproducible FACS protocol has allowed for the isolation of enriched germ cell populations from mouse testis. These included spermatogonia, primary spermatocytes at early (leptonema/zygonema) and later (pachynema/diplonema) stages of prophase I, round spermatids and sperm. This has permitted the study of the dynamics of the 3D conformation of the genome during spermatogenesis at a fine scale.

   1.2 The compartmentalization of the genome during spermatogenesis is highly dynamic, being reflected at different levels. Spermatogonia present a somatic-like 3D genome structure, with clear A/B compartments and TADs. Later on, A/B compartments and TADs are lost during prophase I, where chromatin is reorganized in meiotic chromosomes. The higher-order organization of the chromatin is then re-established in post-meiotic cells, with round spermatids and sperm both presenting A/B compartments and TADs. TADs are reorganized in a cell-specific way, being less stable in round spermatids and found in somatic-like chromosome territories in sperm.

   1.3 There is a fine-tuning between chromatin remodeling and cell-specific gene expression. All analyzed germline populations (pre-meiotic, meiotic and post-meiotic cells) present a correlation between A compartments and transcription, including cell-specific genes and genes essential for fertilization and embryonic development. This puts into relevance the importance of germline transcription regulation for fertility and embryonic development.

   1.4 Meiotic cohesins REC8 and RAD21L present an overlapping distribution along the synaptonemal complex in primary spermatocytes with an average periodicity of 243 kb. This genomic location suggests that cohesins are not only found within the context of the chromosomal axes, but also in the DNA loops that protrude out of the axes.

   1.5 Meiotic cohesins REC8 and RAD21L are found preferentially located in promoter regions of genes with active transcription in primary spermatocytes and round spermatids. This suggests that meiotic cohesins are involved in transcription regulation during meiosis, a role that is preserved in post-meiotic cells.

   1.6 The insulator protein CTCF shows a differential loading pattern in primary spermatocytes, with 20% of the detected peaks overlapping meiotic cohesins, preferentially localized in intergenic regions and associated with high insulator properties. This pattern contrasts with the strong reduction of CTCF peaks observed in round spermatids.

   1.7 The lack of A/B compartments and TADs in the X chromosome in both meiotic and post-meiotic cells is concomitant with the sex body transcriptional silencing during meiosis; MSCI in primary spermatocytes and PMSC in round spermatids. This chromatin remodeling is accompanied by a reduction of meiotic cohesins (REC8 and RAD21L) loading in the X chromosome in both meiotic and post-meiotic cells.

   1.8 Despite the overall X chromosome silencing, some transcripts are found associated to cohesins in their promoter regions. This suggests a role of cohesins in transcription during spermatogenesis.

2. Regarding the study of the effect of Rb fusions on the 3D conformation of the genome in the germ line of natural populations of house mice:

   2.1 The presence of Rb fusions affected the compartmentalization of the genome in germ cells in different degrees depending on the cell type. Primary spermatocytes present a high number of heterologous interactions, suggesting a genome-wide alteration of

chromosomal nuclear disposition. On the contrary, round spermatids show a reduction of inter-chromosomal contacts, indicating that Rb fusions impose physical restrictions in the cell nucleus of post-meiotic cells.

2.2 The remodeling of the nuclear architecture detected in Rb mice exposes chromosomal domains to novel regulatory environments, as exemplified by olfactory receptors genes.

2.3 Robertsonian mice from the BRbS presented a significant reduction of COs when compared to standard mice. This reduction was significant in homozygous metacentrics (quadrivalents), a pattern associated with an increased proportion of chromosomal arms with zero CO and the presence of double centromeric signals. This suggests that centromeric misalignment pose mechanistic disturbances in the formation of COs, displacing them towards distal regions of chromosomal arms.

2.4 Robertsonian fusions in heterozygosis (trivalents), presented different degrees of asynapsis at the centromeric regions, associated with longer chromosomal axes. Given the alteration of chromosomal nuclear disposition observed in prophase I in Rb mice, the detection of elongated chromosomal axes in trivalents can be the result of a delay in the completion of synapsis. As chromosomal axes elongate, chromatin remodeling permits more substrate for the formation of COs.

2.5 The presence of asynaptic trivalents in primary spermatocytes prompts heterologous associations between the sex body and autosomes, resulting in an excessive sex body heterochromatinization (exemplified by H3K9me3). This altered heterochromatinization of the sex body can contribute to transcriptional silencing of essential reproductive genes, thus impairing the individual's fertility.

2.6 The presence of Rb fusions also resulted into a reduction of recombination rates estimated by the analysis of linkage disequilibrium based on SNP genotyping at the population level. This was consistent with high values of genomic divergence (expressed as F_{ST}) as a result of the redistribution of COs along chromosomal arms.

3. Regarding the characterization PRDM9 variability in natural populations of house mice:

3.1 Natural populations of house mouse Rb systems present a high PRDM9 variability (57 different alleles, from which 49 alleles were described for the first time) with contrasting patterns between insular (Madeira) and continental (BRbS) systems. Madeiran Rb mice presented a higher allelic variability (54 different alleles alleles) than mice from the BRbS (13 different alleles).

3.2 Differences in PRDM9 variability between both Rb systems result from their distinct phylogeography patterns. On the one hand, the insularity and holography of Madeira combined with multiple introgression events facilitated the presence of a high number of alleles. On the other hand, the presence of shared alleles in BRbS Rb mice result from a recent expansion, arising from a bottleneck event.

3.3 Both Rb systems present a deficit of PRDM9 heterozygous where the allelic combinations most frequently encountered maintained amino acid sequence similarity from position ZnF3 to ZnF6 in the ZnF array, ZnF involved in DNA-binding. This pattern can be the result of functional constraints that facilitate the accumulation of allelic combinations that maintain recombination hotspot symmetry to ensure efficient meiotic recombination.

APPENDIX

A.1 Supplementary Tables

A.1.1 Source table for antibodies

Antibodies	Source	Dilution	Protocol	Identifier
Anti-mouse SYCP3	Abcam	1:400 | 1:1000	IF | FACS	Cat#ab97672
Anti-Rabbit DMC1	Santa Cruz Biotechnologies	1:100 | 1:100	IF | FACS	Cat#sc-22768
Anti-mouse Cy5	Jackson Immunoresearch	1:200 | 1:1000	IF | FACS	Cat#115-175-146
Anti-rabbit FITC	Jackson Immunoresearch	1:200 | 1:1000	IF | FACS	Cat#111-095-003
Anti-mouse Cy3	Jackson Immunoresearch	1:200	IF	Cat#115-165-003
Anti-human Cy5	Jackson Immunoresearch	1:200	IF	Cat#709-175-149
Anti-mouse FITC	Jackson Immunoresearch	1:200	IF	Cat#115-095-003
Anti-rabbit Cy3	Jackson Immunoresearch	1:200	IF	Cat#111-165-003
Anti-rabbit CTCF	Millipore	1:1200 | 1:2500	IF | WB	Cat#07-729
Anti-rabbit REC8	Courtesy of A.M Pendás	1:100 | 1:2000	IF | WB	N/A
Anti-rabbit RAD21L	Courtesy of A.M Pendás	1:100 | 1:2000	IF | WB	N/A
Anti-HRP-PO	Bio-Rad	1:15000	WB	Cat#1706515
Anti-rabbit ß-tubulin	Abcam	1:5000	WB	Cat#ab108342
Anti-human CREST	Courtesy of M. Fritzler	1:100	IF	N/A
Anti-mouse MLH1	BD Pharmingen	1:50	IF	Cat#551092
Anti-rabbit γH2AX	Sigma-Aldrich	1:300	IF	Cat#H5912
Anti-rabbit H3K9Me3	Abcam	1:300	IF	Cat#ab8898
Anti-mouse α-tubulin	Courtesy of M. Martín	1:1000	IF	N/A
Anti-rabbit SYCP3	Abcam	1:300	IF	Cat#ab15093
Anti-Dig-FITC	Sigma-Aldrich	1:150	FISH	Cat#11207741910

A.1.2 Source table for reagents

Reagent	Source	Catalogue number
5x GC buffer	ThermoFisher	#F519L
Acrylamide/Bis-acrylamide	Sigma-Aldrich	#A3699
Ammonium Persulfate	Sigma-Aldrich	#A3678
AMPure XP Beads	Beckman-Coulter	#A63880
Anti-Cy3 dUTP	EnZo Life Sciencies	#ENZ-42501
Biotin-14-dATP	Life Technologies	#19524-016
Bovine Serum Albumin	New England Biolabs	#B9001S
Buffer, dNTPs and Taq	Takara	#316RR001A
ClarityTM Western ECL substrate	Bio-Rad	#170-5061
Collagenase Type II	Life Technologies	#17101015
Complete Protease inhibitor	Roche	#1187358001
Digoxigenin-11-dUTP	Sigma-Aldrich	Cat#11573152910
DMEM	Gibco	#31966047
DMSO	Sigma-Aldrich	#D2650
DNA Polymerase I	New England Biolabs	#M0210M
DNAse I	Sigma-Aldrich	#DN25-10MG
dNTPs	Roche	#11969064001
DPBS	Gibco	#14190-094
Dynabeads MyOne Streptavidin T1	Life Technologies	#65001
FBS	ThermoFischer	#10270106
Foetal Bovine Serum	ThermoFischer	#10270106
Formamide	Sigma-Aldrich	#47671-1L-F
Formaldehyde	Sigma-Aldrich	#F8775
GBSS	Sigma-Aldrich	#G9779-500ML
GelRed nucleic acid stain	Biotium	#BT-41002-0.5ml
GeneRuler	ThermoFischer	#SM0334
Gentamycin	Life Technologies	#15710049
Giemsa stain	Sigma-Aldrich	#109204
Glycerol	Sigma-Aldrich	#G5516-1L
Glycine	Sigma-Aldrich	#G7126-1KG
Hoechst 33342	Life Technologies	#H3570-10ml
Igepal CA630	Sigma-Aldrich	#I8896-100ML
IP buffer	Diagenode	#C01010173
KaryoMAX^{TM} Colcemid	Gibco	#15212012
Klenow Fragment 3'->5' exo-	New England Biolabs	#M0212M
L-Glutamine	Life Technologies	#25030024
Laemmli	Bio-Rad	#161-0737
Lypsol	Courtesy of H. Schertan	N/A
MboI	New England Biolabs	#R0147M
NEB Buffer 2	New England Biolabs	#B7002S
NEBNext dA-Tailing Module	New England Biolabs	#E6053
NEBNext End Repair Module	New England Biolabs	#E6050
NEBNext High Fidelity Mix	New England Biolabs	#M0541S
NEBNext Quick ligation buffer	New England Biolabs	#B6058S
NEBNext® Multiplex Oligos	New England Biolabs	#E7710S
Nick Translation kit	Abbot Molecular	#06J40-020
Nucleospin Gel and PCR Clean-up kit	Cultek	#22740609.250
Paraformaldehyde	Sigma-Aldrich	#F8775
Pencillin-Streptomycin	Life Technologies	#151140122
Pepsin	Sigma-Aldrich	#10108057001
Phenol/Chloroform/Isoamyl Alcohol	Sigma-Aldrich	#P2069-400ML
PhotoFlo	Kodak	#1464510
Phusion HF DNA Polymerase	ThermoFisher Scientific	#F530L
Pierce's BCA protein kit assay	ThermoFischer	#23227
Pipes	Sigma-Aldrich	#P1851
Ponceau S solution	Sigma-Aldrich	#P710-1L
Precision Plus Protein^{TM}	Bio-Rad	#1610374
Proteinase K (10 mg/ml)	New England Biolabs	#P8107S
Qiagen Midi Plasmid kit	Qiagen	#12143
RNAse A (10 mg/ml)	Life Technologies	#EN0531
Salmon Sperm	Invitrogen	#15632011

Continued on next page

Table S2 – *Continued from previous page*

Reagent	Source	Catalogue number
Seakem Agarose	Lonza	#H350004
Sodium Acetate 3M pH=5.2	Sigma-Aldrich	#S7899
SYBRTM Green I	ThermoFisher Scientific	#S7585
T4 DNA Ligase	New England Biolabs	#M0202M
T4 DNA Ligase Buffer 10X	New England Biolabs	#B0202S
T4 DNA polymerase	New England Biolabs	#M0203L
T4 Polynucleotides Kinase	New England Biolabs	#M0201L
TEMED	Sigma-Aldrich	#T9281
Trans-Blot® Turbo Mini PVDF	Bio-Rad	#1704156
Triton X-100	Sigma-Aldrich	#X100
Trypsin 0.05%	Gibco	#25300062
Trypsin from bovine pancreas	Sigma-Aldrich	#T9935-100MG
Tween-20	Sigma-Aldrich	#P1379
Unblocked Protein A beads	Diagenode	#kch-503-008
Vectashield	Vector laboratories	#H-100

A.1.3 Solutions table

Solution	Protocol	Recipe
4% stacking gel	Western Blot	0.6 M Tris pH=6.8 4% Acrylamide/Bis-acrylamide 10% APS 20% SDS TEMED
8% Running gel mix	Western Blot	8% Acrylamide/Bis-acrylamide 1.5 M Tris pH=8.8 20% SDS 10% APS TEMED
10x Running Buffer	Western Blot	30 g Tris base 144 g glycine 10 g SDS 1L MilliQ water
A-tailing solution	ChIP-seq	1x NEBNext A-tailing buffer NEBNext A-tailing enzyme mix
Adaptor ligation mix	In nuclei Hi-C	1x NEB Quick ligation reaction buffer NEBNext adapter NEB T4 DNA ligase
Antibody solution	3D FISH	1% BSA 4x SSC 0.2% Tween-20
Binding Buffer	In nuclei Hi-C	10 mM Tris-HCl pH=7.5 1mM EDTA 2M NaCl
Blocking solution I	Immunofluorescence	1x PBS 0.05% Tween-20
Blocking solution II	3D FISH	4% BSA 4x SSC 0.2% Tween-20
Blocking solution III	Western Blot	TBST 5% fat-free milk 1x PBS
Carnoy	Chromosome harvest	3:1 Methanol:glacial acetic acid
dATP attachment mix	In nuclei Hi-C	1x NEB Buffer 2 0.5 mM dATP 2.5U NEB Klenow exo minus
Elution Buffer	ChIP-seq	1% SDS 0.1 M NaHCO3
End repair mix	In nuclei Hi-C	1x NEB T4 DNA ligase buffer with 10 mM ATP 0.5 mM dNTP mix 50U NEB T4 PNK 12U T4 DNA polymerase I 5U NEB DNA polymerase I (klenow)
End repair solution	ChIP-seq	1x NEBNext End repair buffer NEBNext End repair enzyme mix

Table S3 – *Continued from previous page*

Solution	Protocol	Recipe
Fixative solution I	Spermatocyte Spreads Enrichment analysis	4% Paraformaldehyde (pH 9.8) 0.15% Triton X-100 MilliQ water
Fixative Solution II	3D FISH	4% Paraformaldehyde (pH 9.8) 0.3x PBS
Fixative solution III	3D FISH	1% Paraformaldehyde 1x PBS
Freezing medium	Cell Culture	1:10 DMSO in FBS
Hybridization Buffer	3D FISH	5.5 mL Formamide 1g Dextrane Sulfate 0.5 mL 20x SSC 1 mL MilliQ water
Hypotonic	Chromosome harvest	0.075 M KCl
Ligation Buffer	In nuclei Hi-C	1x NEB T4 Ligase Buffer 0.8% Triton x-100 1 mg/ml BSA 10000U cohesive ligase
Lysis Buffer	In nuclei Hi-C	10 mM Tris-HCl pH=8.0 10 mM NaCl 0.2% Igepal 1x Protease Inhibitors Cocktail
Lysis Buffer I	ChIP-seq	5mM Pipes 85 mM KCl 0.5% Igepal 1x Protease Inhibitors Cocktail
Lysis Buffer II	ChIP-seq	1% SDS 10 mM EDTA pH=8.0 50 mM Tris-HCl pH=8.1 1x Protease Inhibitors Cocktail
Lysis Buffer III	Genomic DNA extraction	100 mM Tris-HCl pH=8.5 200 mM NaCl 0.2% SDS 5 mM EDTA MilliQ water
Permeabilization solution	3D FISH	0.5% Triton X-100 1x PBS
Protein Digestion Solution	ChIP-seq	16 µl Tris pH=6.5 8 µl 0.5 M ETA pH=8.0 1.6 µl Proteinase K (10 mg/ml)
Reparation Mix	In nuclei Hi-C	1x NEB Buffer 2 0.05 mM dCTP 0.05 mM dTTP 0.05 mM dGTP 50 mM biotin dATP 50 U Kleenow

Continued on next page

Table S3 – *Continued from previous page*

Solution	Protocol	Recipe
RIPA buffer	Whole-protein extraction	10 mM TrisHCl pH=8.0 1mM EDTA 0.5 mM EGTA 1% Triton X-100 0.1% sodium deoxycholate 0.1% SDS 140 mM NaCl
SNET	Genomic DNA extraction	20 mM Tris-HCl pH=8.0 5 mM EDTA pH=8.0 400 mM NaCl 1% SDS (w/v)
Sorenson's Buffer	Chromosome harvest	133 mM Na2HPO4 133 mM KH2PO4
Supplemented DMEM	Cell culture	500 ml DMEM 50 µg/ml Gentamycin (10 mg/ml stock) 1x Penicillin-Streptomycin 20% FBS 2mM L-Glutamine
TE	Genomic DNA extraction	10 mM Tris-HCl pH=7.7-8 1 mM EDTA Up to 1 L MilliQ water
TBST	Western Blot	20 mM Tris 150 mM NaCl 0.1% Tween-20
Washing solution I	Spermatocyte Spreads Enrichment analysis	1 % PhotoFlo MilliQ water
Washing solution II	3D FISH	0.2% Tween-20 4x SSC

A.1.4 Sample tables

A.1.4.1 Mice from the Barcelona Robertsonian System

List of the BRbS samples used in the present work, including information about their origin, population, identification (ID), sex, 2N, and in which studies were included, together with its PRDM9 sequence if available. N/A: not analyzed.

Origin	Population	ID	SEX	2N	Prdm9 study	PRDM9 seq	Recombination study	SNP analysis	Hi-C
BRbS	Ametlla de Segarra	ATM1	N/A	39	yes	10A/10A			
BRbS	Ametlla de Segarra	ATM10	Male	39	yes	10A/10A			
BRbS	Ametlla de Segarra	ATM2	N/A	40	yes	N/A			
BRbS	Ametlla de Segarra	ATM3	N/A	39	yes	10A/10A			
BRbS	Ametlla de Segarra	ATM4	N/A	40	yes	N/A			
BRbS	Ametlla de Segarra	ATM7	N/A	40	yes	10A/10A			
BRbS	Ametlla de Segarra	ATM9	N/A	40	yes	10A/10A			
BRbS	Badalona	926	Male	39	yes	10A/10A			
BRbS	Badalona	927	Male	40	yes	10A/10A			
BRbS	Badalona	928	Male	40	yes	10A/11C			
BRbS	Badalona	929	Male	40	yes	10A/12B			
BRbS	Badalona	938	Male	39	yes	N/A			
BRbS	Badalona	940	Male	39	yes	10A/12D			
BRbS	Badalona	941	Male	40	yes	10A/11C			
BRbS	Badalona	942	Female	40	yes	10A/11C			
BRbS	Badalona	943	Female	40	yes	10A/10A			
BRbS	Badalona	944	Female	39	yes	11C/12D			
BRbS	Badalona	945	Female	40	yes	10A/11C			
BRbS	Calafell	10C	Male	38	yes	10A/10A			
BRbS	Calafell	12C	Female	38	yes	10A/10A			
BRbS	Calafell	13C	Female	37	yes	10A/12E			
BRbS	Calafell	14C	Female	39	yes	10A/10A			
BRbS	Calafell	18C	Male	39	yes	10A/12E			
BRbS	Calafell	1C	Male	37	yes	10A/10A			
BRbS	Calafell	20C	Female	37	yes	N/A			
BRbS	Calafell	2C	Female	37	yes	N/A			
BRbS	Calafell	3C	Male	38	yes	10A/10A			
BRbS	Calafell	4C	Male	39	yes	10A/10A			

Continued on next page

Table S4 – *Continued from previous page*

Origin	Population	ID	SEX	2N	Prdm9 study	PRDM9 seq	Recombination study	SNP analysis	Hi-C
BRbS	Calafell	5C	Female	38	yes	10A/10A			
BRbS	Calafell	6C	Male	37	yes	10A/12E			
BRbS	Calafell	7C	Female	38	yes	10A/10A			
BRbS	Calafell	CASTELL2	Female	38	yes	N/A			
BRbS	Caldes de Montbuí	20160316_01	Male	40	yes	13K/13K	yes		
BRbS	Caldes de Montbuí	20160316_02	Male	40	yes	N/A	yes		
BRbS	Caldes de Montbuí	20160316_03	Male	40	yes	N/A	yes		
BRbS	Caldes de Montbuí	20160316_04	Male	40	yes	10A/11S	yes		
BRbS	Caldes de Montbuí	20160316_05	Male	40	yes	N/A	yes		
BRbS	Caldes de Montbuí	20160316_06	Male	40		N/A	yes		
BRbS	Caldes de Montbuí	20160316_07	Male	40		N/A	yes		
BRbS	Caldes de Montbuí	20160316_08	Male	40	yes	12O/12O	yes		
BRbS	Caldes de Montbuí	20160316_10	Male	40		N/A	yes		
BRbS	Caldes de Montbuí	20160316_11	Female	40	yes	10A/10A			
BRbS	Castellfollit del Boix	1000	Male	40	yes	10A/12B	yes		
BRbS	Castellfollit del Boix	1019	Male	40	yes	10C/10C	yes	yes	
BRbS	Castellfollit del Boix	1020	Male	40	yes	10A/12B	yes	yes	
BRbS	Castellfollit del Boix	1021	Male	40	yes	12B/12B	yes	yes	
BRbS	Castellfollit del Boix	1024	Male	40	yes	10A/10A	yes	yes	
BRbS	Castellfollit del Boix	979	Male	40	yes	12B/12B	yes	yes	
BRbS	Castellfollit del Boix	980	Male	40	yes	10A/10A	yes	yes	
BRbS	Castellfollit del Boix	984	Female	40	yes	10A/10A	yes	yes	
BRbS	Castellfollit del Boix	998	Male	40	yes	10A/12C	yes	yes	
BRbS	Castellfollit del Boix	999	Male	40	yes	10A/12C	yes	yes	
BRbS	Castellfollit del Boix	982	Female	40	yes	10A/10A		yes	
BRbS	Castellfollit del Boix	1001	N/A	40	yes	10A/10A		yes	
BRbS	Castellfollit del Boix	1023	Male	40	yes	10A/12C			
BRbS	Castellfollit del Boix	983	Female	40	yes	N/A			
BRbS	Castellfollit del Boix	1002	Female	40	yes	N/A			
BRbS	Castellfollit del Boix	1015	Female	40	yes	N/A			
BRbS	Castellfollit del Boix	1018	Male	40	yes	N/A			
BRbS	Castellfollit del Boix	1025	Female	40	yes	N/A			
BRbS	Castelldefells	922	Male	29	yes	10A/10A	yes	yes	
BRbS	Castelldefells	923	Male	28	yes	10A/10A	yes	yes	
BRbS	Castelldefells	954	Male	30	yes	10A/10A	yes	yes	Fibroblasts Hi-C
BRbS	Castelldefells	955	Male	32	yes	10A/10A	yes	yes	
BRbS	Castelldefells	956	Male	30	yes	10A/10A	yes	yes	

Origin	Population	ID	SEX	2N	Prdm9 study	PRDM9 seq	Recombination study	SNP analysis	Hi-C
BRbS	Castelldefells	957	Male	33	yes	N/A			
BRbS	Castelldefells	958	Male	31	yes	10A/10A	yes	yes	
BRbS	Castelldefells	959	Male	31	yes	10A/10A	yes	yes	
BRbS	Castelldefells	960	Male	28	yes	10A/10A	yes	yes	
BRbS	Castelldefells	963	Male	29	yes	10A/10A	yes	yes	
BRbS	Castelldefells	965	Male	n.d.	yes	10A/10A			
BRbS	Castelldefells	967	Male	29	yes	10A/10A	yes	yes	
BRbS	Castelldefells	968	Male	30	yes	10A/10A	yes	yes	
BRbS	Castelldefells	970	Male	30	yes	10A/10A	yes	yes	
BRbS	Castelldefells	971	Male	30	yes	10A/10A	yes		
BRbS	Castelldefells	972	Male	31	yes	10A/10A	yes	yes	
BRbS	Castelldefells	CS10	Male	31	yes	10A/10A	yes	yes	
BRbS	Castelldefells	CS12	Female	31	yes	10A/10A			
BRbS	Castelldefells	CS13	Male	30	yes	10A/10A	yes	yes	
BRbS	Castelldefells	CS14	Male	30	yes	10A/10A	yes	yes	
BRbS	Castelldefells (Gavà)	20170721/1	Male	32	yes	10A/10A			
BRbS	Castelldefells (Gavà)	20170721/2	Female	N.D.	yes	10K/10K			
BRbS	Cubelles	CUB1	Female	39	yes	12B/12B			
BRbS	Cubelles	CUB10	Female	34	yes	10A/10A			
BRbS	Cubelles	CUB11	Male	32	yes	10A/10A			
BRbS	Cubelles	CUB12	Male	36	yes	10A/12B			
BRbS	Cubelles	CUB13	Male	33	yes	10A/10A			
BRbS	Cubelles	CUB14	Female	36	yes	10A/10A			
BRbS	Cubelles	CUB15	Female	34	yes	10A/10A			
BRbS	Cubelles	CUB16	Female	35	yes	10A/10A			
BRbS	Cubelles	CUB17	Female	36	yes	N/A			
BRbS	Cubelles	CUB19	Male	34	yes	10A/10A			
BRbS	Cubelles	CUB2	Female	35	yes	10A/12B			
BRbS	El Papiol	PP1	N/A	37	yes	10A/12B			
BRbS	El Papiol	PP2	N/A	38	yes	10A/10A			
BRbS	El Papiol	PP3	N/A	38	yes	12B/12B			
BRbS	El Papiol	PP4	N/A	38	yes	10A/10A			
BRbS	El Papiol	PP5	N/A	38	yes	10A/12B			
BRbS	El Papiol	PP6	N/A	39	yes	10A/10A			
BRbS	El Papiol	PP8	N/A	37	yes	N/A			
BRbS	La Granada	20170703/1	Male	34	yes	10I/10I			
BRbS	La Granada	20170703/2	Male	32	yes	10J/10J			

Origin	Population	ID	SEX	2N	Prdm9 study	PRDM9 seq	Recombination study	SNP analysis	Hi-C
BRbS	La Granada	20170703/3	Male	32	yes	N/A			
BRbS	Les Pobles	PS1	Female	40	yes	10A/10A			
BRbS	Les Pobles	PS10	Female	40	yes	10A/10A			
BRbS	Les Pobles	PS11	Male	39	yes	10A/10A			
BRbS	Les Pobles	PS12	Male	N.D.	yes	10A/10A			
BRbS	Les Pobles	PS13	Male	40	yes	10A/10A			
BRbS	Les Pobles	PS14	Male	40	yes	10A/10A			
BRbS	Les Pobles	PS15	Male	40	yes	N/A			
BRbS	Les Pobles	PS2	Female	39	yes	10A/10A			
BRbS	Les Pobles	PS27	N/A	40	yes	N/A			
BRbS	Les Pobles	PS3	Female	40	yes	10A/10A			
BRbS	Les Pobles	PS4	Female	40	yes	10A/10A			
BRbS	Les Pobles	PS5	Male	N.D.	yes	N/A			
BRbS	Les Pobles	PS6	Male	N.D.	yes	N/A			
BRbS	Les Pobles	PS7	Female	40	yes	N/A			
BRbS	Les Pobles	PS8	Male	40	yes	10A/10A			
BRbS	Les Pobles	PS9	Male	40	yes	10A/10A			
BRbS	Olost	10L	Male	40	yes	10C/10C	yes	yes	
BRbS	Olost	11L	Male	40	yes	12B/12B	yes	yes	
BRbS	Olost	14L	Male	40	yes	8A/8A	yes	yes	
BRbS	Olost	13L	Male	40	yes	N/A			
BRbS	Olost	15L	Female	40	yes	N/A			
BRbS	Olost	17L	Male	40	yes	N/A			
BRbS	Olost	18L	Male	40	yes	N/A			
BRbS	Olost	19L	Male	40	yes	8A/12G			
BRbS	Olost	21L	Male	40	yes	10A/10A	yes	yes	
BRbS	Olost	22L	Male	40	yes	8A/12F			
BRbS	Olost	23L	Male	40	yes	N/A			
BRbS	Olost	4L	Male	40	yes	N/A			
BRbS	Olost	5L	Male	40	yes	10A/12G	yes	yes	
BRbS	Olost	6L	Male	40	yes	8A/8A	yes	yes	
BRbS	Olost	8L	Male	40	yes	N/A			
BRbS	Olost	9L	Male	40	yes	10C/10C	yes	yes	
BRbS	Prat de Llobregat	91	Female	33	yes	N/A			
BRbS	Prat de Llobregat	931	Male	35	yes	N/A			
BRbS	Prat de Llobregat	932	Male	34	yes	10A/10A			
BRbS	Prat de Llobregat	933	Male	33	yes	10A/12B			

Origin	Population	ID	SEX	2N	Prdm9 study	PRDM9 seq	Recombination study	SNP analysis	Hi-C
BRbS	Prat de Llobregat	934	Male	33	yes	N/A			
BRbS	Prat de Llobregat	935	Female	31	yes	10A/12B			
BRbS	Prat de Llobregat	936	Female	31	yes	N/A			
BRbS	Prat de Llobregat	937	Female	32	yes	12B/12B			
BRbS	Prat de Llobregat	951	Male	31	yes	10A/10A			
BRbS	Prat de Llobregat	952	Male	32	yes	10A/10A			
BRbS	Prat de Llobregat	953	Female	32	yes	N/A			
BRbS	Sant Sadurni d'Anoia	SS12	Male	36	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS13	Male	38	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS17	Female	38	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS18	Male	37	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS19	Male	37	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS20	Male	39	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS22	Male	39	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS2	Male	35	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS3	Male	38	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS4	Male	38	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS5	Male	37	yes	10A/10A	yes	yes	
BRbS	Sant Sadurni d'Anoia	SS6	Male	37	yes	10A/10A	yes		
BRbS	Sant Sadurni d'Anoia	SS7	Male	38	yes	10A/10A	yes		
BRbS	Sant Sadurni d'Anoia	SS8	Male	37	yes	10A/10A	yes		
BRbS	Sant Sadurni d'Anoia	SS24	Male	N.D.	yes	N/A			
BRbS	Torreferrussa	900	Male	40	yes	10A/11B			
BRbS	Torreferrussa	902	Female	40	yes	N/A			
BRbS	Torreferrussa	903	Male	40	yes	11B/11B			
BRbS	Torreferrussa	904	Female	40	yes	10A/11B			
BRbS	Torreferrussa	905	Male	40	yes	10B/11B			
BRbS	Torreferrussa	915	Female	40	yes	10B/11B			
BRbS	Torreferrussa	916	Female	40	yes	10B/11B			
BRbS	Torreferrussa	917	Female	40	yes	10A/10A			
BRbS	Torreferrussa	T10	Female	40	yes	10A/11B			
BRbS	Torreferrussa	T11	Female	40	yes	10B/11B			
BRbS	Torreferrussa	T12	Male	40	yes	11B/11B			
BRbS	Torreferrussa	T14	Female	40	yes	N/A			
BRbS	Torreferrussa	T15	Female	40	yes	11B/11B			
BRbS	Torreferrussa	T16	Female	40	yes	10B/11B			
BRbS	Torreferrussa	T17	Male	40	yes	11B/11B			

Table S4 – *Continued from previous page*

Origin	Population	ID	SEX	2N	Prdm9 study	PRDM9 seq	Recombination study	SNP analysis	Hi-C
BRbS	Torreferrussa	T18	Female	40	yes	10A/10A			
BRbS	Torreferrussa	T19	Female	40	yes	11B/11B			
BRbS	Torreferrussa	T22	Male	40	yes	11B/11B			
BRbS	Torreferrussa	T9	Male	40	yes	11B/11B			
BRbS	Viladecans	20170711/03	Female	N.D.	yes	12N/12N			
BRbS	Viladecans	20170711/04	Female	N.D.	yes	N/A			
BRbS	Viladecans	20170711/05	Female	N.D.	yes	13K/13K			
BRbS	Viladecans	20170711/06	Female	N.D.	yes	N/A			
BRbS	Viladecans	20170711/07	Female	N.D.	yes	10A/10A			
BRbS	Viladecans	20170711/09	Female	N.D.	yes	N/A			
BRbS	Viladecans	20170711/10	Female	N.D.	yes	N/A			
BRbS	Viladecans	20170717/01	Female	N.D.	yes	N/A			
BRbS	Viladecans	20170717/02	Male	33	yes	N/A	yes		Germ cells Hi-C #1
BRbS	Viladecans	20170717/03	Male	32	yes	N/A	yes		Germ cells Hi-C #1
BRbS	Viladecans	20170717/04	Male	32	yes	N/A	yes		Germ cells Hi-C #1
BRbS	Viladecans	20170717/05	Male	32	yes	13L/13L	yes		Germ cells Hi-C #1
BRbS	Viladecans	20170717/07	Female	N.D.	yes	10A/10A			
BRbS	Viladecans	20170721/01	Male	32	yes	10A/10A	yes		
BRbS	Viladecans	20170727/01	Male	33	yes	N/A	yes		
BRbS	Vilanova i la Geltrú	20170906/03	Male	32		N/A	yes		
BRbS	Viladecans	20170919/12	Male	31		N/A	yes		Germ cells Hi-C #2
BRbS	Viladecans	20170919/16	Male	31		N/A	yes		Germ cells Hi-C #2
BRbS	Viladecans	20170919/17	Male	32		N/A	yes		
BRbS	Viladecans	20170919/19	Male	31		N/A	yes		Germ cells Hi-C #2
BRbS	Viladecans	20170919/20	Male	31		N/A	yes		Germ cells Hi-C #2
BRbS	Viladecans	20170919/21	Male	32		N/A	yes		
BRbS	Viladecans	946	Male	32	yes	10A/10A	yes		
BRbS	Viladecans	947	Male	32		N/A	yes		

A.1.4.2 Mice from the Madeiran archipelago

List of samples from the Madeiran archipelago used in the present work, including information about their origin, population, identification (ID), sex, 2N, chromosomal race and its PRDM9 sequence if available. N/A: not analyzed. N.D: not determined.

Origin	Population	ID	SEX	2N	Chromosomal race	PRDM9 seq
Madeira Island	Achadas da Cruz	P30.1	Male	24	PADC	10A/12L
Madeira Island	Achadas da Cruz	P30.13	Male	26	PADC	10E/10E
Madeira Island	Achadas da Cruz	P30.15	Male	24	PADC	11R/11R
Madeira Island	Achadas da Cruz	P30.4	Male	26	PADC	12I/12I
Madeira Island	Achadas da Cruz	P39.2	Male	26	PADC	10A/11K
Madeira Island	Achadas da Cruz	P39.4	Male	27	PADC	13C/16A
Madeira Island	Achadas da Cruz	P30.9	Female	N.D.	PADC	10A/13C
Madeira Island	Achadas da Cruz	P39.1	Female	25	PADC	N/A
Madeira Island	Arco da Calheta	P56.2	Male	24	PEDC_SOUTH	13A/13A
Madeira Island	Arco da Calheta	P57.1	Male	24	PEDC_SOUTH	10E/10E
Madeira Island	Arco da Calheta	P57.3	Male	24	PEDC_SOUTH	11K/13C
Madeira Island	Arco da Calheta	P57.4	Female	24	PEDC_SOUTH	13C/13C
Madeira Island	Arco da Calheta	P58.1	Male	24	PEDC_SOUTH	13C/13C
Madeira Island	Arco da Calheta	P58.2	Female	24	PEDC_SOUTH	10A/11P
Madeira Island	Arco da Calheta	P59.1	Male	24	PEDC_SOUTH	10E/13C
Madeira Island	Arco da Calheta	P60.6	Male	23	PEDC_SOUTH	10G/11T
Madeira Island	Arco da Calheta	P62.2	Female	24	PEDC_SOUTH	11K/11K
Madeira Island	Arco da Calheta	P61.12	Male	24	PEDC_SOUTH	11K/15F
Madeira Island	Arco da Calheta	P61.7	Male	24	PEDC_SOUTH	11S/13C
Madeira Island	Arco da Calheta	P56.6	Male	24	PEDC_SOUTH	N/A
Madeira Island	Arco da Calheta	P59.2	Male	24	PEDC_SOUTH	N/A
Madeira Island	Arco da Calheta	P60.4	Female	23	PEDC_SOUTH	N/A
Madeira Island	Arco da Calheta	P62.1	Male	24	PEDC_SOUTH	N/A
Madeira Island	Arco da Calheta	P62.3	Female	24	PEDC_SOUTH	N/A
Madeira Island	Arco da Calheta	P61.6	Female	24	PEDC_SOUTH	N/A
Madeira Island	Canhas	P86.1	Male	24	PLDB	N/A
Madeira Island	Canhas	P86.2	Male	24	PLDB	N/A
Madeira Island	Chão da Ribeira	P69.1	Female	23	PEDC_NORTH	10A/10A
Madeira Island	Chão da Ribeira	P69.2	Male	25	PEDC_NORTH	10A/11K
Madeira Island	Chão da Ribeira	P69.6	Male	24	PEDC_NORTH	10A/13C
Madeira Island	Chão da Ribeira	P83.1	Male	24	PEDC_NORTH	10A/10A

Table S5 – *Continued from previous page*

Origin	Population	ID	SEX	2N	Chromosomal race	PRDM9 seq
Madeira Island	Chão da Ribeira	P83.3	Male	24	PEDC_NORTH	13F/14F
Madeira Island	Chão da Ribeira	P70.1	Female	23	PEDC_NORTH	10A/10A
Madeira Island	Chão da Ribeira	P70.13	Male	23	PEDC_NORTH	10E/10E
Madeira Island	Chão da Ribeira	P70.4	Female	23	PEDC_NORTH	10A/10A
Madeira Island	Chão da Ribeira	P70.6	Male	24	PEDC_NORTH	10E/10E
Madeira Island	Chão da Ribeira	P68.2	Male	25	PEDC_NORTH	13G/13G
Madeira Island	Chão da Ribeira	P68.5	Male	23	PEDC_NORTH	10A/11B
Madeira Island	Chão da Ribeira	P68.8	Female	23	PEDC_NORTH	10A/11K
Madeira Island	Chão da Ribeira	P82.1	Male	24	PEDC_NORTH	N/A
Madeira Island	Chão da Ribeira	P21.7	Female	24	PEDC_NORTH	10E/10E
Madeira Island	Chão da Ribeira	P21.8	Female	23	PEDC_NORTH	10A/13G
Madeira Island	Chão da Ribeira	P21.10	Male	23	PEDC_NORTH	N/A
Madeira Island	Chão da Ribeira	P21.9	Male	23	PEDC_NORTH	N/A
Madeira Island	Chão da Ribeira	P22.2	Female	24	PEDC_NORTH	13D/13D
Madeira Island	Chão da Ribeira	P22.4	Male	23	PEDC_NORTH	13B/15A
Madeira Island	Chão da Ribeira	P22.3	Female	23	PEDC_NORTH	N/A
Madeira Island	Estreito da Calheta	P74.14	Female	24	PEDC_SOUTH	N/A
Madeira Island	Estreito da Calheta	P74.3	Male	24	PEDC_SOUTH	N/A
Madeira Island	Estreito da Calheta	P73.1	Male	25	PEDC_SOUTH	10A/10A
Madeira Island	Estreito da Calheta	P73.6	Female	26	PEDC_SOUTH	11N/11N
Madeira Island	Estreito da Calheta	P73.2	Male	26	PEDC_SOUTH	N/A
Madeira Island	Estreito da Calheta	P73.3	Male	25	PEDC_SOUTH	N/A
Madeira Island	Estreito da Calheta	P73.4	Male	26	PEDC_SOUTH	N/A
Madeira Island	Estreito da Calheta	P5.2	Male	24	PEDC_SOUTH	11B/11B
Madeira Island	Estreito da Calheta	P5.3	Female	26	PEDC_SOUTH	N/A
Madeira Island	Faial	P54.1	Female	22	PSAN	10A/10A
Madeira Island	Faial	P54.2	Male	22	PSAN	11H/11H
Madeira Island	Faial	P54.5	Male	22	PSAN	N/A
Madeira Island	Fajã da Ovelha	P43.2	Male	25	PADC	11N/11N
Madeira Island	Fajã da Ovelha	P43.3	Female	25	PADC	11K/13C
Madeira Island	Fajã da Ovelha	P43.1	Female	26	PADC	N/A
Madeira Island	Levada Grande	P71.1	Female	27	PADC	13J/13J
Madeira Island	Levada Grande	P71.2	Male	27	PADC	14A/14A
Madeira Island	Lombada dos Cedros	P14.1	Female	26	PADC	11K/11K
Madeira Island	Lombada dos Marinheiros	P19.1	Male	26	PADC	13E/13E
Madeira Island	Lombada dos Marinheiros	P19.3	Female	26	PADC	11K/11K
Madeira Island	Lombada dos Marinheiros	P19.6	Male	25	PADC	11K/13C
Madeira Island	Lombada dos Marinheiros	P19.7	Female	26	PADC	11F/11F

Continued on next page

Origin	Population	ID	SEX	2N	Chromosomal race	PRDM9 seq
Madeira Island	Lombada dos Marinheiros	P19.10	Female	25	PADC	N/A
Madeira Island	Lombo da Velha	P44.1	Female	26	Contact zone PEDC/PADC	11B/11B
Madeira Island	Lombo da Velha	P44.3	Female	26	Contact zone PEDC/PADC	11S/11S
Madeira Island	Lombo da Velha	P44.5	Male	24	Contact zone PEDC/PADC	13C/13C
Madeira Island	Lombo das Laranjeiras	P3.1	Female	24	PEDC_SOUTH	11S/11S
Madeira Island	Lombo das Laranjeiras	P3.2	Female	25	PEDC_SOUTH	10E/10E
Madeira Island	Lombo das Laranjeiras	P3.4	Male	24	PEDC_SOUTH	10A/12C
Madeira Island	Lombo do Doutor	P7.1	Female	24	PEDC_SOUTH	N/A
Madeira Island	Lombo do Doutor	P7.2	Female	25	PEDC_SOUTH	10A/11P
Madeira Island	Lombo do Doutor	P37.1	Female	24	PEDC_SOUTH	N/A
Madeira Island	Lombo do Doutor	P37.2	Female	24	PEDC_SOUTH	11U/15B
Madeira Island	Lombo do Doutor	P37.3	Male	24	PEDC_SOUTH	11S/11S
Madeira Island	Lugar da Raposeira	P80.3	Female	26	Contact zone PEDC/PADC	11G/11G
Madeira Island	Lugar da Raposeira	P81.12	Female	25	Contact zone PEDC/PADC	11R/11R
Madeira Island	Madalena do Mar	P11.7	Male	24	PEDC_SOUTH	N/A
Madeira Island	Madalena do Mar	P11.1	Female	24	PEDC_SOUTH	N/A
Madeira Island	Madalena do Mar	P11.5	Male	24	PEDC_SOUTH	11N/11N
Madeira Island	Madalena do Mar	P15.3	Female	24	PEDC	N/A
Madeira Island	Madalena do Mar	P15.1	Female	24	PEDC	11S/11S
Madeira Island	Maloeira	P45.1	Female	25	Contact zone PEDC/PADC	12K/13C
Madeira Island	Maloeira	P45.2	Female	25	Contact zone PEDC/PADC	11K/13C
Madeira Island	Maloeira	P78.1	Female	25	Contact zone PEDC/PADC	11K/13C
Madeira Island	Maloeira	P78.2	Male	26	Contact zone PEDC/PADC	10A/11S
Madeira Island	Maloeira	P78.3	Female	26	Contact zone PEDC/PADC	10A/13C
Madeira Island	Maloeira	P78.7	Male	26	Contact zone PEDC/PADC	11B/11B
Madeira Island	Maloeira	P78.8	Female	26	Contact zone PEDC/PADC	11S/13C
Madeira Island	Maloeira	P12.1	Male	26	Contact zone PEDC/PADC	N/A
Madeira Island	Maloeira	P12.4	Female	25	Contact zone PEDC/PADC	N/A
Madeira Island	Maloeira	P12.10	Female	25	Contact zone PEDC/PADC	10D/10D
Madeira Island	Maloeira	P12.14	Male	24	Contact zone PEDC/PADC	10A/11P
Madeira Island	Maloeira	P12.15	Female	26	Contact zone PEDC/PADC	11R/11R
Madeira Island	Maloeira	P12.2	Female	26	Contact zone PEDC/PADC	10A/11B
Madeira Island	Maloeira	P12.3	Male	26	Contact zone PEDC/PADC	11R/11R
Madeira Island	Maloeira	P12.5	Male	25	Contact zone PEDC/PADC	10A/11B
Madeira Island	Maloeira	P13.1	Male	26	Contact zone PEDC/PADC	N/A
Madeira Island	Maloeira	P13.2	Male	26	Contact zone PEDC/PADC	10E/10E
Madeira Island	Maloeira	P13.4	Female	26	Contact zone PEDC/PADC	10E/10E
Madeira Island	Maloeira	P17.1	Male	26	Contact zone PEDC/PADC	11B/12H

Continued on next page

Table S5 – *Continued from previous page*

Origin	Population	ID	SEX	2N	Chromosomal race	PRDM9 seq
Madeira Island	Maloeira	P17.2	Female	26	Contact zone PEDC/PADC	11S/11S
Madeira Island	Maloeira	P17.3	Female	26	Contact zone PEDC/PADC	11Q/11Q
Madeira Island	Moledos	P38.1	Male	24	PEDC_SOUTH	11B/11B
Madeira Island	Moledos	P38.2	Male	24	PEDC_SOUTH	11B/11B
Madeira Island	Ponta Delgada	P47.1	Male	28	PPOD	13I/13I
Madeira Island	Ponta Delgada	P47.2	Female	28	PPOD	13C/15D
Madeira Island	Ponta Delgada	P47.7	Male	28	PPOD	10A/13C
Madeira Island	Ponta Delgada	P50.1	Female	28	PPOD	11S/11S
Madeira Island	Ponta Delgada	P50.3	Male	28	PPOD	10E/10E
Madeira Island	Ponta Delgada	P50.6	Female	28	PPOD	11S/11S
Madeira Island	Ponta Delgada	P51.2	Female	27	PPOD	10A/10A
Madeira Island	Ponta Delgada	P51.8	Male	28	PPOD	10E/10E
Madeira Island	Ponta Delgada	P51.5	Male	28	PPOD	N/A
Madeira Island	Ponta Delgada	P49.1	Female	28	PPOD	N/A
Madeira Island	Ponta do Pargo	P67.1	Female	25	PADC	N/A
Madeira Island	Ponta do Pargo	P67.2	Male	25	PADC	N/A
Madeira Island	Ponta do Pargo	P67.4	Female	24	PADC	11B/12J
Madeira Island	Ponta do Pargo	P67.5	Male	24	PADC	11O/11O
Madeira Island	Ponta do Pargo	P67.6	Male	24	PADC	11K/13H
Madeira Island	Ponta do Pargo	P67.7	Female	24	PADC	11K/11K
Madeira Island	Ponta do Pargo	P66.12	Male	25	PADC	11K/11K
Madeira Island	Ponta do Pargo	P66.3	Male	24	PADC	11K/11K
Madeira Island	Ponta do Pargo	P66.5	Female	25	PADC	11K/11K
Madeira Island	Ponta do Pargo	P64.1	Male	25	PADC	10A/11K
Madeira Island	Ponta do Pargo	P64.2	Male	24	PADC	11K/11K
Madeira Island	Ponta do Pargo	P64.4	Female	25	PADC	11K/11K
Madeira Island	Ponta do Pargo	P41.1	Female	25	PADC	N/A
Madeira Island	Ponta do Pargo	P41.2	Male	24	PADC	11K/11K
Madeira Island	Ponta do Pargo	P41.3	Female	25	PADC	11K/13C
Madeira Island	Ponta do Pargo	P41.4	Male	24	PADC	10A/11K
Madeira Island	Ponta do Pargo	P40.1	Female	24	PADC	11K/12K
Madeira Island	Ponta do Pargo	P40.2	Male	24	PADC	11B/11B
Madeira Island	Ponta do Pargo	P40.4	Male	24	PADC	11K/11K
Madeira Island	Ponta do Pargo	P40.5	Female	25	PADC	N/A
Madeira Island	Ponta do Pargo	P20.1	Female	25	PADC	11K/11K
Madeira Island	Ponta do Pargo	P20.2	Male	25	PADC	11K/11K
Madeira Island	Ponta do Sol	P10.2	Male	24	PLDB	N/A
Madeira Island	Ponta do Sol	P10.1	Female	24	PLDB	10E/10E

Origin	Population	ID	SEX	2N	Chromosomal race	PRDM9 seq
Madeira Island	Ponta do Sol	P10.3	Female	24	PLDB	10A/10A
Madeira Island	Ponta do Sol	P10.4	Female	24	PLDB	11S/11S
Madeira Island	Porto da Cruz	P55.1	Female	22	PSAN	N/A
Madeira Island	Porto da Cruz	P55.2	Female	22	PSAN	N/A
Madeira Island	Porto da Cruz	P55.3	Male	22	PSAN	10E/10E
Madeira Island	Porto Moniz	P32.1	Female	28	PADC	11K/11K
Madeira Island	Porto Moniz	P32.2	Female	28	PADC	12M/12M
Madeira Island	Porto Moniz	P32.3	Female	28	PADC	11K/11K
Porto Santo island	Camacha	PS4.10	Female	40	STANDARD	11B/12C
Porto Santo island	Camacha	PS4.7	Female	40	STANDARD	11W/12C
Porto Santo island	Camacha	PS4.4	Male	40	STANDARD	N/A
Porto Santo island	Farrobo	PS3.1	Female	40	STANDARD	11V/12C
Porto Santo island	Ponta	PS1.5	Male	40	STANDARD	11S/12C
Porto Santo island	Ponta	PS1.6	Male	40	STANDARD	N/A
Porto Santo island	Ponta	PS1.1	Female	40	STANDARD	N/A
Porto Santo island	Quinta das Palmeiras	PS2.11	Male	40	STANDARD	11F/11F
Porto Santo island	Quinta das Palmeiras	PS2.14	Male	40	STANDARD	11S/12C
Porto Santo island	Quinta das Palmeiras	PS2.3	Female	40	STANDARD	11X/11X
Porto Santo island	Quinta das Palmeiras	PS2.9	Male	40	STANDARD	11S/11S
Madeira Island	Prazeres	P9.2	M	24	Contact zone PEDC/PADC	11S/11S
Madeira Island	Prazeres	P6.1	Female	25	Contact zone PEDC/PADC	11N/11N
Madeira Island	Ribeira da Janela	P72.2	Female	26	PEDC_SOUTH	11K/11K
Madeira Island	Ribeira da Janela	P72.3	Female	26	PEDC_SOUTH	11K/11K
Madeira Island	Ribeira da Janela	P72.4	Male	26	PEDC_SOUTH	11K/11K
Madeira Island	Ribeira da Janela	P33.B1	Female	26	PADC	13C/13C
Madeira Island	Ribeira da Janela	P33.B2	Female	26	PADC	11S/11S
Madeira Island	Ribeira da Janela	P33.B5	Male	26	PADC	16C/16C
Madeira Island	Ribeira da Janela	P72.5	Male	27	PEDC_SOUTH	N/A
Madeira Island	Ribeira da Laje	P24.2	Male	25	PEDC_NORTH	13D/13D
Madeira Island	Ribeira da Laje	P24.3	Female	26	PEDC_NORTH	11L/11L
Madeira Island	Ribeira da Laje	P24.4	Female	26	PEDC_NORTH	13A/13A
Madeira Island	Ribeira da Vaca	P65.17	Female	24	PADC	N/A
Madeira Island	Ribeira da Vaca	P65.1	Male	24	PADC	11I/11I
Madeira Island	Ribeira da Vaca	P65.15	Male	24	PADC	11K/11K
Madeira Island	Ribeira da Vaca	P65.16	Female	24	PADC	11I/11I
Madeira Island	Ribeira Funda	P25.1	Female	26	PEDC_NORTH	N/A
Madeira Island	Ribeira Funda	P25.2	Male	25	PEDC_NORTH	11S/11S
Madeira Island	Ribeira Funda	P25.3	Male	26	PEDC_NORTH	10F/10F

Table S5 – *Continued from previous page*

Origin	Population	ID	SEX	2N	Chromosomal race	PRDM9 seq
Madeira Island	Ribeira Funda	P25.4	Female	26	PEDC_NORTH	10E/10E
Madeira Island	Santa	P31.6	Female	25	PADC	11K/11K
Madeira Island	Santa	P31.1	Female	24	PADC	N/A
Madeira Island	Santa	P31.4	Male	24	PADC	N/A
Madeira Island	Santa	P31.5	Female	26	PADC	N/A
Madeira Island	Santana	P53.1	Female	22	PSAN	10E/10E
Madeira Island	Santana	P53.2	Male	22	PSAN	10A/13C
Madeira Island	Santana	P52.4	Female	22	PSAN	10A/10A
Madeira Island	Santana	P52.5	Male	22	PSAN	10E/10E
Madeira Island	São Vicente	P34.7	Male	26	PSVI	N/A
Madeira Island	São Vicente	P34.3	Female	26	PSVI	11K/13C
Madeira Island	São Vicente	P34.4	Female	27	PSVI	16B/16B
Madeira Island	São Vicente	P34.6	Female	26	PSVI	13B/15E
Madeira Island	Seixal	P23.1	Female	24	PEDC_NORTH	13C/13C
Madeira Island	Sítio da Fajã	P36.10	Male	24	PEDC_SOUTH	11F/11F
Madeira Island	Sítio da Fajã	P36.12	Male	24	PEDC_SOUTH	11K/11K
Madeira Island	Sítio da Fajã	P36.20	Female	24	Híb PEDC_SOUTH/PLDB	11N/11N
Madeira Island	Socorro	P75.2	Female	24	PLDB	N/A
Madeira Island	Socorro	P76.1	Male	24	PLDB	N/A
Madeira Island	Socorro	P75.1	Male	24	PLDB	11J/11J
Madeira Island	Socorro	P75.3	Male	24	PLDB	11K/11K
Madeira Island	Socorro	P75.4	Female	24	PLDB	10A/11K
Madeira Island	Socorro	P76.2	Male	24	PLDB	10A/13C
Madeira Island	Socorro	P76.3	Male	24	Hyb PEDC/PLDB	13J/15C
Madeira Island	Socorro	P76.4	Female	24	PLDB	10H/10H
Madeira Island	Socorro	P76.5	Female	24	PLDB	10A/10A
Madeira Island	Solar da Maloeira	P46.2	Male	25	Contact zone PEDC/PADC	N/A
Madeira Island	Solar da Maloeira	P46.1	Female	26	Contact zone PEDC/PADC	11N/11N
Madeira Island	Solar dos Prazeres	P77.3	Female	26	Contact zone PEDC/PADC	11M/11M

Bibliography

Here are the references in alphabetical order.

Acemel, R. D., Maeso, I., & Gómez-Skarmeta, J. L. (2017). Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *Wiley Interdisciplinary Reviews: Developmental Biology*, 1–19. https://doi.org/10.1002/WDEV.265 (cited on page 9)

Acquaviva, L., Boekhout, M., Karasu, M. E., Brick, K., Pratto, F., Li, T., van Overbeek, M., Kauppi, L., Camerini-Otero, R. D., Jasin, M., & Keeney, S. (2020). Ensuring meiotic DNA break formation in the mouse pseudoautosomal region. *Nature*, 582(7812), 426–431. https://doi.org/10.1038/s41586-020-2327-4 (cited on page 19)

Adolph, S., & Klein, J. (1981). Robertsonian variation in Mus musculus from Central Europe Spain, and Scotland. *The Journal of Heredity*, 72(3), 219–21 (cited on pages 23, 198, 203).

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1), W537–W544. https://doi.org/10.1093/nar/gky379 (cited on page 52)

Agostinho, A., Manneberg, O., van Schendel, R., Hernández-Hernández, A., Kouznetsova, A., Blom, H., Brismar, H., & Höög, C. (2016). High density of REC8 constrains sister chromatid axes and prevents illegitimate synaptonemal complex formation. *EMBO Reports*, 17(6), 901–913. https://doi.org/10.15252/embr.201642030 (cited on pages 15, 195)

Alavattam, K. G., Maezawa, S., Sakashita, A., Khoury, H., Barski, A., Kaplan, N., & Namekawa, S. H. (2019). Attenuated chromatin compartmentalization in meiosis and its maturation in sperm development. *Nature Structural & Molecular Biology*, 26, 175–184. https://doi.org/10.1038/s41594-019-0189-y (cited on pages 119, 194, 195)

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. https://doi.org/10.1101/gr.094052.109 (cited on pages 51, 136)

Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., & Weiner, O. D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife*, 8. https://doi.org/10.7554/eLife.41769 (cited on page 195)

Alexander, T. A., & Machiela, M. J. (2020). LDpop: An interactive online tool to calculate and visualize geographic LD patterns. *BMC Bioinformatics*, 21(1), 1–4. https://doi.org/10.1186/s12859-020-3340-1 (cited on pages 51, 137)

Aloisio, G. M., Nakada, Y., Saatcioglu, H. D., Pena, C. G., Baker, M. D., Tarnawa, E. D., Mukherjee, J., Manjunath, H., Bugde, A., Sengupta, A. L., Amatruda, J. F., Cuevas, I., Hamra, F. K., & Castrillon, D. H. (2014). PAX7 expression defines germline stem cells in the adult testis. *The Journal of Clinical Investigation*, 124(9), 3929–3944. https://doi.org/10.1172/JCI75943 (cited on pages 14, 57)

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 (cited on page 42)

Arndt-Jovin, D. J., & Jovin, T. M. (1977). Analysis and sorting of living cells according to deoxyribonucleic acid content. *The Journal of Histochemistry and Cytochemistry*, 25(7), 585–589. https://doi.org/10.1177/25.7.70450 (cited on page 57)

Arnheim, N., Calabrese, P., & Tiemann-Boege, I. (2007). Mammalian Meiotic Recombination Hot Spots. *Annual Review of Genetics*, 41(1), 369–399. https://doi.org/10.1146/annurev.genet.41.110306.130301 (cited on page 21)

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., . . . Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, *40*(W1), W597–W603. https://doi.org/10.1093/nar/gks400 (cited on page 50)

Arzate-Mejía, R. G., Recillas-Targa, F., & Corces, V. G. (2018). Developing in 3D: the role of CTCF in cell differentiation. *Development*, *145*(6), dev137729. https://doi.org/10.1242/dev.137729 (cited on page 9)

Ashley, T., Gaeth, A. P., Inagaki, H., Seftel, A., Cohen, M. M., Anderson, L. K., Kurahashi, H., & Emanuel, B. S. (2006). Meiotic recombination and spatial proximity in the etiology of the recurrent t(11;22). *American Journal of Human Genetics*, *79*(3), 524–538. https://doi.org/10.1086/507652 (cited on page 199)

Ayala, F. J., & Coluzzi, M. (2005). Chromosome speciation: humans, Drosophila, and mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, *102 Suppl*(Suppl 1), 6535–6542. https://doi.org/10.1073/pnas.0501847102 (cited on page 22)

Baier, B., Hunt, P., Broman, K. W., & Hassold, T. (2014). Variation in genome-wide levels of meiotic recombination is established at the onset of prophase in mammalian males (M. Przeworski, Ed.). *PLoS Genetics*, *10*(1), e1004125. https://doi.org/10.1371/journal.pgen.1004125 (cited on page 15)

Bailey, C., Metierre, C., Feng, Y., Baidya, K., Filippova, G., Loukinov, D., Lobanenkov, V., Semaan, C., & Rasko, J. (2018). CTCF Expression is Essential for Somatic Cell Viability and Protection Against Cancer. *International Journal of Molecular Sciences*, *19*(12), 3832. https://doi.org/10.3390/ijms19123832 (cited on page 9)

Baker, C. L., Kajita, S., Walker, M., Saxl, R. L., Raghupathy, N., Choi, K., Petkov, P. M., & Paigen, K. (2015). PRDM9 Drives Evolutionary Erosion of Hotspots in Mus musculus through Haplotype-Specific Initiation of Meiotic Recombination. *PLoS Genetics*, *11*(1), e1004916. https://doi.org/10.1371/journal.pgen.1004916 (cited on page 203)

Baker, C. L., Walker, M., Kajita, S., Petkov, P. M., & Paigen, K. (2014). PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Research*, *24*(5), 724–732. https://doi.org/10.1101/gr.170167.113 (cited on page 20)

Baker, Z., Schumer, M., Haba, Y., Bashkirova, L., Holland, C., Rosenthal, G. G., & Przeworski, M. (2017). Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife*, *6*, e24133. https://doi.org/10.7554/eLife.24133 (cited on pages 19–21)

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. *Genome Biology*, *8*(9), 227. https://doi.org/10.1186/gb-2007-8-9-227 (cited on pages 17, 196)

Balhorn, R. (2011). Sperm Chromatin: An Overview. In *Sperm chromatin* (pp. 3–18). New York, NY, Springer New York. https://doi.org/10.1007/978-1-4419-6857-9{_}1. (Cited on pages 17, 196, 197)

Bao, J., & Bedford, M. T. (2016). Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. *Reproduction*, *151*(5), R55–R70. https://doi.org/10.1530/REP-15-0562 (cited on pages 16, 17, 196)

Barchi, M., Geremia, R., Magliozzi, R., & Bianchi, E. (2009). Isolation and Analyses of Enriched Populations of Male Mouse Germ Cells by Sedimentation Velocity: The Centrifugal Elutriation. In *Methods in molecular biology* (pp. 299–321). Totowa, NJ, Humana Press. https://doi.org/10.1007/978-1-60761-103-5{_}18. (Cited on page 57)

Barrington, C., Georgopoulou, D., Pezic, D., Varsally, W., Herrero, J., & Hadjur, S. (2019). Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. *Nature Communications*, *10*(1), 1–14. https://doi.org/10.1038/s41467-019-10725-9 (cited on pages 8, 129)

Barrington, C., Pezic, D., & Hadjur, S. (2017). Chromosome structure dynamics during the cell cycle: a structure to fit every phase. *The EMBO Journal*, *36*(18), 2661–2663. https://doi.org/10.15252/embj.201798014 (cited on page 11)

Bashkirova, E., & Lomvardas, S. (2019). Olfactory receptor genes make the case for inter-chromosomal interactions. *Current Opinion in Genetics and Development*, *55*, 106–113. https://doi.org/10.1016/j.gde.2019.07.004 (cited on page 200)

Bastiaan Holwerda, S. J., & de Laat, W. (2013). CTCF: The protein, the binding partners, the binding sites and their chromatin loops. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1620). https://doi.org/10.1098/rstb.2012.0369 (cited on page 9)

Bastos, H., Lassalle, B., Chicheportiche, A., Riou, L., Testart, J., Allemand, I., & Fouchet, P. (2005). Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. *Cytometry Part A*, *65*(1), 40–49. https://doi.org/10.1002/cyto.a.20129 (cited on pages 57, 66)

Bateson, W. (1909). Heredity and variation in modern lights. *Darwin and modern science* (cited on page 22).

Battulin, N., Fishman, V. S., Mazur, A. M., Pomaznoy, M., Khabarova, A. A., Afonnikov, D. A., Prokhortchouk, E. B., & Serov, O. L. (2015). Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. *Genome Biology*, *16*(1), 77. https://doi.org/10.1186/s13059-015-0642-0 (cited on page 197)

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., & de Massy, B. (2010). PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. *Science*, *327*(5967), 836–840. https://doi.org/10.1126/science.1183439 (cited on pages 18–21)

Baudat, F., Imai, Y., & de Massy, B. (2013). Meiotic recombination in mammals: localization and regulation. *Nature Reviews Genetics*, *14*(11), 794–806. https://doi.org/10.1038/nrg3573 (cited on pages 19–21, 123)

Beadle, G. W. (1932). A Possible Influence of the Spindle Fibre on Crossing-Over in Drosophila. *Proceedings of the National Academy of Sciences*, *18*(2), 160 LP –165. https://doi.org/10.1073/pnas.18.2.160 (cited on page 19)

Beagan, J. A., & Phillips-Cremins, J. E. (2020). On the existence and functionality of topologically associating domains. *Nature Genetics*, *52*(1), 8–16. https://doi.org/10.1038/s41588-019-0561-1 (cited on pages 8, 11)

Beliveau, B. J., Boettiger, A. N., Avendaño, M. S., Jungmann, R., McCole, R. B., Joyce, E. F., Kim-Kiselak, C., Bantignies, F., Fonseka, C. Y., Erceg, J., Hannan, M. A., Hoang, H. G., Colognori, D., Lee, J. T., Shih, W. M., Yin, P., Zhuang, X., & Wu, C.-t. (2015). Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nature Communications*, *6*(1), 7147. https://doi.org/10.1038/ncomms8147 (cited on page 4)

Belloc, F., Dumain, P., Boisseau, M. R., Jalloustre, C., Reiffers, J., Bernard, P., & Lacombe, F. (1994). A flow cytometric method using Hoechst 33342 and propidium iodide for simultaneous cell cycle analysis and apoptosis determination in unfixed cells. *Cytometry*, *17*(1), 59–65. https://doi.org/10.1002/cyto.990170108 (cited on page 57)

Bellvé, A. R. (1993). Purification, Culture and Fractionation of Spermatogenic Cells. *Methods in Enzymology*, *225*, 84–113. https://doi.org/10.1016/0076-6879(93)25009-Q (cited on page 67)

Bellvé, A. R., Cavicchia, J., Millette, C., O'Brien, D., Bhatnagar, Y., & Dym, M. (1977). Spermatogenic cells of the prepuberal mouse: isolation and morphological characterization. *The Journal of Cell Biology*, *74*(1), 68–85. https://doi.org/10.1083/jcb.74.1.68 (cited on pages 12, 14, 57)

Belton, J.-m. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi – C : A comprehensive technique to capture the conformation of genomes. *Methods*, *58*(3), 1–9. https://doi.org/10.1016/j.ymeth.2012.05.001 (cited on page 37)

Bendel–Stenzel, M., Anderson, R., Heasman, J., & Wylie, C. (1998). The origin and migration of primordial germ cells in the mouse. *Seminars in Cell & Developmental Biology*, *9*(4), 393–400. https://doi.org/10.1006/scdb.1998.0204 (cited on page 12)

Berg, I. L., Neumann, R., Lam, K.-W. G., Sarbajna, S., Odenthal-Hesse, L., May, C. A., & Jeffreys, A. J. (2010). PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genetics*, *42*(10), 859–863. https://doi.org/10.1038/ng.658 (cited on pages 20, 21)

Berg, I. L., Neumann, R., Sarbajna, S., Odenthal-Hesse, L., Butler, N. J., & Jeffreys, A. J. (2011). Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. *Proceedings of the National Academy of Sciences*, *108*(30), 12378–12383. https://doi.org/10.1073/pnas.1109531108 (cited on page 21)

Berr, A., Pecinka, A., Meister, A., Kreth, G., Fuchs, J., Blattner, F. R., Lysak, M. A., & Schubert, I. (2006). Chromosome arrangement and nuclear architecture but not centromeric sequences are conserved

between Arabidopsis thaliana and Arabidopsis lyrata. *The Plant Journal*, *48*(5), 771–783. https://doi.org/10.1111/j.1365-313X.2006.02912.x (cited on page 3)

Berríos, S. (2017). Nuclear Architecture of Mouse Spermatocytes: Chromosome Topology, Heterochromatin, and Nucleolus. *Cytogenetic and Genome Research*, *151*(2), 61–71. https://doi.org/10.1159/000460811 (cited on page 199)

Berríos, S., Fernández-Donoso, R., & Ayarza, E. (2017). Synaptic configuration of quadrivalents and their association with the XY bivalent in spermatocytes of Robertsonian heterozygotes of Mus domesticus. *Biological Research*, *50*(1), 1–6. https://doi.org/10.1186/s40659-017-0143-6 (cited on pages 25, 199)

Berríos, S., Manieu, C., López-Fenner, J., Ayarza, E., Page, J., González, M., Manterola, M., & Fernández-Donoso, R. (2014). Robertsonian chromosomes and the nuclear architecture of mouse meiotic prophase spermatocytes. *Biological Research*, *47*(1), 16. https://doi.org/10.1186/0717-6287-47-16 (cited on pages 133, 199, 201)

Berríos, S., Manterola, M., Prieto, Z., Lopez-Fenner, J., Page, J. J., Fernandez-Donoso, R., Berríos, S., Manterola, M., Prieto, Z., López-Fenner, J., Page, J. J., & Fernández-Donoso, R. (2010). Model of chromosome associations in mus domesticus spermatocytes. *Biological Research*, *43*(3), 275–286. https://doi.org/10.4067/S0716-97602010000300003 (cited on page 199)

Bhattacharyya, T., Gregorova, S., Mihola, O., Anger, M., Sebestova, J., Denny, P., Simecek, P., & Forejt, J. (2013). Mechanistic basis of infertility of mouse intersubspeci fi c hybrids. *Proc Natl Acad Sci U S A*. https://doi.org/10.1073/pnas.1219126110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1219126110 (cited on page 22)

Bikchurina, T. I., Tishakova, K. V., Kizilova, E. A., Romanenko, S. A., Serdyukova, N. A., Torgasheva, A. A., & Borodin, P. M. (2018). Chromosome Synapsis and Recombination in Male-Sterile and Female-Fertile Interspecies Hybrids of the Dwarf Hamsters (Phodopus, Cricetidae). *Genes*, *9*(5). https://doi.org/10.3390/genes9050227 (cited on page 201)

Billings, T., Parvanov, E. D., Baker, C. L., Walker, M., Paigen, K., & Petkov, P. M. (2013). DNA binding specificities of the long zinc-finger recombination protein PRDM9. *Genome Biology*, *14*(4), R35. https://doi.org/10.1186/gb-2013-14-4-r35 (cited on pages 20, 203)

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature*, *321*(6067), 209–213. https://doi.org/10.1038/321209a0 (cited on page 7)

Blanco, M., & Cocquet, J. (2019). Genetic Factors Affecting Sperm Chromatin Structure. In *Human reproduction* (pp. 1–28). https://doi.org/10.1007/978-3-030-21664-1{_}1. (Cited on page 196)

Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A., & Greider, C. W. (1997). Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell*, *91*(1), 25–34. https://doi.org/10.1016/s0092-8674(01)80006-4 (cited on page 23)

Boateng, K. A., Bellani, M. A., Gregoretti, I. V., Pratto, F., & Camerini-Otero, R. D. (2013). Homologous Pairing Preceding SPO11-Mediated Double-Strand Breaks in Mice. *Developmental Cell*, *24*(2), 196–205. https://doi.org/10.1016/j.devcel.2012.12.002 (cited on pages 14, 18)

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., Imakaev, M., Mirny, L. A., Wu, C.-t. T., & Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, *529*(7586), 418–422. https://doi.org/10.1038/nature16496 (cited on page 8)

Bolcun-Filas, E., & Schimenti, J. C. (2012). Genetics of Meiosis and Recombination in Mice. In *International review of cell and molecular biology* (pp. 179–227). https://doi.org/10.1016/B978-0-12-394309-5.00005-5. (Cited on page 19)

Bompadre, O., & Andrey, G. (2019). Chromatin topology in development and disease. *Current Opinion in Genetics & Development*, *55*, 32–38. https://doi.org/10.1016/j.gde.2019.04.007 (cited on page 119)

Bonev, B., & Cavalli, G. (2016). Organization and function of the 3D genome. *Nature Reviews Genetics*, *17*(11), 661–678. https://doi.org/10.1038/nrg.2016.112 (cited on page 9)

Borde, V., & de Massy, B. (2013). Programmed induction of DNA double strand breaks during meiosis: setting up communication between DNA and the chromosome structure. *Current Opinion in Genetics & Development*, *23*(2), 147–155. https://doi.org/10.1016/j.gde.2012.12.002 (cited on page 19)

Borodin, P. M., Karamysheva, T. V., Belonogova, N. M., Torgasheva, A. A., Rubtsov, N. B., & Searle, J. B. (2008). Recombination map of the common shrew, Sorex araneus (Eulipotyphla, Mammalia). *Genetics*, *178*(2), 621–632. https://doi.org/10.1534/genetics.107.079665 (cited on pages 23, 133)

Boulton, a., Myers, R. S., & Redfield, R. J. (1997). The hotspot conversion paradox and the evolution of meiotic recombination. *PNAS*, *94*(15), 8058–8063. https://doi.org/10.1073/pnas.94.15.8058 (cited on page 21)

Boutanaev, A. M., Mikhaylova, L. M., & Nurminsky, D. I. (2005). The pattern of chromosome folding in interphase is outlined by the linear gene density profile. *Molecular and cellular biology*, *25*(18), 8379–8386. https://doi.org/10.1128/MCB.25.18.8379-8386.2005 (cited on page 4)

Boveri, T. (1909). Die blastomerenkerne von ascaris megalocephala und die theorie der chromosomenindividualität. *Arch Zellforsch*, *3*, 181–268 (cited on page 3).

Boyle, S., Gilchrist, S., Bridger, J. M., Mahy, N. L., Ellis, J. A., & Bickmore, W. A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human molecular genetics*, *10*(3), 211–219. https://doi.org/10.1093/hmg/10.3.211 (cited on page 4)

Branco, M. R., & Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biology*, *4*(5), e138. https://doi.org/10.1371/journal.pbio.0040138 (cited on page 199)

Brand, C. L., & Presgraves, D. C. (2016). Evolution: On the Origin of Symmetry, Synapsis, and Species. *Current Biology*, *26*(8), R325–R328. https://doi.org/10.1016/j.cub.2016.03.014 (cited on page 203)

Braun, R. E. (2001). Packaging paternal chromosomes with protamine. *Nature Genetics*, *28*(1), 10–12. https://doi.org/10.1038/ng0501-10 (cited on page 17)

Brewer, L. R. (2011). Deciphering the structure of DNA toroids. *Integrative Biology*, *3*(5), 540–547. https://doi.org/10.1039/c0ib00128g (cited on page 197)

Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R. D., & Petukhova, G. V. (2012). Genetic recombination is directed away from functional genomic elements in mice. *Nature*, *485*(7400), 642–645. https://doi.org/nature11089[pii]{\textbackslash}r10.1038/nature11089 (cited on page 21)

Brinkley, B. R., Brenner, S. L., Hall, J. M., Tousson, A., Balczon, R. D., & Valdivia, M. M. (1986). Arrangements of kinetochores in mouse cells during meiosis and spermiogenesis. *Chromosoma*, *94*(4), 309–317. https://doi.org/10.1007/BF00290861 (cited on pages 16, 129, 197, 199)

Britton-Davidian, J., Catalan, J., Lopez, J., Ganem, G., Nunes, A. C., Ramalhinho, M. G. M. G., Auffray, J. C., Searle, J. B. J., & Mathias, M. L. (2007). Patterns of genic diversity and structure in a species undergoing rapid chromosomal radiation: an allozyme analysis of house mice from the Madeira archipelago. *Heredity*, *99*(4), 432–442. https://doi.org/10.1038/sj.hdy.6801021 (cited on pages 23–25, 202)

Britton-Davidian, J., Catalan, J., da Graça Ramalhinho, M., Ganem, G., Auffray, J.-C., Capela, R., Biscoito, M., Searle, J. B., & da Luz Mathias, M. (2000). Rapid chromosomal evolution in island mice. *Nature*, *403*(6766), 158–158. https://doi.org/10.1038/35003116 (cited on pages 23–25, 34, 202)

Brown, J. D., & O'Neill, R. J. (2010). Chromosomes, Conflict, and Epigenetics: Chromosomal Speciation Revisited. *Annual Review of Genomics and Human Genetics*, *11*(1), 291–316. https://doi.org/10.1146/annurev-genom-082509-141554 (cited on page 22)

Bryant, J. M., Meyer-Ficca, M. L., Dang, V. M., Berger, S. L., & Meyer, R. G. (2013). Separation of Spermatogenic Cell Types Using STA-PUT Velocity Sedimentation. *Journal of Visualized Experiments*, (80), 1–9. https://doi.org/10.3791/50648 (cited on pages 57, 193)

Buard, J., Rivals, E., Dunoyer de Segonzac, D., Garres, C., Caminade, P., de Massy, B., & Boursot, P. (2014). Diversity of Prdm9 Zinc Finger Array in Wild Mice Unravels New Facets of the Evolutionary Turnover of this Coding Minisatellite (M. Lichten, Ed.). *PLoS ONE*, *9*(1), e85021. https://doi.org/10.1371/journal.pone.0085021 (cited on pages 19, 21, 22, 51, 202)

Burgoyne, P. S. (1982). Genetic homology and crossing over in the X and Y chromosomes of Mammals. *Human Genetics*, *61*(2), 85–90. https://doi.org/10.1007/bf00274192 (cited on page 16)

Burgoyne, P. S., Mahadevaiah, S. K., & Turner, J. M. A. (2009). The consequences of asynapsis for mammalian meiosis. *Nature Reviews Genetics*, *10*(3), 207–216. https://doi.org/10.1038/nrg2505 (cited on pages 16, 25)

Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., & Chen, D. J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *The Journal of Biological Chemistry*, *276*(45), 42462–42467. https://doi.org/10.1074/jbc.C100466200 (cited on page 18)

Bushnell, B. (2014). BBmap: A fast, accurate, splice-aware aligner. https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner. (Cited on page 52)

Busslinger, G. A., Stocsits, R. R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., & Peters, J.-M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*, *544*(7651), 503–507. https://doi.org/10.1038/nature22063 (cited on pages 8, 10)

Bystricky, K., Laroche, T., van Houwe, G., Blaszczyk, M., & Gasser, S. M. (2005). Chromosome looping in yeast. *Journal of Cell Biology*, *168*(3), 375–387. https://doi.org/10.1083/jcb.200409091 (cited on page 3)

Capilla, L. (2015). *On the role of chromosomal rearrangements in evolution : Reconstruction of genome reshuffling in rodents and analysis of Robertsonian fusions in a house mouse chromosomal polymorphism zone* (Doctoral dissertation). Universitat Autònoma de Barcelona. (Cited on page 20).

Capilla, L., Garcia Caldés, M., & Ruiz-Herrera, A. (2016). Mammalian meiotic recombination: a toolbox for genome evolution. *Cytogenetic and Genome Research*, *150*(1), 1–16. https://doi.org/10.1159/000452822 (cited on pages 22, 133, 135, 194, 200)

Capilla, L., Medarde, N., Alemany-Schmidt, A., Oliver-Bonet, M., Ventura, J., & Ruiz-Herrera, A. (2014). Genetic recombination variation in wild Robertsonian mice: on the role of chromosomal fusions and Prdm9 allelic background. *Proceedings. Biological sciences / The Royal Society*, *281*(1786), 1–18. https://doi.org/10.1098/rspb.2014.0297 (cited on pages 19, 21–23, 51, 120, 122, 133, 135, 136, 200, 201)

Cappelletti, E., Piras, F. M., Badiale, C., Bambi, M., Santagostino, M., Vara, C., Masterson, T. A., Sullivan, K. F., Nergadze, S. G., Ruiz-Herrera, A., & Giulotto, E. (2019). CENP-A binding domains and recombination patterns in horse spermatocytes. *Scientific Reports*, *9*(1), 15800. https://doi.org/10.1038/s41598-019-52153-1 (cited on pages 133, 201)

Carrell, D. T., Emery, B. R., & Hammoud, S. (2007). Altered protamine expression and diminished spermatogenesis: what is the link? *Human Reproduction Update*, *13*(3), 313–327. https://doi.org/10.1093/humupd/dml057 (cited on page 17)

Castiglia, R., & Capanna, E. (2000). Contact zone between chromosomal races of Mus musculus domesticus. 2. Fertility and segregation in laboratory-reared and wild mice heterozygous for multiple robertsonian rearrangements. *Heredity*, *85 (Pt 2)*, 147–56 (cited on page 202).

Castiglia, R., & Capanna, E. (2002). Chiasma repatterning across a chromosomal hybrid zone between chromosomal races of Mus musculus domesticus. *Genetica*, *114*(1), 35–40. https://doi.org/10.1023/A:1014626330022 (cited on pages 23, 200)

Cattoglio, C., Pustova, I., Walther, N., Ho, J. J., Hantsche-Grininger, M., Inouye, C. J., Hossain, M. J., Dailey, G. M., Ellenberg, J., Darzacq, X., Tjian, R., & Hansen, A. S. (2019). Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *eLife*, *8*. https://doi.org/10.7554/eLife.40164 (cited on page 8)

Cavé, T., Desmarais, R., Lacombe-Burgoyne, C., & Boissonneault, G. (2019). Genetic Instability and Chromatin Remodeling in Spermatids. *Genes*, *10*(1), 40. https://doi.org/10.3390/genes10010040 (cited on page 17)

Chan, A. H., Jenkins, P. A., & Song, Y. S. (2012). Genome-Wide Fine-Scale Recombination Rate Variation in Drosophila melanogaster. *PLoS Genetics*, *8*(12). https://doi.org/10.1371/journal.pgen.1003090 (cited on pages 51, 137)

Chan, F., Oatley, M. J., Kaucher, A. V., Yang, Q.-E., Bieberich, C. J., Shashikant, C. S., & Oatley, J. M. (2014). Functional and molecular features of the Id4+ germline stem cell population in mouse testes. *Genes & development*, *28*(12), 1351–1362. https://doi.org/10.1101/gad.240465.114 (cited on page 14)

Chen, Y., Zheng, Y., Gao, Y., Lin, Z., Yang, S., Wang, T., Wang, Q., Xie, N., Hua, R., Liu, M., Sha, J., Griswold, M. D., Li, J., Tang, F., & Tong, M. H. (2018). Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Research*, *28*(9), 879–896. https://doi.org/10.1038/s41422-018-0074-y (cited on page 196)

Cheng, C. Y., & Mruk, D. D. (2002). Cell junction dynamics in the testis: Sertoli-germ cell interactions and male contraceptive development. *Physiological Reviews*, *82*(4), 825–874. https://doi.org/10.1152/physrev.00009.2002 (cited on pages 13, 66)

Chiquoine, A. D. (1954). The identification, origin, and migration of the primordial germ cells in the mouse embryo. *The Anatomical Record*, *118*(2), 135–146. https://doi.org/10.1002/ar.1091180202 (cited on page 12)

Chmátal, L., Gabriel, S. I., Mitsainas, G. P., Martínez-Vargas, J., Ventura, J., Searle, J. B., Schultz, R. M., & Lampson, M. A. (2014). Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Current Biology*, *24*(19), 2295–300. https://doi.org/10.1016/j.cub.2014.08.017 (cited on page 25)

Ciancio, G., Pollack, A., Taupier, M. A., Block, N. L., & Irvin, G. L. (1988). Measurement of cell-cycle phase-specific cell death using Hoechst 33342 and propidium iodide: preservation by ethanol fixation. *Journal of Histochemistry & Cytochemistry*, *36*(9), 1147–1152. https://doi.org/10.1177/36.9.2457047 (cited on page 58)

Clermont, Y. (1972). Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiological Reviews*, *52*(1), 198–236. https://doi.org/10.1152/physrev.1972.52.1.198 (cited on page 13)

Clermont, Y., & Leblond, C. P. (1953). Renewal of spermatogonia in the rat. *American Journal of Anatomy*, *93*(3), 475–501. https://doi.org/10.1002/aja.1000930308 (cited on page 14)

Clermont, Y., & Perey, B. (1957). Quantitative study of the cell population of the seminiferous tubules in immature rats. *American Journal of Anatomy*, *100*(2), 241–267. https://doi.org/10.1002/aja.1001000205 (cited on page 12)

Cole, F., Keeney, S., & Jasin, M. (2010). Comprehensive, Fine-Scale Dissection of Homologous Recombination Outcomes at a Hot Spot in Mouse Meiosis. *Molecular Cell*, *39*(5), 700–710. https://doi.org/10.1016/j.molcel.2010.08.017 (cited on page 19)

Comai, G., Boutet, A., Neirijnck, Y., & Schedl, A. (2010). Expression patterns of the Wtx/Amer gene family during mouse embryonic development. *Developmental Dynamics*, *239*(6), 1867–1878. https://doi.org/10.1002/dvdy.22313 (cited on page 197)

Concia, L., Veluchamy, A., Ramirez-Prado, J. S., Martin-Ramirez, A., Huang, Y., Perez, M., Domenichini, S., Rodriguez Granados, N. Y., Kim, S., Blein, T., Duncan, S., Pichot, C., Manza-Mianza, D., Juery, C., Paux, E., Moore, G., Hirt, H., Bergounioux, C., Crespi, M., . . . Benhamed, M. (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biology*, *21*(1), 104. https://doi.org/10.1186/s13059-020-01998-1 (cited on page 193)

Connors, J., Krzywinski, M., Schein, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos : An information aesthetic for comparative genomics. *Genome Research*, *19*(604), 1639–1645. https://doi.org/10.1101/gr.092759.109.19 (cited on page 139)

Coop, G., & Myers, S. R. (2007). Live hot, die young: Transmission distortion in recombination hotspots. *PLoS Genetics*, *3*(3), 0377–0386. https://doi.org/10.1371/journal.pgen.0030035 (cited on page 21)

Cosgrove, M. S., & Wolberger, C. (2005). How does the histone code work? *Biochemistry and Cell Biology*, *83*(4), 468–476. https://doi.org/10.1139/o05-137 (cited on page 5)

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer. (Cited on page 22).

Crane, E., Bian, Q., McCord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., Uzawa, S., Dekker, J., & Meyer, B. J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*, *523*(7559), 240–244. https://doi.org/10.1038/nature14450 (cited on page 8)

Cree, L. H., Balhorn, R., & Brewer, L. R. (2011). Single molecule studies of DNA-protamine interactions. *Protein and peptide letters*, *18*(8), 802–810. https://doi.org/10.2174/092986611795713943 (cited on page 197)

Cremer, M., Grasser, F., Lanctôt, C., Müller, S., Neusser, M., Zinner, R., Solovei, I., & Cremer, T. (2008). Multicolor 3D Fluorescence In Situ Hybridization for Imaging Interphase Chromosomes. In R. Hancock (Ed.), *The nucleus volume 1: Nuclei and subnuclear components* (pp. 205–329). Québec, Humana Press. https://doi.org/10.1007/978-1-59745-406-3. (Cited on page 4)

Cremer, T., Cremer, C., Baumann, H., Luedtke, E. K., Sperling, K., Teuber, V., & Zorn, C. (1982). Rabl's model of the interphase chromosome arrangement tested in Chinise hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Human Genetics*, *60*(1), 46–56. https://doi.org/10.1007/BF00281263 (cited on page 3)

Cremer, T., & Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, 2(4), 292–301. https://doi.org/10.1038/35066075 (cited on pages 3, 4)

Cremer, T., & Cremer, C. (2006). Rise, fall and resurrection of chromosome territories: A historical perspective Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and m. *European Journal of Histochemistry*, 50(4), 223–272. https://doi.org/10.4081/995 (cited on page 3)

Cremer, T., & Cremer, M. (2010). Chromosome territories. *Cold Spring Harbor Perspectives in Biology*, 2(February 2007), 1–23. https://doi.org/10.1038/445379a (cited on pages 3, 194)

Croft, J. A., Bridger, J. M., Boyle, S., Perry, P., Teague, P., & Bickmore, W. A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *The Journal of Cell Biology*, 145(6), 1119–1131. https://doi.org/10.1083/jcb.145.6.1119 (cited on page 4)

da Cruz, I., Rodríguez-Casuriaga, R., Santiñaque, F. F., Farías, J., Curti, G., Capoano, C. A., Folle, G. A., Benavente, R., Sotelo-Silveira, J. R., & Geisinger, A. (2016). Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. *BMC Genomics*, 17(1), 294–313. https://doi.org/10.1186/s12864-016-2618-1 (cited on pages 16, 195–198)

Dadoune, J.-P. (2003). Expression of mammalian spermatozoal nucleoproteins. *Microscopy Research and Technique*, 61(1), 56–75. https://doi.org/10.1002/jemt.10317 (cited on page 17)

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330 (cited on pages 51, 137)

Dapper, A. L., & Payseur, B. A. (2019). Molecular evolution of the meiotic recombination pathway in mammals. *Evolution*, 73(12), 2368–2389. https://doi.org/10.1111/evo.13850 (cited on pages 15, 18, 19)

Davidson, I. F., Goetz, D., Zaczek, M. P., Molodtsov, M. I., Huis in 't Veld, P. J., Weissmann, F., Litos, G., Cisneros, D. A., Ocampo-Hafalla, M., Ladurner, R., Uhlmann, F., Vaziri, A., & Peters, J.-M. (2016). Rapid movement and transcriptional re-localization of human cohesin on DNA. *The EMBO Journal*, 35(24), 2671–2685. https://doi.org/10.15252/embj.201695402 (cited on page 10)

Davies, B., Hatton, E., Altemose, N., Hussin, J. G., Pratto, F., Zhang, G., Hinch, A. G., Moralli, D., Biggs, D., Diaz, R., Preece, C., Li, R., Bitoun, E., Brick, K., Green, C. M., Camerini-Otero, R. D., Myers, S. R., Donnelly, P., Davies, A. B., . . . Donnelly, P. (2016). Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature*, 530(7589), 1–19. https://doi.org/10.1038/nature16931 (cited on pages 21, 203)

De Vries, M., Ramos, L., Housein, Z., & De Boer, P. (2012). Chromatin remodelling initiation during human spermiogenesis. *Biology Open*, 1(5), 446–457. https://doi.org/10.1242/bio.2012844 (cited on page 197)

Deakin, J. E., Potter, S., O'Neill, R., Ruiz-Herrera, A., Cioffi, M. B., Eldridge, M. D., Fukui, K., Marshall Graves, J. A., Griffin, D., Grutzner, F., Kratochvíl, L., Miura, I., Rovatsos, M., Srikulnath, K., Wapstra, E., & Ezaz, T. (2019). Chromosomics: Bridging the Gap between Genomes and Chromosomes. *Genes*, 10(8), 627. https://doi.org/10.3390/genes10080627 (cited on pages 119, 134, 200)

Degl'Innocenti, A., Meloni, G., Mazzolai, B., & Ciofani, G. (2019). A purely bioinformatic pipeline for the prediction of mammalian odorant receptor gene enhancers. *BMC Bioinformatics*, 20(1), 474. https://doi.org/10.1186/s12859-019-3012-1 (cited on pages 131, 132)

Dekker, J., & Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS Letters*, 589(20PartA), 2877–2884. https://doi.org/10.1016/j.febslet.2015.08.044 (cited on page 195)

Dekker, J., Marti-renom, M. A., & Mirny, L. A. (2013). Exploring the three-dimensional organization of genomes : interpreting chromatin interaction data. *Nature Publishing Group*, 14(6), 390–403. https://doi.org/10.1038/nrg3454 (cited on pages 4, 5, 7, 193)

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, 295(5558), 1306–11. https://doi.org/10.1126/science.1067799 (cited on pages 4, 37)

Del Prete, S., Mikulski, P., Schubert, D., & Gaudin, V. (2015). One, Two, Three: Polycomb Proteins Hit All Dimensions of Gene Regulation. *Genes*, 6(3), 520–542. https://doi.org/10.3390/genes6030520 (cited on page 8)

Delaneau, O., Zazhytska, M., Borel, C., Giannuzzi, G., Rey, G., Howald, C., Kumar, S., Ongen, H., Popadin, K., Marbach, D., Ambrosini, G., Bielser, D., Hacker, D., Romano, L., Ribaux, P., Wiederkehr, M., Falconnet, E., Bucher, P., Bergmann, S., . . . Dermitzakis, E. T. (2019). Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science*, *364*(6439). https://doi.org/10.1126/science.aat8266 (cited on page 193)

Delaneau, O., Marchini, J., & Zagury, J.-F. (2012). A linear complexity phasing method for thousands of genomes. *Nature Methods*, *9*(2), 179–181. https://doi.org/10.1038/nmeth.1785 (cited on pages 51, 137)

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., & Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell*, *13*(5), 602–616. https://doi.org/10.1016/j.stem.2013.08.013 (cited on page 197)

de Rooij, D. G. (2017). The nature and dynamics of spermatogonial stem cells. *Development*, *144*(17), 3022–3030. https://doi.org/10.1242/dev.146571 (cited on pages 57, 60, 67, 194)

de Rooij, D. G., & Russell, L. (2000). All you wanted to know about spermatogonia but were afraid to ask. *Journal of Andrology*, *21*, 776–798. https://doi.org/10.1002/j.1939-4640.2000.tb03408.x (cited on page 14)

Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W.-L., Timmermann, B., Wittler, L., Vingron, M., Mundlos, S., & Ibrahim, D. M. (2019). Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nature Genetics*, *51*(8), 1263–1271. https://doi.org/10.1038/s41588-019-0466-z (cited on page 193)

de Wit, E., Vos, E. S., Holwerda, S. J., Valdes-Quezada, C., Verstegen, M. J., Teunissen, H., Splinter, E., Wijchers, P. J., Krijger, P. H., & de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Molecular Cell*, *60*(4), 676–684. https://doi.org/10.1016/j.molcel.2015.09.023 (cited on page 8)

Dixon, J. R., Gorkin, D. U., & Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. *Molecular Cell*, *62*(5), 668–680. https://doi.org/10.1016/j.molcel.2016.05.018 (cited on page 8)

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082 (cited on pages 4, 8, 9, 119)

Dobigny, G., Britton-Davidian, J., & Robinson, T. J. (2017). Chromosomal polymorphism in mammals: an evolutionary perspective. *Biological Reviews*, *92*(1), 1–21. https://doi.org/10.1111/brv.12213 (cited on page 201)

Dobzhansky, T. (1936). Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids. *Genetics*, *21*(2), 113–135 (cited on page 22).

Dong, P., Tu, X., Chu, P.-Y., Lu, P., Zhu, N., Grierson, D., Du, B., Li, P., & Zhong, S. (2017). 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Molecular Plant*, *10*(12), 1497–1509. https://doi.org/10.1016/j.molp.2017.11.005 (cited on page 6)

Dorsett, D., & Merkenschlager, M. (2013). Cohesin at active genes: a unifying theme for cohesin and gene expression from model organisms to humans. *Current Opinion in Cell Biology*, *25*(3), 327–333. https://doi.org/10.1016/j.ceb.2013.02.003 (cited on page 196)

Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., Weintraub, A. S., Schujiers, J., Lee, T. I., Zhao, K., & Young, R. A. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, *159*(2), 374–387. https://doi.org/10.1016/j.cell.2014.09.030 (cited on pages 8, 10)

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., Shendure, J., Fields, S., Blau, C. A., & Noble, W. S. (2010). A three-dimensional model of the yeast genome. *Nature*, *465*(7296), 363–367. https://doi.org/10.1038/nature08973 (cited on pages 6, 8)

Dumas, D., Catalan, J., & Britton-davidian, J. (2015). Reduced recombination patterns in Robertsonian hybrids between chromosomal races of the house mouse: chiasma analyses. *Heredity*, *114*(1), 56–64. https://doi.org/10.1038/hdy.2014.69 (cited on page 200)

Dumas, D., & Britton-Davidian, J. (2002). Chromosomal Rearrangements and Evolution of Recombination: Comparison of Chiasma Distribution Patterns in Standard and Robertsonian Populations of the House Mouse. *Genetics*, *162*(3), 1355–1366 (cited on pages 22, 23, 200).

Dunleavy, J. E. M., O'Connor, A. E., & O'Bryan, M. K. (2019). An optimised STAPUT method for the purification of mouse spermatocyte and spermatid populations. *Molecular Human Reproduction*, 25(11), 675–683. https://doi.org/10.1093/molehr/gaz056 (cited on page 57)

Dunn, K. L., Zhao, H., & Davie, J. R. (2003). The insulator binding protein CTCF associates with the nuclear matrix. *Experimental Cell Research*, 288(1), 218–223. https://doi.org/10.1016/s0014-4827(03)00185-x (cited on page 9)

Dzur-Gejdosova, M., Simecek, P., Gregorova, S., Bhattacharyya, T., & Forejt, J. (2012). Dissecting the genetic architecture of F1 hybrid sterility in house mice. *Evolution*, 66(11), 3321–35. https://doi.org/10.1111/j.1558-5646.2012.01684.x (cited on page 22)

Ea, V., Baudement, M.-O., Lesne, A., & Forné, T. (2015). Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. *Genes*, 6(3), 734–50. https://doi.org/10.3390/genes6030734 (cited on page 5)

Eagen, K. P. (2018). Principles of Chromosome Architecture Revealed by Hi-C. *Trends in Biochemical Sciences*, 43(6), 469–478. https://doi.org/10.1016/j.tibs.2018.03.006 (cited on pages 4, 5, 7, 8)

Eram, M. S., Bustos, S. P., Lima-Fernandes, E., Siarheyeva, A., Senisterra, G., Hajian, T., Chau, I., Duan, S., Wu, H., Dombrovski, L., Schapira, M., Arrowsmith, C. H., & Vedadi, M. (2014). Trimethylation of histone H3 lysine 36 by human methyltransferase PRDM9 protein. *The Journal of Biological Chemistry*, 289(17), 12177–12188. https://doi.org/10.1074/jbc.M113.523183 (cited on page 20)

Eres, I. E., Luo, K., Hsiao, C. J., Blake, L. E., & Gilad, Y. (2019). Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLoS Genetics*, 15(7), e1008278–e1008278. https://doi.org/10.1371/journal.pgen.1008278 (cited on page 5)

Ernst, C., Eling, N., Martinez-Jimenez, C. P., Marioni, J. C., & Odom, D. T. (2019). Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. *Nature Communications*, 10(1), 1251. https://doi.org/10.1038/s41467-019-09182-1 (cited on pages 16, 195, 198)

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x (cited on pages 51, 136)

Fallahi, M., Getun, I. V., Wu, Z. K., & Bois, P. R. J. (2010). A global expression switch marks pachytene initiation during mouse male meiosis. *Genes*, 1, 469–483. https://doi.org/10.3390/genes1030469 (cited on pages 57, 193)

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving Inversions. *Trends in Ecology and Evolution*, 34(3), 239–248. https://doi.org/10.1016/j.tree.2018.12.005 (cited on page 22)

Faria, R., & Navarro, A. (2010). Chromosomal speciation revisited: Rearranging theory with pieces of evidence. *Trends in Ecology and Evolution*, 25(11), 660–669. https://doi.org/10.1016/j.tree.2010.07.008 (cited on pages 22, 23, 133, 201, 202)

Farré, M., Micheletti, D., & Ruiz-Herrera, A. (2013). Recombination rates and genomic shuffling in human and vhimpanzee—A new twist in the chromosomal speciation theory. *Molecular Biology and Evolution*, 30(4), 853–864. https://doi.org/10.1093/molbev/mss272 (cited on pages 22, 23, 120, 133, 201)

Farré, M., Robinson, T. J., & Ruiz-Herrera, A. (2015). An Integrative Breakage Model of genome architecture, reshuffling and evolution. *BioEssays*, 37(5), 479–488. https://doi.org/10.1002/bies.201400174 (cited on pages 22, 119, 134, 200)

Fedyk, S., Pavlova, S. V., Chtnicki, W., & Searle, J. B. (2019). Chromosomal Hybrid Zones. In J. Zima, J. B. Searle, & P. D. Polly (Eds.), *Shrews, chromosomes and speciation* (pp. 271–312). Cambridge, Cambridge University Press. https://doi.org/DOI:10.1017/9780511895531.009. (Cited on page 134)

Feng, J., Liu, T., Qin, B., Zhang, Y., & Liu, X. S. (2012). Identifying ChIP-seq enrichment using MACS. *Nature Protocols*, 7(9), 1728–1740. https://doi.org/10.1038/nprot.2012.101 (cited on page 52)

Feng, S., Cokus, S. J., Schubert, V., Zhai, J., Pellegrini, M., & Jacobsen, S. E. (2014). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. *Molecular Cell*, 55(5), 694–707. https://doi.org/10.1016/j.molcel.2014.07.008 (cited on pages 6, 8)

Finch, J. T., & Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, 73(6), 1897–1901. https://doi.org/10.1073/pnas.73.6.1897 (cited on page 4)

Flachs, P., Mihola, O., Šimeček, P., Gregorová, S., Schimenti, J. C., Matsui, Y., Baudat, F., de Massy, B., Piálek, J., Forejt, J., & Trachtulec, Z. (2012). Interallelic and Intergenic Incompatibilities of the Prdm9 (Hst1) Gene in Mouse Hybrid Sterility. *PLoS Genetics*, *8*(11), e1003044. https://doi.org/10.1371/journal.pgen.1003044 (cited on page 22)

Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., Suva, M. L., & Bernstein, B. E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, *529*(7584), 110–114. https://doi.org/10.1038/nature16490 (cited on page 8)

Flegel, C., Vogel, F., Hofreuter, A., Schreiner, B. S. P., Osthold, S., Veitinger, S., Becker, C., Brockmeyer, N. H., Muschol, M., Wennemuth, G., Altmüller, J., Hatt, H., & Gisselmann, G. (2016). Characterization of the olfactory receptors expressed in human spermatozoa. *Frontiers in Molecular Biosciences*, *2*, 73. https://doi.org/10.3389/fmolb.2015.00073 (cited on page 134)

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., Razin, S. V., Mirny, L. A., Tachibana-Konwalski, K. K., Ulyanov, S. V., Abdennur, N., Razin, S. V., Mirny, L. A., & Tachibana-Konwalski, K. K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*, *544*, 110–114. https://doi.org/10.1038/nature21711 (cited on page 197)

Forejt, J. (1996). Hybrid sterility in the mouse. *Trends in Genetics*, *12*(10), 412–417. https://doi.org/10.1016/0168-9525(96)10040-8 (cited on pages 19, 22)

Forejt, J., & Iványi, P. (1974). Genetic studies on male sterility of hybrids between laboratory and wild mice (Mus musculus L.) (2009/04/14). *Genetical Research*, *24*(2), 189–206. https://doi.org/DOI:10.1017/S0016672300015214 (cited on page 19)

Forejt, J., Vincek, V., Klein, J., Lehrach, H., & Loudova-Mickova, M. (1991). Genetic mapping of the t-complex region on mouse chromosome 17 including the Hybrid sterility-1 gene. *Mammalian Genome*, *1*(2), 84–91. https://doi.org/10.1007/bf02443783 (cited on page 19)

Förster, D. W., Günduz, İ., Nunes, A. C., Gabriel, S., Ramalhinho, M. G., Mathias, M. L., Britton-Davinian, J., & Searle, J. B. (2009). Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology*, *18*(21), 4477–4494. https://doi.org/10.1111/j.1365-294X.2009.04344.x (cited on pages 24, 25, 202)

Förster, D. W., Mathias, M. L., Britton-Davidian, J., & Searle, J. B. (2013). Origin of the chromosomal radiation of Madeiran house mice: a microsatellite analysis of metacentric chromosomes. *Heredity*, *110*(4), 380–388. https://doi.org/10.1038/hdy.2012.107 (cited on pages 23, 24, 203)

Franchini, P., Colangelo, P., Solano, E., Capanna, E., Verheyen, E., & Castiglia, R. (2010). Reduced gene flow at pericentromeric loci in a hybrid zone involving chromosomal races of the house mouse Mus musculus domesticus. *Evolution*, *64*(7). https://doi.org/10.1111/j.1558-5646.2010.00964.x (cited on page 23)

Franchini, P., Kautt, A. F., Nater, A., Antonini, G., Castiglia, R., Meyer, A., & Solano, E. (2020). Reconstructing the evolutionary history of chromosomal races on islands: a genome-wide analysis of natural house mouse populations (R. Rogers, Ed.). *Molecular Biology and Evolution*, 1–27. https://doi.org/10.1093/molbev/msaa118 (cited on page 203)

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509 (cited on page 137)

Franke, M., & Gómez-Skarmeta, J. L. (2018). An evolutionary perspective of regulatory landscape dynamics in development and disease. *Current Opinion in Cell Biology*, *55*, 24–29. https://doi.org/10.1016/j.ceb.2018.06.009 (cited on page 119)

Franklin, R. E., & Gosling, R. G. (1953). Molecular Configuration in Sodium Thymonucleate. *Nature*, *171*(4356), 740–741. https://doi.org/10.1038/171740a0 (cited on page 3)

Fraune, J., Schramm, S., Alsheimer, M., & Benavente, R. (2012). The mammalian synaptonemal complex: Protein components, assembly and role in meiotic recombination. *Experimental Cell Research*, *318*(12), 1340–1346. https://doi.org/10.1016/j.yexcr.2012.02.018 (cited on pages 15, 17)

Fritz, A. J., Sehgal, N., Pliss, A., Xu, J., & Berezney, R. (2019). Chromosome territories and the global regulation of the genome. *Genes, Chromosomes and Cancer*, *58*(7), 407–426. https://doi.org/10.1002/gcc.22732 (cited on page 194)

Froenicke, L., Anderson, L. K., Wienberg, J., & Ashley, T. (2002). Male mouse recombination maps for each autosome identified by chromosome painting. *The American Journal of Human Genetics*, *71*(6), 1353–1368. https://doi.org/10.1086/344714 (cited on page 135)

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, *15*(9), 2038–2049. https://doi.org/10.1016/j.celrep.2016.04.085 (cited on pages 10, 193, 195)

Fujihara, Y., Oji, A., Kojima-Kita, K., Larasati, T., & Ikawa, M. (2018). Co-expression of sperm membrane proteins CMTM2A and CMTM2B is essential for ADAM3 localization and male fertility in mice. *Journal of Cell Science*, *131*(19), jcs221481. https://doi.org/10.1242/jcs.221481 (cited on page 194)

Furey, T. S. (2012). ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nature Reviews Genetics*, *13*(12), 840–852. https://doi.org/10.1038/nrg3306 (cited on page 45)

Galazka, J. M., Klocko, A. D., Uesaka, M., Honda, S., Selker, E. U., & Freitag, M. (2016). Neurospora chromosomes are organized by blocks of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. *Genome Research*, *26*(8), 1069–1080. https://doi.org/10.1101/gr.203182.115 (cited on pages 6, 8)

Garagna, S., Zuccotti, M., Thornhill, A., Fernandez-Donoso, R., Berrios, S., Capanna, E., & Redi, C. A. (2001). Alteration of nuclear architecture in male germ cells of chromosomally derived subfertile mice. *Journal of Cell Science*, *114*(Pt 24), 4429–34 (cited on pages 23, 199).

Garagna, S., Broccoli, D., Redi, C. A., Searle, J. B., Cooke, H. J., & Capanna, E. (1995). Robertsonian metacentrics of the house mouse lose telomeric sequences but retain some minor satellite DNA in the pericentromeric area. *Chromosoma*, *103*(10), 685–692. https://doi.org/10.1007/BF00344229 (cited on page 23)

Garagna, S., Page, J., Fernandez-Donoso, R., Zuccotti, M., & Searle, J. B. (2014). The robertsonian phenomenon in the house mouse: mutation, meiosis and speciation. *Chromosoma*, *123*(6), 529–544. https://doi.org/10.1007/s00412-014-0477-6 (cited on pages 23, 25, 201)

Gassei, K., Ehmcke, J., Dhir, R., & Schlatt, S. (2010). Magnetic activated cell sorting allows isolation of spermatogonia from adult primate testes and reveals distinct GFRa1-positive subpopulations in men. *Journal of Medical Primatology*, *39*(2), 83–91. https://doi.org/10.1111/j.1600-0684.2009.00397.x (cited on pages 14, 57)

Gassei, K., Ehmcke, J., & Stefan, S. (2009). Efficient enrichment of undifferentiated GFR alpha 1+ spermatogonia from immature rat testis by magnetic activated cell sorting. *Cell and Tissue Research*, *337*(1), 177–183. https://doi.org/10.1007/s00441-009-0799-5 (cited on pages 57, 193)

Gaysinskaya, V., & Bortvin, A. (2015). Flow cytometry of murine spermatocytes. *Current Protocols in Cytometry*, *2015*(April), 1–7. https://doi.org/10.1002/0471142956.cy0744s72 (cited on pages 58, 66, 193)

Gaysinskaya, V., Soh, I. Y., van der Heijden, G. W., & Bortvin, A. (2014). Optimized flow cytometry isolation of murine spermatocytes. *Cytometry Part A*, *85*(6), 556–565. https://doi.org/10.1002/cyto.a.22463 (cited on page 57)

Gel, B., & Serra, E. (2017). KaryoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*, *33*(19), 3088–3090. https://doi.org/10.1093/bioinformatics/btx346 (cited on page 52)

George, P., Kinney, N. A., Liang, J., Onufriev, A. V., & Sharakhov, I. V. (2020). Three-dimensional Organization of Polytene Chromosomes in Somatic and Germline Tissues of Malaria Mosquitoes. *Cells*, *9*(2). https://doi.org/10.3390/cells9020339 (cited on page 193)

Getun, I. V., Torres, B., & Bois, P. R. (2011). Flow Cytometry Purification of Mouse Meiotic Cells. *Journal of Visualized Experiments*, (50), 4–6. https://doi.org/10.3791/2602 (cited on pages 57, 58, 66, 193)

Gibcus, J. H., & Dekker, J. (2013). The Hierarchy of the 3D Genome. *Molecular Cell*, *49*(5), 773–782. https://doi.org/10.1016/J.MOLCEL.2013.02.011 (cited on pages 4, 5)

Gibcus, J. H., Samejima, K., Goloborodko, A., Samejima, I., Naumova, N., Nuebler, J., Kanemaki, M. T., Xie, L., Paulson, J. R., Earnshaw, W. C., Mirny, L. A., & Dekker, J. (2018). A pathway for mitotic chromosome formation. *Science*, *359*, 652. https://doi.org/10.1126/science.aao6135 (cited on pages 11, 194)

Gilbert, S. F. (2005). Mechanisms for the environmental regulation of gene expression: ecological aspects of animal development. *Journal of Biosciences*, *30*(1), 65–74. https://doi.org/10.1007/bf02705151 (cited on page 7)

Gilles, A. M., Imhoff, J. M., & Keil, B. (1979). alpha-Clostripain. Chemical characterization, activity, and thiol content of the highly active form of clostripain. *Journal of Biological Chemistry*, *254*(5), 1462–1468 (cited on page 58).

Giménez, M. D., Panithanarak, T., Hauffe, H. C., & Searle, J. B. (2016). Empirical demonstration of hybrid chromosomal races in house mice. *Evolution*, *70*(7), 1651–1658. https://doi.org/10.1111/evo.12970 (cited on page 203)

Giménez, M. D., White, T. A., Hauffe, H. C., Panithanarak, T., & Searle, J. B. (2013). Understanding the basis of diminished gene flow between hybridizing chromosomal races of the house mouse. *Evolution*, *67*(5). https://doi.org/10.1111/evo.12054 (cited on page 202)

Ginsburg, M., Snow, M. H., & McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development*, *110*(2), 521–528 (cited on page 12).

Go, V. L. W., Vernon, R. G., & Fritz, I. B. (1971). Studies on Spermatogenesis in Rats. I. Application of the Sedimentation Velocity Technique to an Investigation of Spermatogenesis. *Canadian Journal of Biochemistry*, *49*(7), 753–760. https://doi.org/10.1139/o71-106 (cited on page 57)

Gómez-H, L., Felipe-Medina, N., Sánchez-Martín, M., Davies, O. R., Ramos, I., García-Tuñón, I., de Rooij, D. G., Dereli, I., Tóth, A., Barbero, J. L., Benavente, R., Llano, E., & Pendas, A. M. (2016). C14ORF39/SIX6OS1 is a constituent of the synaptonemal complex and is essential for mouse fertility. *Nature Communications*, *7*(May), 13298. https://doi.org/10.1038/ncomms13298 (cited on page 15)

Gottesfeld, J. M., & Forbes, D. J. (1997). Mitotic repression of the transcriptional machinery. *Trends in Biochemical Sciences*, *22*(6), 197–202. https://doi.org/10.1016/s0968-0004(97)01045-1 (cited on page 195)

Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x (cited on pages 51, 137)

Govin, J., Caron, C., Lestrat, C., Rousseaux, S., & Khochbin, S. (2004). The role of histones in chromatin remodelling during mammalian spermiogenesis. *European journal of biochemistry*, *271*(17), 3459–3469. https://doi.org/10.1111/j.1432-1033.2004.04266.x (cited on page 16)

Grabske, R. J., Lake, S., Gledhill, B. L., & Meistrich, M. L. (1975). Centrifugal elutriation: Separation of spermatogenic cells on the basis of sedimentation velocity. *Journal of Cellular Physiology*, *86*(1), 177–189. https://doi.org/10.1002/jcp.1040860119 (cited on page 57)

Grasser, F., Neusser, M., Fiegler, H., Thormeyer, T., Cremer, M., Carter, N. P., Cremer, T., & Müller, S. (2008). Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. *Journal of Cell Science*, *121*(11), 1876–1886. https://doi.org/10.1242/jcs.026989 (cited on page 4)

Gray, S., & Cohen, P. E. (2016). Control of Meiotic Crossovers: From Double-Stand Break Formation to Designation. *Annual Review of Genetics*, *50*(1), 120215–035111. https://doi.org/10.1146/annurev-genet-120215-035111 (cited on page 19)

Gregorova, S., & Forejt, J. (2000). PWD/Ph and PWK/Ph inbred mouse strains of Mus m. musculus subspecies–a valuable resource of phenotypic variations and genomic polymorphisms. *Folia Biologica*, *46*(1), 31–41 (cited on page 19).

Gregorova, S., Gergelits, V., Chvatalova, I., Bhattacharyya, T., Valiskova, B., Fotopulosova, V., Jansa, P., Wiatrowska, D., & Forejt, J. (2018). Modulation of Prdm9-controlled meiotic chromosome asynapsis overrides hybrid sterility in mice. *eLife*, *7*. https://doi.org/10.7554/eLife.34282 (cited on page 22)

Grob, S., Schmid, M. W., & Grossniklaus, U. (2014). Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. *Molecular Cell*, *55*(5), 678–693. https://doi.org/10.1016/j.molcel.2014.07.009 (cited on page 6)

Groeneveld, L. F., Atencia, R., Garriga, R. M., & Vigilant, L. (2012). High diversity at PRDM9 in chimpanzees and bonobos. *PloS One*, *7*(7), e39064. https://doi.org/10.1371/journal.pone.0039064 (cited on page 21)

Grubert, F., Zaugg, J. B., Kasowski, M., Ursu, O., Spacek, D. V., Martin, A. R., Greenside, P., Srivas, R., Phanstiel, D. H., Pekowska, A., Heidari, N., Euskirchen, G., Huber, W., Pritchard, J. K., Bustamante, C. D., Steinmetz, L. M., Kundaje, A., & Snyder, M. (2015). Genetic Control of Chromatin States

in Humans Involves Local and Distal Chromosomal Interactions. *Cell*, *162*(5), 1051–1065. https://doi.org/10.1016/j.cell.2015.07.048 (cited on page 193)

Gruhn, J. R., Rubio, C., Broman, K. W., Hunt, P. A., & Hassold, T. (2013). Cytological Studies of Human Meiosis: Sex-Specific Differences in Recombination Originate at, or Prior to, Establishment of Double-Strand Breaks. *PLOS ONE*, *8*(12), e85075 (cited on page 17).

Guerrero, R. F., & Kirkpatrick, M. (2014). Local adaptation and the evolution of chromosome fusions. *Evolution*, *68*(10), 2747–2756. https://doi.org/10.1111/evo.12481 (cited on page 22)

Gündüz, I., López-Fuster, M. J., Ventura, J., & Searle, J. B. (2001a). Clinal analysis of a chromosomal hybrid zone in the house mouse. *Genetical Research*, *77*(1), 41–51. https://doi.org/10.1017/s0016672300004808 (cited on pages 23, 24, 202, 203)

Gündüz, I., Auffray, J.-C., Britton-Davidian, J., Ganem, G., Ramalhinho, M. G., Mathias, M. L., & Searle, J. B. (2001b). Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology*, *10*, 2023–2029 (cited on pages 198, 203).

Gündüz, İ., Pollock, C. L., Giménez, M. D., Förster, D. W., White, T. A., Sans-Fuentes, M. A., Hauffe, H. C., Ventura, J., López-Fuster, M. J., & Searle, J. B. (2010). Staggered Chromosomal Hybrid Zones in the House Mouse: Relevance to Reticulate Evolution and Speciation. *Genes*, *1*(2), 193–209. https://doi.org/10.3390/genes1020193 (cited on page 23)

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M. Q., Ren, B., Krainer, A. R., Maniatis, T., & Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, *162*(4), 900–910. https://doi.org/10.1016/j.cell.2015.07.038 (cited on pages 9, 10)

Haaf, T., & Ward, D. C. (1995). Higher Order Nuclear Structure in Mammalian Sperm Revealed by in Situ Hybridization and Extended Chromatin Fibers. *Experimental Cell Research*, *219*(2), 604–611. https://doi.org/10.1006/excr.1995.1270 (cited on page 197)

Habermann, F. A., Cremer, M., Walter, J., Kreth, G., von Hase, J., Bauer, K., Wienberg, J., Cremer, C., Cremer, T., & Solovei, I. (2001). Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Research*, *9*(7), 569–584. https://doi.org/10.1023/a:1012447318535 (cited on page 4)

Hahne, F., & Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. In E. Mathé & S. Davis (Eds.), *Statistical genomics* (pp. 335–351). New York, Humana Press. https://doi.org/https://doi.org/10.1007/978-1-4939-3578-9{_}16. (Cited on page 52)

Hale, D. W. (1986). Heterosynapsis and suppression of chiasmata within heterozygous pericentric inversions of the Sitka deer mouse. *Chromosoma*, *94*(6), 425–32 (cited on page 22).

Hall, T. A. (1999). BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. *Nucleic Acids Symposium Series*, *41*, 95–98. https://doi.org/10.14601/ (cited on page 50)

Hammoud, S. S., Low, D. H. P., Yi, C., Carrell, D. T., Guccione, E., & Cairns, B. R. (2014). Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell*, *15*(2), 239–253. https://doi.org/10.1016/j.stem.2014.04.006 (cited on pages 14, 16, 17, 60, 196, 197)

Han, J., Zhang, Z., & Wang, K. (2018). 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. *Molecular Cytogenetics*, *11*, 21. https://doi.org/10.1186/s13039-018-0368-2 (cited on page 5)

Handel, M. A., & Schimenti, J. C. (2010). Genetics of mammalian meiosis : regulation , dynamics and impact on fertility. *Nature Reviews Genetics*, *11*(2), 124–136. https://doi.org/10.1038/nrg2723 (cited on pages 15, 194)

Hansen, A. S., Cattoglio, C., Darzacq, X., & Tjian, R. (2018). Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus*, *9*(1), 20–32. https://doi.org/10.1080/19491034.2017.1389365 (cited on pages 5, 8, 9)

Hao, S.-L., Ni, F.-D., & Yang, W.-X. (2019). The dynamics and regulation of chromatin remodeling during spermiogenesis. *Gene*, *706*(May), 201–210. https://doi.org/10.1016/j.gene.2019.05.027 (cited on page 196)

Hara, K., Nakagawa, T., Enomoto, H., Suzuki, M., Yamamoto, M., Simons, B. D., & Yoshida, S. (2014). Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. *Cell Stem Cell*, *14*(5), 658–672. https://doi.org/10.1016/j.stem.2014.01.019 (cited on page 14)

Hassold, T., Sherman, S., & Hunt, P. (2000). Counting cross-overs: characterizing meiotic recombination in mammals. *Human Molecular Genetics*, *9*(16), 2409–19 (cited on page 202).

Hauffe, H. C., Gimenez, M. D., Vega, R., White, T. A., & Searle, J. B. (2011). Properties of a hybrid zone between highly distinct chromosomal races of the house mouse (Mus musculus domesticus) in Northern Italy, and comparisons with other hybrid zones. *Cytogenetic and Genome Research*, *134*(3), 191–199. https://doi.org/10.1159/000327717 (cited on pages 23, 203)

Hauffe, H. C., & Searle, J. B. (1998). Chromosomal heterozygosity and fertility in house mice (Mus musculus domesticus) from Northern Italy. *Genetics*, *150*(3), 1143–1154 (cited on page 25).

Hauffe, H. C., Giménez, M. D., & Searle, J. B. (2012). Chromosomal hybrid zones in the house mouse. In J. Piálek, M. Macholán, P. Munclinger, & S. J. E. Baird (Eds.), *Evolution of the house mouse* (pp. 407–430). Cambridge, Cambridge University Press. https://doi.org/DOI:10.1017/CBO9781139044547.018. (Cited on pages 23, 203)

Hayashi, K., & Matsui, Y. (2006). Meisetz, a novel histone tri-methyltransferase, regulates meiosis-specific epigenesis. *Cell Cycle*, *5*(6), 615–620. https://doi.org/10.4161/cc.5.6.2572 (cited on page 19)

Hayashi, K., Yoshida, K., & Matsui, Y. (2005). A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature*, *438*(7066), 374–378. https://doi.org/10.1038/nature04112 (cited on page 20)

He, Z., Henricksen, L. A., Wold, M. S., & Ingles, C. J. (1995). RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature*, *374*(6522), 566–569. https://doi.org/10.1038/374566a0 (cited on page 18)

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., & Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(43), 17507–17512. https://doi.org/10.1073/pnas.1111941109 (cited on page 9)

Helsel, A. R., & Oatley, J. M. (2017). Transplantation as a Quantitative Assay to Study Mammalian Male Germline Stem Cells. *Methods in Molecular Biology*, *1463*, 155–172. https://doi.org/10.1007/978-1-4939-4017-2{_}12 (cited on page 14)

Helsel, A. R., Yang, Q.-E., Oatley, M. J., Lord, T., Sablitzky, F., & Oatley, J. M. (2017). ID4 levels dictate the stem cell state in mouse spermatogonia. *Development*, *144*(4), 624–634. https://doi.org/10.1242/dev.146928 (cited on pages 14, 57, 60)

Henderson, K. A., & Keeney, S. (2005). Synaptonemal complex formation: where does it start? *BioEssays*, *27*(10), 995–998. https://doi.org/10.1002/bies.20310 (cited on pages 15, 18)

Hernández-Hernández, A., Lilienthal, I., Fukuda, N., Galjart, N., & Höög, C. (2016). CTCF contributes in a critical way to spermatogenesis and male fertility. *Scientific Reports*, *6*, 28355. https://doi.org/10.1038/srep28355 (cited on page 197)

Herrán, Y., Gutiérrez-Caballero, C., Sánchez-Martín, M., Hernández, T., Viera, A., Barbero, J. L., de Álava, E., de Rooij, D. G., Suja, J. Á., Llano, E., & Pendás, A. M. (2011). The cohesin subunit RAD21L functions in meiotic synapsis and exhibits sexual dimorphism in fertility. *The EMBO Journal*, *30*(15), 3091–3105. https://doi.org/10.1038/emboj.2011.222 (cited on page 195)

Hnisz, D., Day, D. S., & Young, R. A. (2016a). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell*, *167*(5), 1188–1200. https://doi.org/10.1016/j.cell.2016.10.024 (cited on page 10)

Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A.-L., Bak, R. O., Li, C. H., Goldmann, J., Lajoie, B. R., Fan, Z. P., Sigova, A. A., Reddy, J., Borges-Rivera, D., Lee, T. I., Jaenisch, R., Porteus, M. H., Dekker, J., & Young, R. A. (2016b). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, *351*(6280), 1454 LP –1458. https://doi.org/10.1126/science.aad9024 (cited on page 198)

Horn, P. J., & Peterson, C. L. (2002). Chromatin Higher Order Folding–Wrapping up Transcription. *Science*, *297*(5588), 1824 LP –1827. https://doi.org/10.1126/science.1074200 (cited on page 4)

Hou, C., Li, L., Qin, Z. S., & Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Molecular Cell*, *48*(3), 471–484. https://doi.org/10.1016/j.molcel.2012.08.031 (cited on pages 4–6, 8)

Hoyer-Fender, S., Singh, P. B., & Motzkus, D. (2000). The murine heterochromatin protein M31 is associated with the chromocenter in round spermatids and Is a component of mature spermatozoa. *Experimental Cell Research*, *254*(1), 72–9. https://doi.org/10.1006/excr.1999.4729 (cited on pages 129, 197, 199)

Huckins, C. (1971). The spermatogonial stem cell population in adult rats. I. Their morphology, proliferation and maturation. *The Anatomical record*, *169*(3), 533–557. https://doi.org/10.1002/ar.1091690306 (cited on page 14)

Hud, N. V., Allen, M. J., Downing, K. H., Lee, J., & Balhorn, R. (1993). Identification of the elemental packing unit of DNA in mammalian sperm cells by atomic force microscopy. *Biochemical and Biophysical Research Communications*, *193*(3), 1347–1354. https://doi.org/10.1006/bbrc.1993.1773 (cited on page 196)

Imai, Y., Baudat, F., Taillepierre, M., Stanzione, M., Toth, A., & de Massy, B. (2017). The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. *Chromosoma*. https://doi.org/10.1007/s00412-017-0631-z (cited on page 20)

Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., Dekker, J., & Mirny, L. A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods*, *9*, 999 (cited on page 8).

Ishiguro, K.-i., Kim, J., Fujiyama-Nakamura, S., Kato, S., & Watanabe, Y. (2011). A new meiosis-specific cohesin complex implicated in the cohesin code for homologous pairing. *EMBO Reports*, *12*(3), 267–275. https://doi.org/10.1038/embor.2011.2 (cited on page 195)

Ishiguro, K.-i., Kim, J., Shibuya, H., Hernández-Hernández, A., Suzuki, A., Fukagawa, T., Shioi, G., Kiyonari, H., Li, X. C., Schimenti, J., Höög, C., & Watanabe, Y. (2014). Meiosis-specific cohesin mediates homolog recognition in mouse spermatocytes. *Genes & Development*, *28*(6), 594–607. https://doi.org/10.1101/gad.237313.113 (cited on page 195)

Ishiguro, K.-i., & Watanabe, Y. (2016). The cohesin REC 8 prevents illegitimate inter-sister synaptonemal complex assembly. *EMBO Reports*, *17*(6), 783–784. https://doi.org/10.15252/embr.201642544 (cited on page 15)

Jackson, D. A., Hassan, A. B., Errington, R. J., & Cook, P. R. (1993). Visualization of focal sites of transcription within human nuclei. *The EMBO Journal*, *12*(3), 1059–1065 (cited on page 4).

Jasin, M., & Rothstein, R. (2013). Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harbor Perspectives in Biology*, *5*(11). https://doi.org/10.1101/cshperspect.a012740 (cited on page 200)

Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., Weintraub, A. S., Hnisz, D., Pegoraro, G., Lee, T. I., Misteli, T., Jaenisch, R., & Young, R. A. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell*, *18*(2), 262–275. https://doi.org/10.1016/j.stem.2015.11.007 (cited on page 8)

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C.-A., Schmitt, A. D., Espinoza, C. A., & Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, *503*(7475), 290–294. https://doi.org/10.1038/nature12644 (cited on page 195)

Jung, Y. H., Sauria, M. E., Lyu, X., Cheema, M. S., Ausio, J., Taylor, J., & Corces, V. G. (2017). Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Reports*, *18*(6), 1366–1382. https://doi.org/10.1016/j.celrep.2017.01.034 (cited on page 197)

Kanke, M., Tahara, E., Huis In't Veld, P. J., & Nishiyama, T. (2016). Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *The EMBO journal*, *35*(24), 2686–2698. https://doi.org/10.15252/embj.201695756 (cited on page 10)

Kauppi, L., Barchi, M., Baudat, F., Romanienko, P. J., Keeney, S., & Jasin, M. (2011). Distinct properties of the XY pseudoautosomal region crucial for male meiosis. *Science*, *331*(6019), 916–920. https://doi.org/10.1126/science.1195774 (cited on page 19)

Kauppi, L., Barchi, M., Lange, J., Baudat, F., Jasin, M., & Keeney, S. (2013). Numerical constraints and feedback control of double-strand breaks in mouse meiosis. *Genes and Development*, *27*(8), 873–886. https://doi.org/10.1101/gad.213652.113 (cited on page 18)

Kaye, J. S., & McMaster-Kaye, R. (1974). Histones of spermatogenous cells in the house cricket. *Chromosoma*, *46*(4), 397–419. https://doi.org/10.1007/bf00331629 (cited on page 57)

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., & Yao, X. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Resource 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell*, *170*(2), 367–381. https://doi.org/10.1016/j.cell.2017.06.029 (cited on page 197)

Keeney, S., Giroux, C. N., & Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, *88*(3), 375–384. https://doi.org/10.1016/S0092-8674(00)81876-0 (cited on pages 15, 18)

Keeney, S., Lange, J., & Mohibullah, N. (2014). Self-Organization of Meiotic Recombination Initiation: General Principles and Molecular Pathways. *Annual Review of Genetics*, *48*(1), 187–214. https://doi.org/10.1146/annurev-genet-120213-092304 (cited on page 15)

Kim, J., Farré, M., Auvil, L., Capitanu, B., Larkin, D. M., Ma, J., & Lewin, H. A. (2017). Reconstruction and evolutionary history of eutherian chromosomes. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1702012114 (cited on page 196)

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V., & Ren, B. (2007). Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell*, *128*(6), 1231–1245. https://doi.org/10.1016/j.cell.2006.12.048 (cited on page 9)

King, M. (1993). *Species Evolution: The Role of Chromosome Change*. Cambridge University Press. (Cited on page 22).

Kleckner, N. (2006). Chiasma formation: Chromatin/axis interplay and the role(s) of the synaptonemal complex. *Chromosoma*, *115*(3), 175–194. https://doi.org/10.1007/s00412-006-0055-7 (cited on pages 17, 19)

Kleckner, N., Storlazzi, A., & Zickler, D. (2003). Coordinate variation in meiotic pachytene SC length and total crossover/chiasma frequency under conditions of constant DNA length. *Trends in Genetics*, *19*(11), 623–628. https://doi.org/10.1016/j.tig.2003.09.004 (cited on pages 17, 19, 201)

Koehler, D., Zakhartchenko, V., Froenicke, L., Stone, G., Stanyon, R., Wolf, E., Cremer, T., & Brero, A. (2009). Changes of higher order chromatin arrangements during major genome activation in bovine preimplantation embryos. *Experimental Cell Research*, *315*(12), 2053–2063. https://doi.org/10.1016/j.yexcr.2009.02.016 (cited on page 4)

Kojic, A., Cuadrado, A., Koninck, M. D., & Rodríguez-corsino, M. (2017). Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nature Structural & Molecular Biology*. https://doi.org/10.1038/s41594-018-0070-4 (cited on page 10)

Komai, Y., Tanaka, T., Tokuyama, Y., Yanai, H., Ohe, S., Omachi, T., Atsumi, N., Yoshida, N., Kumano, K., Hisha, H., Matsuda, T., & Ueno, H. (2014). Bmi1 expression in long-term germ stem cells. *Scientific Reports*, *4*, 6175. https://doi.org/10.1038/srep06175 (cited on pages 14, 57)

Kono, H., Tamura, M., Osada, N., Suzuki, H., Abe, K., Moriwaki, K., Ohta, K., & Shiroishi, T. (2014). Prdm9 polymorphism unveils mouse evolutionary tracks. *DNA Research*, *21*(3), 315–326. https://doi.org/10.1093/dnares/dst059 (cited on pages 19, 21, 51, 202)

Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpflin, R., Markowski, J., Ibrahim, D. M., Acuna-Hidalgo, R., Despang, A., Andrey, G., Wittler, L., Timmermann, B., Vingron, M., & Mundlos, S. (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nature Cell Biology*, *21*(3), 305–310. https://doi.org/10.1038/s41556-019-0273-x (cited on pages 119, 198)

Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T.-H. S., Parsi, K. M., Yang, L., Maehr, R., Mirny, L. A., Dekker, J., & Rando, O. J. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell*, 1–12. https://doi.org/10.1016/j.molcel.2020.03.003 (cited on pages 5, 10, 11, 195)

Krivega, I., & Dean, A. (2017). CTCF fences make good neighbours. *Nature Cell Biology*, *19*(8), 883–885. https://doi.org/10.1038/ncb3584 (cited on page 8)

Kumar, R., Ghyselinck, N., Ishiguro, K.-i., Watanabe, Y., Kouznetsova, A., Höög, C., Strong, E., Schimenti, J., Daniel, K., Toth, A., de Massy, B., Hoog, C., Strong, E., Schimenti, J., Daniel, K., Toth, A., & de Massy, B. (2015). MEI4 - a central player in the regulation of meiotic DNA double-strand break formation in the mouse. *Journal of Cell Science*, *128*(9), 1800–11. https://doi.org/10.1242/jcs.165464 (cited on page 18)

Kumar, Y., Sengupta, D., & Bickmore, W. A. (2020). Recent advances in the spatial organization of the mammalian genome. *Journal of Biosciences*, *45*(1), 18. https://doi.org/10.1007/s12038-019-9968-1 (cited on pages 5, 7)

Kuo, L. J., & Yang, L.-X. (2008). Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In vivo*, *22*(3), 305–309 (cited on page 18).

Lajoie, B. R., Dekker, J., & Kaplan, N. (2015). The Hitchhiker's guide to Hi-C analysis: Practical guidelines. *Methods*, *72*, 65–75. https://doi.org/10.1016/j.ymeth.2014.10.031 (cited on pages 5–8)

Lam, D. M., & Bruce, W. R. (1971). The biosynthesis of protamine during spermatogenesis of the mouse: extraction, partial characterization, and site of synthesis. *Journal of Cellular Physiology*, *78*(1), 13–24. https://doi.org/10.1002/jcp.1040780104 (cited on page 57)

Lam, D. M., Furrer, R., & Bruce, W. R. (1970). The separation, physical characterization, and differentiation kinetics of spermatogonial cells of the mouse. *Proceedings of the National Academy of Sciences of the United States of America*, *65*(1), 192–199. https://doi.org/10.1073/pnas.65.1.192 (cited on page 57)

Lam, I., & Keeney, S. (2015). Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harbor Perspectives in Biology*, *7*(1), a016634. https://doi.org/10.1101/cshperspect.a016634 (cited on pages 18, 123)

Latt, S. A., & Stetten, G. (1976). Spectral studies on 33258 Hoechst and related bisbenzimidazole dyes useful for fluorescent detection of deoxyribonucleic acid synthesis. *The Journal of Histochemistry and Cytochemistry*, *24*(1), 24–33. https://doi.org/10.1177/24.1.943439 (cited on page 57)

Laugsch, M., Bartusel, M., Rehimi, R., Alirzayeva, H., Karaolidou, A., Crispatzu, G., Zentis, P., Nikolic, M., Bleckwehl, T., Kolovos, P., van Ijcken, W. F., Šarić, T., Koehler, K., Frommolt, P., Lachlan, K., Baptista, J., & Rada-Iglesias, A. (2019). Modeling the pathological long-range regulatory effects of human structural variation with patient-specific hiPSCs. *Cell Stem Cell*, *24*(5), 736–752. https://doi.org/10.1016/j.stem.2019.03.004 (cited on page 119)

Lawrence, M., Gentleman, R., & Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*, *25*(14), 1841–1842. https://doi.org/10.1093/bioinformatics/btp328 (cited on page 52)

Lee, J., & Hirano, T. (2011). RAD21L, a novel cohesin subunit implicated in linking homologous chromosomes in mammalian meiosis. *The Journal of Cell Biology*, *192*(2), 263–276. https://doi.org/10.1083/jcb.201008005 (cited on pages 15, 17, 195)

Lee, J., Iwai, T., Yokota, T., & Yamashita, M. (2003). Temporally and spatially selective loss of Rec8 protein from meiotic chromosomes during mammalian meiosis. *Journal of Cell Science*, *116*(13), 2781 LP –2790. https://doi.org/10.1242/jcs.00495 (cited on page 195)

Leidescher, S., Nübler, J., Feodorova, Y., Hildebrand, E., Bultmann, S., Link, S., Thanisch, K., Dekker, J., Leonhardt, H., Mirny, L., & Solovei, I. (2020). Spatial organization of transcribed eukaryotic genes. *bioRxiv* (cited on pages 195, 196).

Lesch, B. J., Dokshin, G. a., Young, R. a., McCarrey, J. R., & Page, D. C. (2013). A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(40), 16061–6. https://doi.org/10.1073/pnas.1315204110 (cited on page 197)

Lesch, B. J., & Page, D. C. (2014). Poised chromatin in the mammalian germ line. *Development*, *141*(19), 3619–26. https://doi.org/10.1242/dev.113027 (cited on page 197)

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 (cited on page 52)

Lichter, P., Cremer, T., Borden, J., Manuelidis, L., & Ward, D. C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Human Genetics*, *80*(3), 224–234. https://doi.org/10.1007/bf01790090 (cited on page 3)

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369 (cited on pages 4, 5, 7, 8, 119, 193, 194)

Lima, A. C., Jung, M., Rusch, J., Usmani, A., Lopes, A. M., & Conrad, D. F. (2016). Multispecies Purification of Testicular Germ Cells. *Biology of Reproduction*, *95*(4), 85–85. https://doi.org/10.1095/biolreprod.116.140566 (cited on pages 57, 66, 67)

Liu, Y., Niu, M., Yao, C., Hai, Y., Yuan, Q., Liu, Y., Guo, Y., Li, Z., & He, Z. (2015). Fractionation of human spermatogenic cells using STA-PUT gravity sedimentation and their miRNA profiling. *Scientific Reports*, *5*, 8084. https://doi.org/10.1038/srep08084 (cited on page 57)

Llano, E., Herrán, Y., García-Tuñón, I., Gutiérrez-Caballero, C., de Álava, E., Barbero, J. L., Schimenti, J., de Rooij, D. G., Sánchez-Martín, M., & Pendás, A. M. (2012). Meiotic cohesin complexes are essential for the formation of the axial element in mice. *The Journal of Cell Biology*, *197*(7), 877–885. https://doi.org/10.1083/jcb.201201100 (cited on pages 17, 195)

Loir, M., & Lanneau, M. (1974). Separation of ram spermatids by sedimentation at unit gravity. *Experimental Cell Research*, *83*(2), 319–327. https://doi.org/https://doi.org/10.1016/0014-4827(74)90345-0 (cited on page 57)

Longhese, M. P., Bonetti, D., Guerini, I., Manfrini, N., & Clerici, M. (2009). DNA double-strand breaks in meiosis: checking their formation, processing and repair. *DNA repair*, *8*(9), 1127–1138. https://doi.org/10.1016/j.dnarep.2009.04.005 (cited on pages 15, 18)

Louie, A. J., & Dixon, G. H. (1972). Synthesis, acetylation, and phosphorylation of histone IV and its binding to DNA during spermatogenesis in trout. *Proceedings of the National Academy of Sciences of the United States of America*, *69*(7), 1975–1979. https://doi.org/10.1073/pnas.69.7.1975 (cited on page 57)

Lucchesi, J. C., & Suzuki, D. T. (1968). The interchromosomal control of recombination. *Annual Review of Genetics*, *2*(1), 53–86. https://doi.org/10.1146/annurev.ge.02.120168.000413 (cited on page 119)

Lukassen, S., Bosch, E., Ekici, A. B., & Winterpacht, A. (2018). Single-cell RNA sequencing of adult mouse testes. *Scientific Data*, *5*, 1–7. https://doi.org/10.1038/sdata.2018.192 (cited on page 67)

Luo, Z., Wang, X., Jiang, H., Wang, R., Chen, J., Chen, Y., Xu, Q., Cao, J., Gong, X., Wu, J., Yang, Y., Li, W., Han, C., Cheng, C. Y., Rosenfeld, M. G., Sun, F., & Song, X. (2020). Reorganized 3D Genome Structures Support Transcriptional Regulation in Mouse Spermatogenesis. *iScience*, *23*(4), 101034. https://doi.org/10.1016/j.isci.2020.101034 (cited on page 195)

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., . . . Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004 (cited on pages 193, 198)

Lynn, A., Ashley, T., & Hassold, T. (2004). Variation in human meiotic recombination. *Annual Review of Genomics and Human Genetics*, *5*(1), 317–349. https://doi.org/10.1146/annurev.genom.4.070802.110217 (cited on pages 133, 201)

Maeshima, K., Imai, R., Tamura, S., & Nozaki, T. (2014). Chromatin as dynamic 10-nm fibers. *Chromosoma*, *123*(3), 225–237. https://doi.org/10.1007/s00412-014-0460-2 (cited on page 4)

Malik, S.-B., Ramesh, M. A., Hulstrand, A. M., & Logsdon, J. M. J. (2007). Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. *Molecular Biology and Evolution*, *24*(12), 2827–2841. https://doi.org/10.1093/molbev/msm217 (cited on page 18)

Manterola, M., Page, J., Vasco, C., Berríos, S., Parra, M. T., Viera, A., Rufas, J. S., Zuccotti, M., Garagna, S., & Fernández-Donoso, R. (2009). A High Incidence of Meiotic Silencing of Unsynapsed Chromatin Is Not Associated with Substantial Pachytene Loss in Heterozygous Male Mice Carrying Multiple Simple Robertsonian Translocations (R. S. Hawley, Ed.). *PLoS Genetics*, *5*(8), e1000625. https://doi.org/10.1371/journal.pgen.1000625 (cited on page 25)

Manuelidis, L. (1985). Individual interphase chromosome domains revealed by in situ hybridization. *Human Genetics*, *71*(4), 288–293. https://doi.org/10.1007/bf00388453 (cited on page 3)

Marco-Sola, S., Sammeth, M., Guigó, R., & Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. *Nature Methods*, *9*, 1185–1188 (cited on page 138).

Margolin, G., Khil, P. P., Kim, J., Bellani, M. A., & Camerini-Otero, R. D. (2014). Integrated transcriptome analysis of mouse spermatogenesis. *BMC Genomics*, *15*(1), 1–19. https://doi.org/10.1186/1471-2164-15-39 (cited on pages 16, 198, 199)

Martínez-Vargas, J., Muñoz-Muñoz, F., Medarde, N., López-Fuster, M. J., & Ventura, J. (2014). Effect of chromosomal reorganizations on morphological covariation of the mouse mandible: insights

from a Robertsonian system of Mus musculus domesticus. *Frontiers in Zoology*, *11*(1), 51. https://doi.org/10.1186/s12983-014-0051-3 (cited on page 24)

Mather, K. (1938). Crossing-over. *Biological Reviews*, *13*(3), 252–292. https://doi.org/10.1111/j.1469-185X.1938.tb00516.x (cited on page 19)

Mathias, M. d. L., & Mira, A. (1992). On the origin and colonization of house mice in the Madeira Islands. *Biological Journal of the Linnean Society*, *46*(1-2), 13–24. https://doi.org/10.1111/j.1095-8312.1992.tb00846.x (cited on pages 24, 202)

Matveevsky, S., Kolomiets, O., Bogdanov, A., Alpeeva, E., & Bakloushinskaya, I. (2020). Meiotic Chromosome Contacts as a Plausible Prelude for Robertsonian Translocations. *Genes*, *11*(4), 386. https://doi.org/10.3390/genes11040386 (cited on pages 134, 199, 201)

Maya-Miles, D., Andújar, E., Pérez-Alegre, M., Murillo-Pineda, M., Barrientos-Moreno, M., Cabello-Lobato, M. J., Gómez-Marín, E., Morillo-Huesca, M., & Prado, F. (2019). Crosstalk between chromatin structure, cohesin activity and transcription. *Epigenetics and Chromatin*, *12*(1), 1–18. https://doi.org/10.1186/s13072-019-0293-6 (cited on pages 196, 198)

Mayer, R., Brero, A., von Hase, J., Schroeder, T., Cremer, T., & Dietzel, S. (2005). Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biology*, *6*, 44. https://doi.org/10.1186/1471-2121-6-44 (cited on page 4)

Mays-Hoopes, L. L., Bolen, J., Riggs, a. D., & Singer-Sam, J. (1995). Preparation of spermatogonia, spermatocytes, and round spermatids for analysis of gene expression using fluorescence-activated cell sorting. *Biology of Reproduction*, *53*(5), 1003–1011. https://doi.org/10.1095/biolreprod53.5.1003 (cited on page 66)

McCord, R. P., Kaplan, N., & Giorgetti, L. (2020). Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Molecular Cell*, *77*(4), 688–708. https://doi.org/10.1016/j.molcel.2019.12.021 (cited on pages 4, 193)

McLaren, A. (1984). Meiosis and differentiation of mouse germ cells. *Symposia of the Society for Experimental Biology*, *38*, 7–23 (cited on page 12).

McLaren, A. (2003). Primordial germ cells in the mouse. *Developmental Biology*, *262*(1), 1–15. https://doi.org/10.1016/S0012-1606(03)00214-8 (cited on page 12)

McNicoll, F., Stevense, M., & Jessberger, R. (2013). *Cohesin in Gametogenesis* (Vol. 102). (Cited on pages 10, 15).

McPherson, S. M., & Longo, F. J. (1993). Nicking of rat spermatid and spermatozoa DNA: possible involvement of DNA topoisomerase II. *Developmental Biology*, *158*(1), 122–130. https://doi.org/10.1006/dbio.1993.1173 (cited on page 17)

Medarde, N., López-Fuster, M. J., Muñoz-Muñoz, F., & Ventura, J. (2012). Spatio-temporal variation in the structure of a chromosomal polymorphism zone in the house mouse. *Heredity*, *109*(2), 78–89. https://doi.org/10.1038/hdy.2012.16 (cited on pages 23, 120, 135, 202, 203)

Medarde, N., Merico, V., López-Fuster, M. J., Zuccotti, M., Garagna, S., & Ventura, J. (2015). Impact of the number of Robertsonian chromosomes on germ cell death in wild male house mice. *Chromosome Research*, *23*(2), 159–169. https://doi.org/10.1007/s10577-014-9442-8 (cited on pages 23, 25, 133, 199, 201, 202)

Medarde, N., Muñoz-Muñoz, F., López-Fuster, M., & Ventura, J. (2013). Variational modularity at the cell level: insights from the sperm head of the house mouse. *BMC Evolutionary Biology*, *13*(1), 179. https://doi.org/10.1186/1471-2148-13-179 (cited on pages 25, 134, 199)

Meistrich, M. L., & Hess, R. A. (2013). Assessment of spermatogenesis through staging of seminiferous tubules. *Methods in Molecular Biology*, *927*, 299–307. https://doi.org/10.1007/978-1-62703-038-0{_}27 (cited on pages 13, 16)

Meistrich, M. (1972). Separation of mouse spermatogenic cells by velocity sedimentation. *Journal of Cellular Physiology*, *80*(2), 299–312. https://doi.org/10.1002/jcp.1040800218 (cited on pages 57, 67)

Meistrich, M., Bruce, W., & Clermont, Y. (1973). Cellular composition of fractions of mouse testis cells following velocity sedimentation separation. *Experimental Cell Research*, *79*(1), 213–227. https://doi.org/10.1016/0014-4827(73)90507-7 (cited on page 57)

Merico, V., Pigozzi, M., Esposito, A., Merani, M., & Garagna, S. (2003). Meiotic recombination and spermatogenic impairment in Mus musculus domesticus carrying multiple simple Robertsonian translocations.

Cytogenetic and Genome Research, *103*(3-4), 321–329. https://doi.org/10.1159/000076820 (cited on pages 122, 200)

Merico, V., Gimenez, M. D., Vasco, C., Zuccotti, M., Searle, J. B., Hauffe, H. C., & Garagna, S. (2013). Chromosomal speciation in mice: a cytogenetic analysis of recombination. *Chromosome Research*, *21*(5), 523–533. https://doi.org/10.1007/s10577-013-9377-5 (cited on pages 122, 136, 200)

Merkenschlager, M., & Nora, E. P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annual Review of Genomics and Human Genetics*, *17*(1), 17–43. https://doi.org/10.1146/annurev-genom-083115-022339 (cited on pages 8–11)

Meyer-Ficca, M., Muller-Navia, J., & Scherthan, H. (1998). Clustering of pericentromeres initiates in step 9 of spermiogenesis of the rat (Rattus norvegicus) and contributes to a well defined genome architecture in the sperm nucleus. *Journal of Cell Science, 111 (Pt 1*, 1363–1370 (cited on page 197).

Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C., & Forejt, J. (2009). A Mouse Speciation Gene Encodes a Meiotic Histone H3 Methyltransferase. *Science*, *323*(5912), 373–5. https://doi.org/10.1126/science.1163601 (cited on pages 19, 22)

Miller, D., Brinkworth, M., & Iles, D. (2010). Paternal DNA packaging in spermatozoa: more than the sum of its parts? DNA, histones, protamines and epigenetics. *Reproduction*, *139*(2), 287–301. https://doi.org/10.1530/REP-09-0281 (cited on pages 17, 196)

Miltenyi, S., Müller, W., Weichel, W., & Radbruch, A. (1990). High gradient magnetic cell separation with MACS. *Cytometry*, *11*(2), 231–238. https://doi.org/10.1002/cyto.990110203 (cited on page 57)

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H. D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., & Grewal, S. I. S. (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature*, *516*(7531), 432–435. https://doi.org/10.1038/nature13833 (cited on pages 6, 8)

Moens, P. B., Kolas, N. K., Tarsounas, M., Marcon, E., Cohen, P. E., & Spyropoulos, B. (2002). The time course and chromosomal localization of recombination-related proteins at meiosis in the mouse are compatible with models that can resolve the early DNA-DNA interactions without reciprocal recombination. *Journal of Cell Science*, *115*(8), 1611–1622 (cited on page 201).

Moens, P. B., Marcon, E., Shore, J. S., Kochakpour, N., & Spyropoulos, B. (2007). Initiation and resolution of interhomolog connections: Crossover and non-crossover sites along mouse synaptonemal complexes. *Journal of Cell Science*, *120*(6), 1017–1027. https://doi.org/10.1242/jcs.03394 (cited on page 19)

Molnar, M., & Kleckner, N. (2008). Examination of interchromosomal interactions in vegetatively growing diploid Schizosaccharomyces pombe cells by Cre/loxP site-specific recombination. *Genetics*, *178*(1), 99–112. https://doi.org/10.1534/genetics.107.082826 (cited on page 3)

Monahan, K., Horta, A., & Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature*, *565*(7740), 448–453. https://doi.org/10.1038/s41586-018-0845-0 (cited on pages 134, 200)

Monesi, V. (1962). Autoradiographic study of dna synthesis and the cell cycle in spermatogonia and spermatocytes of mouse testis using tritiated thymidine. *The Journal of Cell Biology*, *14*(1), 1–18. https://doi.org/10.1083/jcb.14.1.1 (cited on page 14)

Moraru, M., & Schalch, T. (2019). Chromatin fiber structural motifs as regulatory hubs of genome function? *Essays in Biochemistry*, *63*(1), 123–132. https://doi.org/10.1042/EBC20180065 (cited on page 4)

Morgan, A. P., Fu, C.-P., Kao, C.-Y., Welsh, C. E., Didion, J. P., Yadgary, L., Hyacinth, L., Ferris, M. T., Bell, T. A., Miller, D. R., Giusti-Rodriguez, P., Nonneman, R. J., Cook, K. D., Whitmire, J. K., Gralinski, L. E., Keller, M., Attie, A. D., Churchill, G. A., Petkov, P., . . . Pardo-Manuel de Villena, F. (2016a). The Mouse Universal Genotyping Array: From substrains to subspecies. *Genes|Genomes|Genetics*, *6*(2), 263–279. https://doi.org/10.1534/g3.115.022087 (cited on page 51)

Morgan, A. P., Holt, J. M., McMullan, R. C., Bell, T. A., Clayshulte, A. M.-F., Didion, J. P., Yadgary, L., Thybert, D., Odom, D. T., Flicek, P., McMillan, L., & de Villena, F. P.-M. (2016b). The Evolutionary Fates of a Large Segmental Duplication in Mouse. *Genetics*, *204*(1), 267–85. https://doi.org/10.1534/genetics.116.191007 (cited on page 136)

Mruk, D. D., & Cheng, C. Y. (2015). The mammalian blood-testis barrier: Its biology and regulation. *Endocrine Reviews*, *36*(5), 564–591. https://doi.org/10.1210/er.2014-1101 (cited on pages 13, 14)

Mu, X., Murakami, H., Mohibullah, N., & Keeney, S. (2020). Chromosome-autonomous feedback down-regulates meiotic DSB competence upon synaptonemal complex formation. *bioRxiv*, 1–39. https://doi.org/https://doi.org/10.1101/2020.05.11.089367 (cited on page 19)

Muller, H. (1942). Isolating mechanisms, evolution, and temperature. *Biol. Symp.*, *6*, 71–125 (cited on page 22).

Muller, H., Scolari, V. F., Agier, N., Piazza, A., Thierry, A., Mercy, G., Descorps-Declere, S., Lazar-Stefanita, L., Espeli, O., Llorente, B., Fischer, G., Mozziconacci, J., & Koszul, R. (2018). Characterizing meiotic chromosomes' structure and pairing using a designer sequence optimized for Hi-C. *Molecular Systems Biology*, *14*(7), 1–19. https://doi.org/10.15252/msb.20188293 (cited on page 194)

Muller, H. J. (1916). The Mechanism of Crossing-Over. *The American Naturalist*, *50*(592), 193–221. https://doi.org/10.1086/279534 (cited on page 19)

Muñoz-Fuentes, V., Di Rienzo, A., & Vilà, C. (2011). Prdm9, a Major Determinant of Meiotic Recombination Hotspots, Is Not Functional in Dogs and Their Wild Relatives, Wolves and Coyotes (D. S. Dawson, Ed.). *PLoS ONE*, *6*(11), e25498. https://doi.org/10.1371/journal.pone.0025498 (cited on page 19)

Muñoz-Muñoz, F., Sans-Fuentes, M. A., Lopez-Fuster, M. J., & Ventura, J. (2006). Variation in fluctuating asymmetry levels across a Robertsonian polymorphic zone of the house mouse. *Journal of Zoological Systematics and Evolutionary Research*, *44*(3), 236–250. https://doi.org/10.1111/j.1439-0469.2006.00357.x (cited on page 24)

Muñoz-Muñoz, F., Sans-Fuentes, M. A., López-Fuster, M. J., & Ventura, J. (2003). Non-metric morphological divergence in the western house mouse, Mus musculus domesticus, from the Barcelona chromosomal hybrid zone. *Biological Journal of the Linnean Society*, *80*(2), 313–322. https://doi.org/10.1046/j.1095-8312.2003.00240.x (cited on page 24)

Myers, S., Bowden, R., Tumian, A., Bontrop, R. E., Freeman, C., MacFie, T. S., McVean, G., & Donnelly, P. (2010). Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, *327*(5967), 876–879. https://doi.org/10.1126/science.1182363 (cited on pages 18, 19)

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A., & Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, *502*(7469), 59–64. https://doi.org/10.1038/nature12593 (cited on pages 6, 7, 11)

Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., Mendelson Cohen, N., Wingett, S., Fraser, P., & Tanay, A. (2017). Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*, *547*(7661), 61–67. https://doi.org/10.1038/nature23001 (cited on pages 8, 11)

Nakagawa, T., Sharma, M., Nabeshima, Y.-i., Braun, R. E., & Yoshida, S. (2010). Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment. *Science*, *328*(5974), 62–67. https://doi.org/10.1126/science.1182868 (cited on page 14)

Namekawa, S. H., Park, P. J., Zhang, L.-F., Shima, J. E., McCarrey, J. R., Griswold, M. D., & Lee, J. T. (2006). Postmeiotic Sex Chromatin in the Male Germline of Mice. *Current Biology*, *16*(7), 660–667. https://doi.org/10.1016/j.cub.2006.01.066 (cited on pages 16, 17, 198)

Nanda, I., Schneider-Rasp, S., Winking, H., & Schmid, M. (1995). Loss of telomeric sites in the chromosomes of Mus musculus domesticus (Rodentia: Muridae) during Robertsonian rearrangements. *Chromosome Research*, *3*(7), 399–409. https://doi.org/10.1007/BF00713889 (cited on page 23)

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. A., & Dekker, J. (2013). Organization of the mitotic chromosome. *Science*, *342*(6161), 948–53. https://doi.org/10.1126/science.1236083 (cited on pages 8, 11, 193, 194)

Navarro, A., & Barton, N. H. (2003). Chromosomal Speciation and Molecular Divergence–Accelerated Evolution in Rearranged Chromosomes. *Science*, *300*(5617), 321–324. https://doi.org/10.1126/science.1080600 (cited on pages 22, 23)

Navarro, A., Betrán, E., Barbadilla, A., & Ruiz, A. (1997). Recombination and Gene Flux Caused by Gene Conversion and Crossing Over in Inversion Heterokaryotypes. *Genetics*, *146*(2) (cited on page 23).

Navone, F., Consalez, G. G., Sardella, M., Caspani, E., Pozzoli, O., Frassoni, C., Morlacchi, E., Sitia, R., Sprocati, T., & Cabibbo, A. (2001). Expression of KIF3C kinesin during neural development and in vitro neuronal differentiation. *Journal of Neurochemistry*, *77*(3), 741–753. https://doi.org/10.1046/j.1471-4159.2001.00277.x (cited on page 197)

Nebel, B. R., Amarose, A. P., & Hackett, E. M. (1961). Calendar of Gametogenic Development in the Prepuberal Male Mouse. *Science*, *134*(3482), 832–833. https://doi.org/10.1126/science.134.3482.832 (cited on page 12)

Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, *98*(21), 12084–12088. https://doi.org/10.1073/PNAS.221274498 (cited on pages 23, 202)

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Bluthgen, N., Dekker, J., & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*(7398), 381–385. https://doi.org/10.1038/nature11049 (cited on pages 4, 8)

Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., & Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, *169*(5), 930–944. https://doi.org/10.1016/j.cell.2017.05.004 (cited on page 196)

Norton, H. K., & Phillips-Cremins, J. E. (2017). Crossed wires: 3D genome misfolding in human disease. *Journal of Cell Biology*, *216*(11), 3441–3452. https://doi.org/10.1083/jcb.201611001 (cited on page 9)

Nurick, I., Shamir, R., & Elkon, R. (2018). Genomic meta-analysis of the interplay between 3D chromatin organization and gene expression programs under basal and stress conditions. *Epigenetics and Chromatin*, *11*(1), 1–14. https://doi.org/10.1186/s13072-018-0220-2 (cited on pages 7, 193, 194)

Oakberg, E. F. (1956a). Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. *The American Journal of Anatomy*, *99*, 507–516. https://doi.org/10.1002/aja.1000990307 (cited on page 13)

Oakberg, E. F. (1971). Spermatogonial stem-cell renewal in the mouse. *The Anatomical Record*, *169*(3), 515–531. https://doi.org/10.1002/ar.1091690305 (cited on page 14)

Oakberg, E. F. (1956b). A description of spermiogenesis in the mouse and its use in analysis of the cycle of the seminiferous epithelium and germ cell renewal. *American Journal of Anatomy*, *99*(3), 391–413. https://doi.org/10.1002/aja.1000990303 (cited on page 16)

O'Donnell, L., Stanton, P., & de Kretser, D. M. (2000). *Endocrinology of the Male Reproductive System and Spermatogenesis* (K. R. Feingold, B. Anawalt, A. Boyce, G. Chrousos, K. Dungan, A. Grossman, J. M. Hershman, G. Kaltsas, C. Koch, P. Kopp, M. Korbonits, R. McLachlan, J. E. Morley, M. New, L. Perreault, J. Purnell, R. Rebar, F. Singer, D. L. Trence, . . . D. P. Wilson, Eds.). (Cited on page 14).

Ohlsson, R., Renkawitz, R., & Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends in Genetics*, *17*(9), 520–527. https://doi.org/10.1016/S0168-9525(01)02366-6 (cited on page 9)

O'Keefe, R. T., Henderson, S. C., & Spector, D. L. (1992). Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *The Journal of Cell Biology*, *116*(5), 1095–1110. https://doi.org/10.1083/jcb.116.5.1095 (cited on page 4)

Oliva, R. (2006). Protamines and male infertility. *Human Reproduction Update*, *12*(4), 417–435. https://doi.org/10.1093/humupd/dml009 (cited on page 17)

Oliver, P. L., Goodstadt, L., Bayes, J. J., Birtle, Z., Roach, K. C., Phadnis, N., Beatson, S. A., Lunter, G., Malik, H. S., & Ponting, C. P. (2009). Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa (M. W. Nachman, Ed.). *PLoS Genetics*, *5*(12), e1000753. https://doi.org/10.1371/journal.pgen.1000753 (cited on page 21)

Ortiz-Barrientos, D., Engelstädter, J., & Rieseberg, L. H. (2016). Recombination Rate Evolution and the Origin of Species. *Trends in Ecology & Evolution*, *31*(3), 226–236. https://doi.org/10.1016/j.tree.2015.12.016 (cited on page 21)

Otto, S. P., & Payseur, B. A. (2019). Crossover Interference: Shedding Light on the Evolution of Recombination. *Annual Review of Genetics*, *53*(1), 19–44. https://doi.org/10.1146/annurev-genet-040119-093957 (cited on page 19)

Paigen, K., & Petkov, P. M. (2018). PRDM9 and Its Role in Genetic Recombination. *Trends in Genetics*, *20*(4), 1–10. https://doi.org/10.1016/j.tig.2017.12.017 (cited on page 203)

Palozola, K. C., Donahue, G., Liu, H., Grant, G. R., Becker, J. S., Cote, A., Yu, H., Raj, A., & Zaret, K. S. (2017). Mitotic transcription and waves of gene reactivation during mitotic exit. *Science*, *358*(6359), 119 LP –122. https://doi.org/10.1126/science.aal4671 (cited on page 195)

Panigrahi, A. K., Zhang, N., Otta, S. K., & Pati, D. (2012). A cohesin–RAD21 interactome. *Biochemical Journal*, *442*(3), 661–670. https://doi.org/10.1042/BJ20111745 (cited on page 196)

Parada, L. A., & Misteli, T. (2002). Chromosome positioning in the interphase nucleus. *Trends in Cell Biology*, *12*(9), 425–432. https://doi.org/10.1016/S0962-8924(02)02351-6 (cited on page 4)

Parvanov, E. D., Petkov, P. M., & Paigen, K. (2010). Prdm9 controls activation of mammalian recombination hotspots. *Science*, *327*(5967), 835. https://doi.org/10.1126/science.1181495 (cited on pages 18–21, 47)

Parvanov, E. D., Tian, H., Billings, T., Saxl, R. L., Spruce, C., Aithal, R., Krejci, L., Paigen, K., & Petkov, P. M. (2017). PRDM9 interactions with other proteins provide a link between recombination hotspots and the chromosomal axis in meiosis. *Molecular Biology of the Cell*, *28*(3), 488–499. https://doi.org/10.1091/mbc.E16-09-0686 (cited on page 20)

Patel, A., Horton, J. R., Wilson, G. G., Zhang, X., & Cheng, X. (2016). Structural basis for human PRDM9 action at recombination hot spots. *Genes and Development*, *30*(3), 257–265. https://doi.org/10.1101/gad.274928.115 (cited on page 20)

Patel, A., Zhang, X., Blumenthal, R. M., & Cheng, X. (2017). Structural basis of human PR/SET domain 9 (PRDM9) allele C-specific recognition of its cognate DNA sequence. *The Journal of Biological Chemistry*, *292*(39), 15994–16002. https://doi.org/10.1074/jbc.M117.805754 (cited on page 20)

Patel, L., Kang, R., Rosenberg, S. C., Qiu, Y., Raviram, R., Chee, S., Hu, R., Ren, B., Cole, F., & Corbett, K. D. (2019). Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. *Nature Structural & Molecular Biology*, *26*, 164–174. https://doi.org/10.1038/s41594-019-0187-0 (cited on pages 119, 126, 129, 194–196)

Paytuví-Gallart, A. (2019). *Development and application of integrative tools for the functional and structural analyses of genomes* (Doctoral dissertation). (Cited on pages 8, 52, 119).

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M. A., Fuchs, J., & Schubert, I. (2004). Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. *Chromosoma*, *113*(5), 258–269. https://doi.org/10.1007/s00412-004-0316-2 (cited on page 3)

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*(5), 946–952. https://doi.org/10.1111/1755-0998.12129 (cited on page 137)

Peric-Hupkes, D., & van Steensel, B. (2010). Role of the Nuclear Lamina in Genome Organization and Gene Expression. *Cold Spring Harbor Symposia on Quantitative Biology*, *75*, 517–524. https://doi.org/10.1101/sqb.2010.75.014 (cited on page 4)

Petes, T. D. (2001). Meiotic recombination hot spots and cold spots. *Nature Reviews Genetics*, *2*(5), 360–369. https://doi.org/10.1038/35072078 (cited on page 122)

Phillips, B. T., Gassei, K., & Orwig, K. E. (2010). Spermatogonial stem cell regulation and spermatogenesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *365*(1546), 1663–1678. https://doi.org/10.1098/rstb.2010.0026 (cited on pages 12–14, 66, 194)

Phillips, J. E., & Corces, V. G. (2009). CTCF: Master Weaver of the Genome. *Cell*, *137*(7), 1194–1211. https://doi.org/10.1016/j.cell.2009.06.001 (cited on page 9)

Phillips-Cremins, J. E., & Corces, V. G. (2013). Chromatin Insulators: Linking Genome Organization to Cellular Function. *Molecular Cell*, *50*(4), 461–474. https://doi.org/10.1016/j.molcel.2013.04.018 (cited on page 196)

Phillips-Cremins, J. E., Sauria, M. E., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S., Ong, C. T., Hookway, T. A., Guo, C., Sun, Y., Bland, M. J., Wagstaff, W., Dalton, S., McDevitt, T. C., Sen, R., Dekker, J., Taylor, J., & Corces, V. G. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, *153*(6), 1281–1295. https://doi.org/10.1016/j.cell.2013.04.053 (cited on page 8)

Piálek, J., Hauffe, H. C., & Searle, J. B. (2005). Chromosomal variation in the house mouse. *Biological Journal of the Linnean Society*, *84*(3), 535–563. https://doi.org/10.1111/j.1095-8312.2005.00454.x (cited on pages 23–25, 120, 134)

Picot, J., Guerin, C. L., Le Van Kim, C., & Boulanger, C. M. (2012). Flow cytometry: Retrospective, fundamentals and recent instrumentation. *Cytotechnology*, *64*(2), 109–130. https://doi.org/10.1007/s10616-011-9415-0 (cited on pages 35, 36)

Pinkel, D., Straume, T., & Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(9), 2934–2938. https://doi.org/10.1073/pnas.83.9.2934 (cited on page 3)

Pittman, D. L., Weinberg, L. R., & Schimenti, J. C. (1998). Identification, characterization, and genetic mapping of Rad51d, a new mouse and human RAD51/RecA-related gene. *Genomics*, *49*(1), 103–111. https://doi.org/10.1006/geno.1998.5226 (cited on page 18)

Platz, R. D., Grimes, S. R., Meistrich, M. L., & Hnilica, L. S. (1975). Changes in nuclear proteins of rat testis cells separated by velocity sedimentation. *The Journal of biological chemistry*, *250*(15), 5791–5800 (cited on page 57).

Ponting, C. P. (2011). What are the genomic drivers of the rapid evolution of PRDM9? *Trends in Genetics*, *27*(5), 165–171. https://doi.org/10.1016/j.tig.2011.02.001 (cited on page 21)

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K., Thurman, R. E., Cheng, Y., Gulsoy, G., Dennis, J. H., Snyder, M. P., Stamatoyannopoulos, J. A., Taylor, J., Hardison, R. C., Kahveci, T., . . . Gilbert, D. M. (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature*, *515*(7527), 402–405. https://doi.org/10.1038/nature13986 (cited on page 8)

Powers, N. R., Parvanov, E. D., Baker, C. L., Walker, M., Petkov, P. M., & Paigen, K. (2016). The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS Genetics*, *12*(6), e1006146. https://doi.org/10.1371/journal.pgen.1006146 (cited on page 20)

Pratto, F., Brick, K., Khil, P., Smagulova, F., Petukhova, G. V., & Camerini-Otero, R. D. (2014). DNA recombination. Recombination initiation maps of individual human genomes. *Science*, *346*(6211), 1256442. https://doi.org/10.1126/science.1256442 (cited on page 20)

Pugacheva, E. M., Rivero-Hinojosa, S., Espinoza, C. A., Méndez-Catalá, C. F., Kang, S., Suzuki, T., Kosaka-Suzuki, N., Robinson, S., Nagarajan, V., Ye, Z., Boukaba, A., Rasko, J. E., Strunnikov, A. V., Loukinov, D., Ren, B., & Lobanenkov, V. V. (2015). Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. *Genome Biology*, *16*(1), 1–24. https://doi.org/10.1186/s13059-015-0736-8 (cited on page 197)

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795 (cited on pages 51, 136, 137)

Qiao, H., Chen, J. K., Reynolds, A., Höög, C., Paddy, M., & Hunter, N. (2012). Interplay between Synaptonemal Complex, Homologous Recombination, and Centromeres during Mammalian Meiosis. *PLoS Genetics*, *8*(6), e1002790. https://doi.org/10.1371/journal.pgen.1002790 (cited on page 200)

R Core Team. (2018). R: a language and environment for statistical computing. *Vienna, Austria*, R Foundation for Statistical Computing. https://www.r-project.org/. (Cited on page 51)

Rabl, C., & Gegenbauer, C. (1885). Morphologisches Jahrbuch. . *Gegangaur, ed*, *10*, 214–330 (cited on page 3).

Racko, D., Benedetti, F., Dorier, J., & Stasiak, A. (2019). Are TADs supercoiled? *Nucleic Acids Research*, *47*(2), 521–532. https://doi.org/10.1093/nar/gky1091 (cited on page 10)

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., Habermann, B., Akhtar, A., & Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature communications*, *9*(1), 189. https://doi.org/10.1038/s41467-017-02525-w (cited on page 52)

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(1), 160–165. https://doi.org/10.1093/nar/gkw257 (cited on page 52)

Rankin, S. (2015). Complex elaboration: Making sense of meiotic cohesin dynamics. *FEBS Journal*, *282*(13), 2413–2430. https://doi.org/10.1111/febs.13301 (cited on page 17)

Rao, S. S. P., Huang, S.-C., Glenn, B., Hilaire, S., Casellas, R., Lander, E. S., Lieberman, E., Correspondence, A., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D., Huang, X., Shamim, M. S., Shin, J., Turner, D., . . . Lieberman Aiden, E. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell*, *171*, 305–320. https://doi.org/10.1016/j.cell.2017.09.026 (cited on pages 9, 10, 195)

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021 (cited on pages 4–10, 37, 119, 125, 129, 193, 194)

Rathke, C., Baarends, W. M., Awe, S., & Renkawitz-Pohl, R. (2014). Chromatin dynamics during spermiogenesis. *Biochimica et Biophysica Acta*, *1839*(3), 155–168. https://doi.org/10.1016/j.bbagrm.2013.08.004 (cited on pages 17, 196)

Reeves, A. (2001). MicroMeasure: a new computer program for the collection and analysis of cytogenetic data. *Genome*, *44*(3), 439–43 (cited on pages 44, 136).

Reig-Viader, R., Garcia-Caldés, M., & Ruiz-Herrera, A. (2016). Telomere homeostasis in mammalian germ cells: a review. *Chromosoma*, *125*, 337–351. https://doi.org/10.1007/s00412-015-0555-4 (cited on pages 12, 14, 15)

Ribagorda, M., Berríos, S., Solano, E., Ayarza, E., Martín-Ruiz, M., Gil-Fernández, A., Parra, M. T., Viera, A., Rufas, J. S., Capanna, E., Castiglia, R., Fernández-Donoso, R., & Page, J. (2019). Meiotic behavior of a complex hexavalent in heterozygous mice for Robertsonian translocations: insights for synapsis dynamics. *Chromosoma*, *128*(2), 149–163. https://doi.org/10.1007/s00412-019-00695-8 (cited on pages 25, 125)

Ribatti, D. (2018). An historical note on the cell theory. *Experimental Cell Research*, *364*(1), 1–4. https://doi.org/10.1016/j.yexcr.2018.01.038 (cited on page 3)

Ribeiro, J., Abby, E., Livera, G., & Martini, E. (2016). RPA homologs and ssDNA processing during meiotic recombination. *Chromosoma*, *125*(2), 265–276. https://doi.org/10.1007/s00412-015-0552-7 (cited on page 18)

Riccardi, C., & Nicoletti, I. (2006). Analysis of apoptosis by propidium iodide staining and flow cytometry. *Nature Protocols*, *1*(3), 1458–1461. https://doi.org/10.1038/nprot.2006.238 (cited on page 58)

Richardson, B. E., & Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nature Reviews Molecular Cell Biology*, *11*(1), 37–49. https://doi.org/10.1038/nrm2815 (cited on page 12)

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, *16*(7), 351–358. https://doi.org/10.1016/B978-0-12-800049-6.00074-3 (cited on pages 23, 125, 201, 202)

Rivero-Hinojosa, S., Kang, S., Lobanenkov, V. V., & Zentner, G. E. (2017). Testis-specific transcriptional regulators selectively occupy BORIS-bound CTCF target regions in mouse male germ cells. *Scientific Reports*, *7*, 41279. https://doi.org/10.1038/srep41279 (cited on page 197)

Robert, T., Nore, A., Brun, C., Maffre, C., Crimi, B., Bourbon, H.-M., & de Massy, B. (2016). The TopoVIB-Like protein family is required for meiotic DNA double-strand break formation. *Science*, *351*(6276), 943–949. https://doi.org/10.1126/science.aad5309 (cited on page 18)

Robertson, W. R. B. (1916). Chromosome studies. I. Taxonomic relationships shown in the chromosomes of tettigidae and acrididae: V-shaped chromosomes and their significance in acrididae, locustidae, and gryllidae: Chromosomes and variation. *Journal of Morphology*, *27*(2), 179–331. https://doi.org/10.1002/jmor.1050270202 (cited on pages 23, 119)

Robinson, T. J. (1995). King, M., Species evolution: the role of chromosome change. (M. King, Ed.). *Systematic Biology*, *44*(4), 578. https://doi.org/10.2307/2413666 (cited on pages 23, 119)

Rodríguez-Casuriaga, R., Santiñaque, F. F., Folle, G. A., Souza, E., López-Carro, B., & Geisinger, A. (2014). Rapid preparation of rodent testicular cell suspensions and spermatogenic stages purification by flow cytometry using a novel blue-laser-excitable vital dye. *MethodsX*, *1*(October), 239–243. https://doi.org/10.1016/j.mex.2014.10.002 (cited on page 57)

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., & Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *The Journal of Biological Chemistry*, *273*(10), 5858–5868. https://doi.org/10.1074/jbc.273.10.5858 (cited on page 18)

Romanienko, P. J., & Camerini-Otero, R. D. (2000). The mouse Spo11 gene is required for meiotic chromosome synapsis. *Molecular Cell*, *6*(5), 975–987. https://doi.org/10.1016/S1097-2765(00)00097-6 (cited on pages 15, 18)

Romrell, L. J., Bellvé, A. R., & Fawcett, D. W. (1976). Separation of mouse spermatogenic cells by sedimentation velocity. A morphological characterization. *Developmental Biology*, *49*(1), 119–131. https://doi.org/10.1016/0012-1606(76)90262-1 (cited on pages 13, 57, 67)

Rong, M., Matsuda, A., Hiraoka, Y., & Lee, J. (2016). Meiotic cohesin subunits RAD21L and REC8 are positioned at distinct regions between lateral elements and transverse filaments in the synaptonemal complex of mouse spermatocytes. *Journal of Reproduction and Development*, *62*(6), 623–630. https://doi.org/10.1262/jrd.2016-127 (cited on page 195)

Rong, M., Miyauchi, S., & Lee, J. (2017). Ectopic expression of meiotic cohesin RAD21L promotes adjacency of homologous chromosomes in somatic cells. *Journal of Reproduction and Development*, *63*(3), 227–234. https://doi.org/10.1262/jrd.2016-171 (cited on page 198)

Roosen-Runge, E. C., & Giesel, L. O. (1950). Quantitative studies on spermatogenesis in the albino rat. *American Journal of Anatomy*, *87*(1), 1–30. https://doi.org/10.1002/aja.1000870102 (cited on page 14)

Rowley, M. J., & Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nature Reviews Genetics*, *19*(12), 789–800. https://doi.org/10.1038/s41576-018-0060-8 (cited on pages 7, 9–11, 193)

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., Wang, P., Ruan, Y., & Corces, V. G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Molecular Cell*, *67*(5), 837–852. https://doi.org/10.1016/j.molcel.2017.07.022 (cited on pages 5–7)

Royo, H., Polikiewicz, G., Mahadevaiah, S. K., Prosser, H., Mitchell, M., Bradley, A., de Rooij, D. G., Burgoyne, P. S., & Turner, J. M. A. (2010). Evidence that meiotic sex chromosome inactivation is essential for male fertility. *Current Biology*, *20*(23), 2117–2123. https://doi.org/10.1016/j.cub.2010.11.010 (cited on page 199)

Ruiz-Herrera, A., Vozdova, M., Fernández, J., Sebestova, H., Capilla, L., Frohlich, J., Vara, C., Hernández-Marsal, A., Sipek, J., Robinson, T. J., & Rubes, J. (2017). Recombination correlates with synaptonemal complex length and chromatin loop size in bovids—insights into mammalian meiotic chromosomal organization. *Chromosoma*, *126*(5), 615–631. https://doi.org/10.1007/s00412-016-0624-3 (cited on pages 17–19, 44, 120, 122, 123, 133, 135, 136, 196, 200, 201)

Sanborn, A. L., Rao, S. S. P., Huang, S.-C. C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J., Geeting, K. P., Gnirke, A., Melnikov, A., McKenna, D., Stamenova, E. K., Lander, E. S., & Aiden, E. L. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), E6456–E6465. https://doi.org/10.1073/pnas.1518552112 (cited on page 10)

Sanborn, B. M., Steinberger, A., Meistrich, M. L., & Steinberger, E. (1975). Androgen binding sites in testis cell fractions as measured by a nuclear exchange assay. *Journal of Steroid Biochemistry*, *6*(11), 1459–1465. https://doi.org/https://doi.org/10.1016/0022-4731(75)90197-1 (cited on page 57)

Sánchez-Guillén, R. A., Capilla, L., Reig-Viader, R., Martínez-Plana, M., Pardo-Camacho, C., Andrés-Nieto, M., Ventura, J., & Ruiz-Herrera, A. (2015). On the origin of Robertsonian fusions in nature: evidence of telomere shortening in wild house mice. *Journal of Evolutionary Biology*, *28*(1), 241–249. https://doi.org/10.1111/jeb.12568 (cited on pages 23, 34, 135, 198, 203)

Sandor, C., Li, W., Coppieters, W., Druet, T., Charlier, C., & Georges, M. (2012). Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle. *PLoS Genetics*, *8*(7), e1002854. https://doi.org/10.1371/journal.pgen.1002854 (cited on page 21)

Sans-Fuentes, M. A., García-Valero, J., Ventura, J., & López-Fuster, M. J. (2010). Spermatogenesis in house mouse in a Robertsonian polymorphism zone. *Reproduction*, *140*(4), 569–581. https://doi.org/10.1530/REP-10-0237 (cited on pages 25, 134, 198–200, 202, 203)

Sans-Fuentes, M. A., López-Fuster, M. J., Ventura, J., Díez-Noguera, A., & Cambras, T. (2005). Effect of Robertsonian Translocations on the Motor Activity Rhythm in the House Mouse. *Behavior Genetics*, *35*(5), 603–613. https://doi.org/10.1007/s10519-005-5375-5 (cited on pages 23, 24)

Sans-Fuentes, M. A., Muñoz-Muñoz, F., Ventura, J., & López-Fuster, M. J. (2007). Rb(7.17), a rare Robertsonian fusion in wild populations of the house mouse. *Genetical Research*, *89*(4), 207–213. https://doi.org/10.1017/S0016672307008993 (cited on page 23)

SAS Institute Inc. (2019). JMP package. Cary, NC. https://www.jmp.com/en_us/home.html. (Cited on pages 51, 137)

Sassone-Corsi, P. (2002). Unique Chromatin Remodeling and Transcriptional Regulation in Spermatogenesis. *Science*, *296*(5576), 2176–2178. https://doi.org/10.1126/science.1070963 (cited on page 16)

Schalbetter, S. A., Fudenberg, G., Baxter, J., Pollard, K. S., & Neale, M. J. (2019). Principles of meiotic chromosome assembly revealed in S. cerevisiae. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-12629-0 (cited on pages 194, 195)

Scherthan, H., Weich, S., Schwegler, H., Heyting, C., Härle, M., & Cremer, T. (1996). Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing. *The Journal of Cell Biology*, *134*(5), 1109–1125. https://doi.org/10.1083/jcb.134.5.1109 (cited on pages 14, 15, 23)

Scriven, P. (2001). Robertsonian translocations–reproductive risks and indications for preimplantation genetic diagnosis. *Human Reproduction*, *16*(11), 2267–2273. https://doi.org/10.1093/humrep/16.11.2267 (cited on pages 25, 119)

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brannstrom, A., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., . . . Widmer, A. (2014). Genomics and the origin of species. *Nat Rev Genet*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644 (cited on pages 22, 23)

Segura, J., Ferretti, L., Ramos-Onsins, S., Capilla, L., Farré, M., Reis, F., Oliver-Bonet, M., Fernández-Bellón, H., Garcia, F., Garcia-Caldés, M., Robinson, T. J., & Ruiz-Herrera, A. (2013). Evolution of recombination in eutherian mammals: insights into mechanisms that affect recombination rates and crossover interference. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1771), 20131945. https://doi.org/10.1098/rspb.2013.1945 (cited on pages 44, 135, 136, 202)

Serra, F., Baù, D., Goodstadt, M., Castillo, D., Filion, G. J., & Marti-Renom, M. A. (2017). Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors (A. Prlic, Ed.). *PLOS Computational Biology*, *13*(7), e1005665. https://doi.org/10.1371/journal.pcbi.1005665 (cited on pages 8, 52, 138)

Sexton, T., Kurukuti, S., Mitchell, J. A., Umlauf, D., Nagano, T., & Fraser, P. (2012a). Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. *Nature Protocols*, *7*(7), 1335–1350. https://doi.org/10.1038/nprot.2012.071 (cited on page 7)

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., & Cavalli, G. (2012b). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, *148*(3), 458–472. https://doi.org/10.1016/j.cell.2012.01.010 (cited on pages 4–6, 8, 193)

Sharma, R., & Agarwal, A. (2011). Spermatogenesis: An Overview. In *Sperm chromatin* (pp. 19–44). New York, NY, Springer New York. https://doi.org/10.1007/978-1-4419-6857-9{_}2. (Cited on pages 12, 13, 16, 17)

Shaw, P. J., Abranches, R., Paula Santos, A., Beven, A. F., Stoger, E., Wegel, E., & Gonzalez-Melendi, P. (2002). The architecture of interphase chromosomes and nucleolar transcription sites in plants. *Journal of Structural Biology*, *140*(1-3), 31–38. https://doi.org/10.1016/s1047-8477(02)00537-3 (cited on page 3)

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., & Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–120. https://doi.org/10.1038/nature11243 (cited on page 8)

Shirakawa, T., Yaman-Deveci, R., Tomizawa, S.-i., Kamizato, Y., Nakajima, K., Sone, H., Sato, Y., Sharif, J., Yamashita, A., Takada-Horisawa, Y., Yoshida, S., Ura, K., Muto, M., Koseki, H., Suda, T., & Ohbo, K. (2013). An epigenetic switch is crucial for spermatogonia to exit the undifferentiated state toward a Kit-positive identity. *Development*, *140*(17), 3565 LP –3576. https://doi.org/10.1242/dev.094045 (cited on page 194)

Shopland, L. S., Lynch, C. R., Peterson, K. A., Thornton, K., Kepper, N., Hase, J. v., Stein, S., Vincent, S., Molloy, K. R., Kreth, G., Cremer, C., Bult, C. J., & O'Brien, T. P. (2006). Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. *The Journal of Cell Biology*, *174*(1), 27–38. https://doi.org/10.1083/jcb.200603083 (cited on page 4)

Sikorska, N., & Sexton, T. (2019). Defining functionally relevant spatial chromatin domains: it's a TAD complicated. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2019.12.006 (cited on page 8)

Sin, H.-S., Barski, A., Zhang, F., Kartashov, A. V., Nussenzweig, A., Chen, J., Andreassen, P. R., & Namekawa, S. H. (2012). RNF8 regulates active epigenetic modifications and escape gene activation from inactive sex chromosomes in post-meiotic spermatids. *Genes and Development*, *26*(24), 2737–2748. https://doi.org/10.1101/gad.202713.112 (cited on page 17)

Sin, H.-S., Kartashov, A. V., Hasegawa, K., Barski, A., & Namekawa, S. H. (2015). Poised chromatin and bivalent domains facilitate the mitosis-to-meiosis transition in the male germline. *BMC Biology*, *13*, 53. https://doi.org/10.1186/s12915-015-0159-8 (cited on pages 17, 197)

Skibbens, R. V. (2016). Of Rings and Rods: Regulating Cohesin Entrapment of DNA to Generate Intra- and Intermolecular Tethers. *PLoS Genetics*, *12*(10), 1–12. https://doi.org/10.1371/journal.pgen.1006337 (cited on page 10)

Slijepcevic, P. (1998). Telomeres and mechanisms of Robertsonian fusion. *Chromosoma*, *107*(2), 136–140. https://doi.org/10.1007/s004120050289 (cited on page 23)

Smagulova, F., Brick, K., Pu, Y., Camerini-Otero, R. D., & Petukhova, G. V. (2016). The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes and Development*, *30*(3), 266–280. https://doi.org/10.1101/gad.270009.115 (cited on page 22)

Smagulova, F., Brick, K., Pu, Y., Sengupta, U., Camerini-Otero, R. D., & Petukhova, G. V. (2013). Suppression of genetic recombination in the pseudoautosomal region and at subtelomeres in mice with a hypomorphic Spo11 allele. *BMC Genomics*, *14*, 493. https://doi.org/10.1186/1471-2164-14-493 (cited on page 18)

Smagulova, F., Gregoretti, I. V., Brick, K., Khil, P., Camerini-Otero, R. D., & Petukhova, G. V. (2011). Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature*, *472*(7343), 375–378. https://doi.org/10.1038/nature09869 (cited on pages 20, 21, 203)

Smit, A., Hubley, R., & Green, P. (2015). RepeatMasker Open-4.0. (Cited on page 139).

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S. M., Schroth, G. P., Tanay, A., & Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO Journal*, *32*(24), 3119–3129. https://doi.org/10.1038/emboj.2013.237 (cited on pages 10, 196)

Sohni, A., Tan, K., Song, H. W., Burow, D., de Rooij, D. G., Laurent, L., Hsieh, T. C., Rabah, R., Hammoud, S. S., Vicini, E., & Wilkinson, M. F. (2019). The Neonatal and Adult Human Testis Defined at the Single-Cell Level. *Cell Reports*, *26*(6), 1501–1517. https://doi.org/10.1016/j.celrep.2019.01.045 (cited on page 67)

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., Dym, M., deMassy, B., Mikkelsen, T. S., & Kaessmann, H. (2013). Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. *Cell Reports*, *3*(6), 2179–2190. https://doi.org/10.1016/j.celrep.2013.05.031 (cited on pages 16, 198, 199)

Spindler, M. C., Filbeck, S., Stigloher, C., & Benavente, R. (2019). Quantitative basis of meiotic chromosome synapsis analyzed by electron tomography. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-52455-4 (cited on page 195)

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N., & De Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes and Development*, *20*(17), 2349–2354. https://doi.org/10.1101/gad.399506 (cited on page 9)

Spruce, C., Dlamini, S., Ananda, G., Bronkema, N., Tian, H., Paigen, K., Carter, G. W., & Baker, C. L. (2020). HELLS and PRDM9 form a pioneer complex to open chromatin at meiotic recombination hot spots. *Genes and Development*, *34*(5-6), 398–412. https://doi.org/10.1101/gad.333542.119 (cited on page 20)

Stack, S. M., Brown, D. B., & Dewey, W. C. (1977). Visualization of interphase chromosomes. *Journal of Cell Science*, *26*(1), 281 LP –299 (cited on page 3).

Stanzione, M., Baumann, M., Papanikos, F., Dereli, I., Lange, J., Ramlal, A., Tränkner, D., Shibuya, H., De Massy, B., Watanabe, Y., Jasin, M., Keeney, S., & Tóth, A. (2016). Meiotic DNA break formation requires the unsynapsed chromosome axis-binding protein IHO1 (CCDC36) in mice. *Nature Cell Biology*, *18*(11), 1208–1220. https://doi.org/10.1038/ncb3417 (cited on page 18)

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *372*(1736), 20160455. https://doi.org/10.1098/rstb.2016.0455 (cited on pages 19, 21)

Steger, K. (1999). Transcriptional and translational regulation of gene expression in haploid spermatids. *Anatomy and Embryology*, *199*(6), 471–487. https://doi.org/10.1007/s004290050245 (cited on page 17)

Stigler, J., Camdere, G. O., Koshland, D. E., & Greene, E. C. (2016). Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. *Cell Reports*, *15*(5), 988–998. https://doi.org/10.1016/j.celrep.2016.04.003 (cited on page 10)

Striedner, Y., Schwarz, T., Welte, T., Futschik, A., Rant, U., & Tiemann-Boege, I. (2017). The long zinc finger domain of PRDM9 forms a highly stable and long-lived complex with its DNA recognition sequence. *Chromosome Research*. https://doi.org/10.1007/s10577-017-9552-1 (cited on page 48)

Struijk, R. B., De Winter-Korver, C. M., van Daalen, S. K., Hooibrink, B., Repping, S., & van Pelt, A. M. (2019). Simultaneous Purification of Round and Elongated Spermatids from Testis Tissue Using a FACS-Based DNA Ploidy Assay. *Cytometry Part A*, *95*(3), 309–313. https://doi.org/10.1002/cyto.a.23698 (cited on pages 57, 67)

Subramanian, V. V., & Hochwagen, A. (2014). The Meiotic Checkpoint Network: Step-by-Step through Meiotic Prophase. *Cold Spring Harbor Perspectives in Biology*, *6*(10), a016675–a016675. https://doi.org/10.1101/cshperspect.a016675 (cited on page 198)

Sun, F., Xu, Q., Zhao, D., & Degui Chen, C. (2015a). Id4 Marks Spermatogonial Stem Cells in the Mouse Testis. *Scientific Reports*, *5*, 17594. https://doi.org/10.1038/srep17594 (cited on page 14)

Sun, F., Fujiwara, Y., Reinholdt, L. G., Hu, J., Saxl, R. L., Baker, C. L., Petkov, P. M., Paigen, K., & Handel, M. A. (2015b). Nuclear localization of PRDM9 and its role in meiotic chromatin modifications and homologous synapsis. *Chromosoma*, *124*(3), 397–415. https://doi.org/10.1007/s00412-015-0511-3 (cited on page 20)

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., & Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Research*, *24*(3), 390–400. https://doi.org/10.1101/gr.163519.113 (cited on page 193)

Syrjänen, J. L., Heller, I., Candelli, A., Davies, O. R., Peterman, E. J. G., Wuite, G. J. L., & Pellegrini, L. (2017). Single-molecule observation of DNA compaction by meiotic protein SYCP3. *eLife*, *6*. https://doi.org/10.7554/eLife.22582 (cited on page 17)

Szabo, Q., Bantignies, F., & Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Science Advances*, *5*(4). https://doi.org/10.1126/sciadv.aaw1668 (cited on pages 7, 8, 193)

Talbert, P. B., & Henikoff, S. (2010). Centromeres convert but don't cross. *PLoS Biology*, *8*(3), 1–5. https://doi.org/10.1371/journal.pbio.1000326 (cited on page 19)

Tatsura, H., Nagao, H., Tamada, A., Sasaki, S., Kohri, K., & Mori, K. (2001). Developing germ cells in mouse testis express pheromone receptors. *FEBS Letters*, *488*(3), 139–144. https://doi.org/10.1016/S0014-5793(00)02411-X (cited on page 200)

Terrenoire, E., McRonald, F., Halsall, J. a., Page, P., Illingworth, R. S., Taylor, a. M. R., Davison, V., O'Neill, L. P., & Turner, B. M. (2010). Immunostaining of modified histones defines high-level features of the human metaphase epigenome. *Genome Biology*, *11*(11), R110. https://doi.org/10.1186/gb-2010-11-11-r110 (cited on page 7)

Thomas, J. H., Emerson, R. O., & Shendure, J. (2009). Extraordinary molecular evolution in the PRDM9 fertility gene. *PloS one*, *4*(12), e8505. https://doi.org/10.1371/journal.pone.0008505 (cited on page 21)

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. https://doi.org/10.1093/bib/bbs017 (cited on page 52)

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K.,

Diegel, M., Dunn, D., Ebersol, A. K., . . . Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, *489*(7414), 75–82. https://doi.org/10.1038/nature11232 (cited on page 195)

Tiang, C.-L., He, Y., & Pawlowski, W. P. (2012). Chromosome Organization and Dynamics during Interphase, Mitosis, and Meiosis in Plants. *Plant Physiology*, *158*(1), 26 LP –34. https://doi.org/10.1104/pp.111.187161 (cited on page 4)

Tiemann-Boege, I., Calabrese, P., Cochran, D. M., Sokol, R., & Arnheim, N. (2006). High-resolution recombination patterns in a region of human chromosome 21 measured by sperm typing. *PLoS Genetics*, *2*(5), 682–692. https://doi.org/10.1371/journal.pgen.0020070 (cited on page 48)

Tiemann-Boege, I., Schwarz, T., Striedner, Y., & Heissl, A. (2017). The consequences of sequence erosion in the evolution of recombination hotspots. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *372*(1736), 20160462. https://doi.org/10.1098/rstb.2016.0462 (cited on pages 21, 203)

Torgasheva, A. A., Rubtsov, N. B., & Borodin, P. M. (2013). Recombination and synaptic adjustment in oocytes of mice heterozygous for a large paracentric inversion. *Chromosome Research*, *21*(1), 37–48. https://doi.org/10.1007/s10577-012-9336-6 (cited on pages 23, 120)

Trovero, M. F., Rodríguez-Casuriaga, R., Romeo, C., Santiñaque, F. F., François, M., Folle, G. A., Benavente, R., Sotelo-Silveira, J. R., & Geisinger, A. (2020). Revealing stage-specific expression patterns of long noncoding RNAs along mouse spermatogenesis. *RNA Biology*, *17*(3), 350–365. https://doi.org/10.1080/15476286.2019.1700332 (cited on page 196)

Tseng, Y. T. Y.-T., Liao, H.-F. H. F., Yu, C. Y. C.-Y., Mo, C. F. C.-F., & Lin, S.-P. S. P. S.-P. (2015). Epigenetic factors in the regulation of prospermatogonia and spermatogonial stem cells. *Reproduction*, *150*(3), R77–R91. https://doi.org/10.1530/REP-14-0679 (cited on pages 14, 194)

Turner, J. M. A. (2007). Meiotic sex chromosome inactivation. *Development*, *134*(10), 1823–1831. https://doi.org/10.1242/dev.000018 (cited on page 198)

Turner, J. M. A., Mahadevaiah, S. K., Ellis, P. J. I., Mitchell, M. J., & Burgoyne, P. S. (2006). Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids. *Developmental Cell*, *10*(4), 521–529. https://doi.org/10.1016/j.devcel.2006.02.009 (cited on pages 16, 198)

Turner, J. M. (2015). Meiotic Silencing in Mammals. *Annual Review of Genetics*, *49*(1), 395–412. https://doi.org/10.1146/annurev-genet-112414-055145 (cited on pages 16, 17, 198)

Turner, L. M., White, M. a., Tautz, D., & Payseur, B. a. (2014). Genomic Networks of Hybrid Sterility. *PLoS Genetics*, *10*(2), 18–22. https://doi.org/10.1371/journal.pgen.1004162 (cited on page 22)

Ullastres, A., Farré, M., Capilla, L., & Ruiz-Herrera, A. (2014). Unraveling the effect of genomic structural changes in the rhesus macaque - implications for the adaptive role of inversions. *BMC Genomics*, *15*(1), 530. https://doi.org/10.1186/1471-2164-15-530 (cited on page 120)

Vallot, A., & Tachibana, K. (2020). The emergence of genome architecture and zygotic genome activation. *Current Opinion in Cell Biology*, *64*, 50–57. https://doi.org/10.1016/j.ceb.2020.02.002 (cited on pages 197, 198)

Vanderhaeghen, P., Schurmans, S., Vassart, G., & Parmentier, M. (1997). Specific repertoire of olfactory receptor genes in the male germ cells of several mammalian species. *Genomics*, *39*(3), 239–246. https://doi.org/10.1006/geno.1996.4490 (cited on page 200)

van Steensel, B., & Furlong, E. E. (2019). The role of transcription in shaping the spatial organization of the genome. *Nature Reviews Molecular Cell Biology*, *20*(6), 327–337. https://doi.org/10.1038/s41580-019-0114-6 (cited on pages 194–197)

Vara, C., Capilla, L., Ferretti, L., Ledda, A., Sánchez-Guillén, R. A., Gabriel, S. I., Albert-Lizandra, G., Florit-Sabater, B., Bello-Rodríguez, J., Ventura, J., Searle, J. B., Mathias, M. L., & Ruiz-Herrera, A. (2019a). PRDM9 diversity at fine geographical scale reveals contrasting evolutionary patterns and functional constraints in natural populations of house mice (B. Chang, Ed.). *Molecular Biology and Evolution*, *36*(8), 1686–1700. https://doi.org/10.1093/molbev/msz091 (cited on pages 33, 120, 135, 136)

Vara, C., Paytuví-Gallart, A., Cuartero, Y., Le Dily, F., Garcia, F., Salvà-Castro, J., Gómez-H, L., Julià, E., Moutinho, C., Aiese Cigliano, R., Sanseverino, W., Fornas, O., Pendás, A. M., Heyn, H., Waters, P. D., Marti-Renom, M. A., & Ruiz-Herrera, A. (2019b). Three-dimensional genomic structure and cohesin

occupancy correlate with transcriptional activity during spermatogenesis. *Cell Reports*, *28*(2), 352–367. https://doi.org/10.1016/j.celrep.2019.06.037 (cited on pages 44, 119, 120, 125, 126, 129, 131, 134, 137–139)

Vernet, N., Mahadevaiah, S. K., Ellis, P. J. I., de Rooij, D. G., & Burgoyne, P. S. (2012). Spermatid development in XO male mice with varying Y chromosome short-arm gene content: evidence for a Y gene controlling the initiation of sperm morphogenesis. *Reproduction*, *144*(4), 433–445. https://doi.org/10.1530/REP-12-0158 (cited on page 199)

Vian, L., Pkowska, A., Rao, S. S., Kieffer-Kwon, K.-R., Jung, S., Baranello, L., Huang, S.-C., El Khattabi, L., Dose, M., Pruett, N., Sanborn, A. L., Canela, A., Maman, Y., Oksanen, A., Resch, W., Li, X., Lee, B., Kovalchuk, A. L., Tang, Z., . . . Casellas, R. (2018). The energetics and physiological impact of cohesin extrusion. *Cell*, *173*(5), 1165–1178. https://doi.org/10.1016/j.cell.2018.03.072 (cited on page 129)

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A., & Hadjur, S. (2015). Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Reports*, *10*(8), 1297–1309. https://doi.org/10.1016/j.celrep.2015.02.004 (cited on pages 5, 8)

Walker, W. H. (2011). Testosterone signaling and the regulation of spermatogenesis. *Spermatogenesis*, *1*(2), 116–120. https://doi.org/10.4161/spmg.1.2.16956 (cited on page 13)

Wallace, B., Searle, J., & Everett, C. (1992). Male meiosis and gametogenesis in wild house mice (Mus musculus domesticus) from a chromosomal hybrid zone; a comparison between "simple" Robertsonian heterozygotes and homozygotes. *Cytogenetic and Genome Research*, *61*(3), 211–220. https://doi.org/10.1159/000133410 (cited on page 202)

Wallace, B., Searle, J., & Everett, C. (2002). The effect of multiple simple Robertsonian heterozygosity on chromosome pairing and fertility of wild-stock house mice (Mus musculus domesticus). *Cytogenetic and Genome Research*, *96*(1-4), 276–286. https://doi.org/10.1159/000063054 (cited on pages 25, 200, 202)

Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., Lanz, C., & Weigel, D. (2015a). Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. *Genome Research*, *25*(2), 246–256. https://doi.org/10.1101/gr.170332.113 (cited on page 6)

Wang, J., Meng, X., Chen, H., Yuan, C., Li, X., Zhou, Y., & Chen, M. (2016a). Exploring the mechanisms of genome-wide long-range interactions: interpreting chromosome organization. *Briefings in Functional Genomics*, *15*(5), 385–395. https://doi.org/10.1093/bfgp/elv062 (cited on pages 7, 197)

Wang, S., Veller, C., Sun, F., Ruiz-Herrera, A., Shang, Y., Liu, H., Zickler, D., Chen, Z., Kleckner, N., & Zhang, L. (2019a). Per-nucleus crossover covariation and implications for evolution. *Cell*, *177*(2), 326–338. https://doi.org/10.1016/j.cell.2019.02.021 (cited on pages 17–19, 120, 122, 123, 133, 196, 200, 201)

Wang, S., Zickler, D., Kleckner, N., & Zhang, L. (2015b). Meiotic crossover patterns: Obligatory crossover, interference and homeostasis in a single process. *Cell Cycle*, *14*(3), 305–314. https://doi.org/10.4161/15384101.2014.991185 (cited on pages 120, 123, 133, 200, 201)

Wang, S., Su, J.-H. J.-h. J.-H., Beliveau, B. J., Bintu, B., Moffitt, J. R., Wu, C.-t. C.-t. C.-t., & Zhuang, X. (2016b). Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, *353*(6299), 598–602. https://doi.org/10.1126/science.aaf8084 (cited on page 197)

Wang, Y., Wang, H., Zhang, Y., Du, Z., Si, W., Fan, S., Qin, D., Wang, M., Duan, Y., Li, L., Jiao, Y., Li, Y., Wang, Q., Shi, Q., Wu, X., & Xie, W. (2019b). Reprogramming of Meiotic Chromatin Architecture during Spermatogenesis. *Molecular Cell*, *73*(3), 547–561. https://doi.org/10.1016/j.molcel.2018.11.019 (cited on pages 119, 194, 195)

Wang, Y., Lu, P., Jiang, L., Wu, B., & Zhou, B. (2019c). Control of sinus venous valve and sinoatrial node development by endocardial NOTCH1. *Cardiovascular Research*. https://doi.org/10.1093/cvr/cvz249 (cited on page 197)

Ward, W. S., & Coffey, D. S. (1991). DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biology of Reproduction*, *44*(4), 569–574. https://doi.org/10.1095/biolreprod44.4.569 (cited on page 16)

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0 (cited on page 3)

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370 (cited on page 137).

Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends in Ecology and Evolution*, *33*(6), 427–440. https://doi.org/10.1016/j.tree.2018.04.002 (cited on page 22)

White, M. (1969). Chromosomal rearrangements and speciation. *Annual Review of Genetics*, *3*, 75–98 (cited on page 22).

White, M. (1973). *Animal Cytology and Evolution*. Cambridge University Press. (Cited on page 22).

White, M. (1978). *Modes of Speciation* (W. Freeman, Ed.). (Cited on pages 22, 125).

White, T. A., & Searle, J. B. (2007). Genetic diversity and population size: island populations of the common shrew, Sorex araneus. *Molecular Ecology*, *16*(10), 2005–2016. https://doi.org/10.1111/j.1365-294X.2007.03296.x (cited on page 23)

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. (Cited on page 52).

Wilch, E. S., & Morton, C. C. (2018). Historical and Clinical Perspectives on Chromosomal Translocations. In Y. Zhang (Ed.), *Chromosome translocation* (pp. 1–14). Singapore, Springer Singapore. https://doi.org/10.1007/978-981-13-0593-1{_}1. (Cited on pages 22, 25, 119)

Winkel, K., Alsheimer, M., Ollinger, R., & Benavente, R. (2009). Protein SYCP2 provides a link between transverse filaments and lateral elements of mammalian synaptonemal complexes. *Chromosoma*, *118*(2), 259–267. https://doi.org/10.1007/s00412-008-0194-0 (cited on page 15)

Wischnitzer, S. (1973). The Submicroscopic Morphology of the Interphase Nucleus. In G. H. Bourne & J. F. B. T. I. R. o. C. Danielli (Eds.). Academic Press. https://doi.org/10.1016/S0074-7696(08)61933-6. (Cited on page 3)

Wistuba, J., Stukenborg, J., & Luetjens, C. (2007). Mammalian Spermatogenesis. In *Functional development and embryology* (pp. 99–117). https://doi.org/10.1016/S1067-5701(96)80011-8. (Cited on pages 13, 14, 194)

Wu, H., Mathioudakis, N., Diagouraga, B., Dong, A., Dombrovski, L., Baudat, F., Cusack, S., de Massy, B., & Kadlec, J. (2013). Molecular basis for the regulation of the H3K4 methyltransferase activity of PRDM9. *Cell reports*, *5*(1), 13–20. https://doi.org/10.1016/j.celrep.2013.08.035 (cited on page 20)

Xu, H., Beasley, M. D., Warren, W. D., van der Horst, G. T. J., & McKay, M. J. (2005). Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis. *Developmental Cell*, *8*(6), 949–961. https://doi.org/10.1016/j.devcel.2005.03.018 (cited on page 17)

Yamada, S., Kim, S., Tischfield, S. E., Jasin, M., Lange, J., & Keeney, S. (2017). Genomic and chromatin features shaping meiotic double-strand break formation and repair in mice. *Cell Cycle*, *16*(20), 1870–1884. https://doi.org/10.1080/15384101.2017.1361065 (cited on page 22)

Yang, T., Zhang, F., Yardimci, G. G., Hardison, R. C., Stafford Noble, W., Yue, F., & Li, Q. (2017). HiCRep : assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Research*, *27*, 1939–1949 (cited on page 52).

Yang, Y., Zhang, Y., Ren, B., Dixon, J. R., & Ma, J. (2019). Comparing 3D Genome Organization in Multiple Species Using Phylo-HMRF. *Cell Systems*, *8*(6), 494–505. https://doi.org/10.1016/j.cels.2019.05.011 (cited on page 5)

Yoshida, K., Kondoh, G., Matsuda, Y., Habu, T., Nishimune, Y., & Morita, T. (1998). The mouse RecA-like gene Dmc1 is required for homologous chromosome synapsis during meiosis. *Mol Cell*, *1*(5), 707–718. https://doi.org/S1097-2765(00)80070-2[pii] (cited on pages 18, 62)

Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, *31*(14), 2382–2383. https://doi.org/10.1093/bioinformatics/btv145 (cited on page 52)

Zalensky, A. O., Allen, M. J., Kobayashi, A., Zalenskaya, I. A., Balhorn, R., & Bradbury, E. M. (1995). Well-defined genome architecture in the human sperm nucleus. *Chromosoma*, *103*(9), 577–590. https://doi.org/10.1007/BF00357684 (cited on page 197)

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., To, J. K., . . . Flicek, P. (2017). Ensembl 2018. *Nucleic Acids Research*, *46*(D1), D754–D761. https://doi.org/10.1093/nar/gkx1098 (cited on page 42)

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E. G., Blüthgen, N., Stadler, M., Tiana, G., & Giorgetti, L. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Research*, *27*(3), 479–490. https://doi.org/10.1101/gr.212803.116 (cited on page 195)

Zhang, Y., McCord, R. P., Ho, Y. J., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., Becker, M. S., Alt, F. W., & Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, *148*(5), 908–921. https://doi.org/10.1016/j.cell.2012.02.002 (cited on pages 5, 6)

Zhu, L. J., Gazin, C., Lawson, N. D., Pagès, H., Lin, S. M., Lapointe, D. S., & Green, M. R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, *11*(1), 237. https://doi.org/10.1186/1471-2105-11-237 (cited on page 52)

Zickler, D., & Kleckner, N. (1999). Meiotic Chromosomes: Integrating Structure and Function. *Annual Review of Genetics*, *33*(1), 603–754. https://doi.org/0066-4197 (cited on pages 15, 17, 19, 133, 194)

Zickler, D., & Kleckner, N. (2015). Recombination, Pairing, and Synapsis of Homologs during Meiosis. *Cold Spring Harbor Perspectives in Biology*, *7*(6), a016626. https://doi.org/10.1101/cshperspect.a016626 (cited on pages 15, 17, 19, 133, 200, 201)

Zickler, D., & Kleckner, N. (2016). A few of our favorite things: Pairing, the bouquet, crossover interference and evolution of meiosis. *Seminars in Cell and Developmental Biology*, *54*, 135–148. https://doi.org/10.1016/j.semcdb.2016.02.024 (cited on pages 17, 19)

Zlatanova, J., & Caiafa, P. (2009). CTCF and its protein partners: divide and rule? *Journal of Cell Science*, *122*(9), 1275–1284. https://doi.org/10.1242/jcs.039990 (cited on page 9)

Acknowledgements

En primer lugar, querría agradecer profundamente a la Dra. Aurora Ruiz-Herrera, sin la cual esta tesis no hubiese sido posible. Aurora, gracias por darme la oportunidad de hacer ciencia y ver "el diamante en bruto" que hay en mí, por la paciencia y tu dedicación, por enseñarme que nunca es suficiente, que es "cuestión de organizarse", por tirar del carro mientras me enseñabas a llevar las riendas. No sólo has sido mi mentora y mi supervisora, sino que también me has moldeado como científica y como persona, enseñándome a "investigar". Aunque nuestros "registros" sean distintos y muchas veces compliquen bastante nuestro mutuo entendimiento, has sacado la mejor versión de mi. De nuevo, gracias. Espero que este sea el principio de una *life-long* collaboration.

También querría agradecer efusivamente a mi familia por su apoyo incondicional. A mis padres, por darme todos los recursos que estaban en sus manos para tener todas las oportunidades posibles, por ser mi equipo, por transmitirme su pasión por aprender y porque sin ellos no sería la persona que soy ahora. Mamá, gracias por tu apoyo incondicional, por empujarme siempre hacia arriba y enseñarme a luchar contra el conformismo. Gracias por ser mi mayor crítica y inspirarme a mejorar siempre, por aportarme tu perspectiva "más humana" y hacerme ver las cosas de otra manera. Papa, gràcies per transmetre'm la teva passió per la biologia i a perseguir la curiositat. Gràcies per donar-me les eines per enfrontar-me al món, fer-me veure el valor de l'esforç i donar-me alè quan ja no me'n quedava. Gràcies també per obrir-me els ulls i fomentar el meu esperit crític. También quería agradecer especialmente a mis tías Fini y Meli por transmitirme su pasión por la biología, aunque como ya se vaticinaba en nuestras excursiones por bosques asturianos, he acabado muy alejada de la botánica.

A mis compañeros de laboratorio, Ana, Andros,Yan, Lucía y Laia. A Ana y Andros, gracias por las infinitas comidas en las que se forjó nuestra amistad, nuestras discusiones y nuestros momentos eurovisivos. Ana, como bien sabes somos personas muy diferentes, pero nuestras aficiones comunes nos han permitido disfrutar de muchos momentos que me han ayudado a sobrellevar los momentos más bajos del doctorado. Tenemos que volver a un espectáculo de RuPaul y conseguiremos ir a un concierto de Panic at the Disco! Andros, no hay persona con la que haya discutido más y que me genere tan altos grados de desesperación. Sin embargo, también eres la persona con la que más me río y cuando tienes tus ataques de genialidad, me aportas nuevas ideas que nunca se me hubieran ocurrido. Gracias por ser mi amigo a tu manera y por compartir este camino conmigo. Yan, I am also grateful to you for all the moments we have shared in these years, I have learnt a lot from you and I hope we keep in touch! También querría agradecer a Laia y Lucía por su ayuda en el lab en mi última fase de la tesis, y por compartir una nueva pasión por los geckos.

Quiero agradecer también el apoyo de mis amigas, las "chochis". Es poco común mantener una amistad desde el parvulario y que dure casi 30 años (¡y los que nos quedan!). Gracias por intentar entender mi camino en ciencia y poner la mejor cara posible cuando os contaba de que iba mi tesis, gracias por todas las cervezas de los viernes y compartir esas bravas sabor pescado, por todos los bailes y los viajes. Gracias Corti, Larry, Helena, Pati, Gemma, Irene, Núria, Barbi y Bea por ser todas distintas, y por aportarme cada una vuestra visión del mundo. No hubiera podido sobrevivir a la última etapa de la tesis sin los "Super Powers" del SAF. Muchas gracias por todos los *combats* que hemos disfrutado juntos, los *pumps* que hemos sufrido y por todas las cervezas que hemos compartido en la vila. Quiero agradecer especialmente a Bea, Andrés y Alex por los momentos que hemos compartido y por seguir compartiendo el camino (y birras/radlers de limón!). Gracias también a Llorenç, Esther y Víctor por seguir formando parte de mi vida, chistorras gitanas locas desde la uni. Esther, gracias por los tips para usar el LaTeX!

Quisiera también agradecer al Dr. Alberto M. Pendás por recibirme con los brazos abiertos en su laboratorio y por las discusiones sobre el trabajo a lo largo del proyecto. Also, thanks to Dr. Irene Tiemann-Boege for welcoming and supervising me during my visit to her lab in Linz and for her thoughtful insights on my research. También querría dar las gracias al Dr. Jacint Ventura por guiarme en mis primeros pasos en el trabajo de campo y por su ayuda en el proyecto. Gracias también a la Dra. Yasmina Cuartero y al Dr. François Ledily por abrirme las puertas del mundo de la genómica y a dar mis primeros pasos en técnicas de secuenciación

masiva. También agradecer al Dr. Marc Martí-Renom por su input durante el proyecto y sin el cual este no habría sido posible. Gracias también a mi compañero, el Dr. Andreu Paytuví, este es un proyecto que iniciamos conjuntamente y que sin él, no hubiese sido posible. No querría olvidarme del *Servei d'Estabulari* de la UAB, gracias por enseñarme lo que es la paciencia.

Quiero hacer una mención especial a la unidad de citometría del PRBB, al jefe de la unidad Oscar Fornas, gràcies per ajudar-nos a desenvolupar el projecte, per les múltiples converses i brain storming de com millorar els resultats (i alguna conversa sobre pokemons també). A Eva Julià, per la teva ajuda i els teus consells, a Erika por su paciencia en explicarme las "intríngulis" de la citometría y nuestro momentos musicales y a Álex, gràcies també per la teva ajuda. Quería hacer una mención al programa de la cadena SER "La Vida Moderna", gracias por alegrarme los incontables trayectos entre la UAB y el PRBB para hacer los experimentos de FACS.

Finally, I would like to mention that this thesis was finished amidst the covid-19 global pandemic crisis, so thank you SARS-CoV-2 for guest-starring the last "episode" of my thesis , forcing the lockdown on me so I could completely focus on becoming a doctor.