Text in context: Chromatin effects in gene regulation

Marc Corrales Berjano

TESI DOCTORAL UPF / 2017

DIRECTOR DE LA TESI

Dr. Guillaume Filion

DEPARTAMENT

Gene Regulation, Stem cells and Cancer

Genome architecture

Center for Genomic Regulation (CRG)

a mi familia,

Aknowledgements

I want to thank the members of the doctoral board, present and substitutes, for accepting to analize my work and defend my thesis.

I also want to thank my thesis commitee Dr. Miguel Beato, Dr. Jordi Garcia-Ojalvo and Dr. Tanya Vavouri for guidance and valuable scientific imput whenever I needed it along these years.

I want to thank my mentor Dr. Guillaume Filion, not only for giving me the opportunity to carry this work in a wonderful environement, but for teaching me. No matter the tigh schedule he has allowed me to sit by his side (literally) reviewing a script at the computer, calculating the probaility of plasmid recircularization or deriving a statistical formula at the white board but also helping me to prepare and electroporate a bacterial library in what seemed a contraction of space-time.

I want to thank the community of the CRG, always ready to discuss science and life in the corridors, the data seminars, symposiums, the beer sessions or following a mail asking for advice or materials. People at the administration helped enormously to navigate the tortuous bureaucratic world, maximizing the time one can dedicate toscience. ¡Dónde estaría yo sin tu puerta siempre abierta, Imma, madre en funciones de todo phD student!

I want to thank the CRG for funding my phD and for constituting what I envisage as the perfect environment to learn to be a scientist.

I want to thank all the members past and present of the “Filionetis” family. It is a real pleasure to work by your side and you manage to make the lab home, human and scientifically speaking. All words used to capture the impact that you had in my growth as a scientist, colleague and individual will fall too short.

Quiero agradecer a mi familia ese apoyo incondicional que solo puede brindar una familia, sin abandonar el sentido común para tratar de enseñarme a ser mi mejor versión como profesional y como persona. Os lo debo todo y os quiero.

Luis, has sido amigo e inspiración desde la carrera. En lo intelectual y en lo personal. Se que aunque nos separe tiempo y geografía siempre que vuelva a sentarme delante tuyo reaparecerá esa complicidad que discurre entre palabras y retoques de bigote.

Marta, me es imposible escribir sobre todo lo que me has dado estos años, en cada uno de los aspectos a los que solo se llega y aprecia con la intimidad. Gracias por entrar en mi vida, por contribuir a esto que es Marc y haberme acompañado pacientemente en este viaje.

Ruggero! Otra clase de castellano: Nunca es tarde si la dicha es buena. Aprecio muchísimo que llegaras y te instalaras tan rápido en mi vida. La verdad es que es fácil cuando uno aúna las cualidades intelectuales y humanas que te caracterizan. Gracias por ser mi amigo, por los miles de consejos y conversaciones, por corregirme la tesis y por las canciones. Sospecho sobre la coincidencia de que vivas en el 112.

Alba, entrando y saliendo, tu sempiterna sonrisa, sensibilidad y buen humor me han salvado de mas de un descalabro este difícil verano, amiga mía. Se que has llegado para quedarte y que vas a seguir enriqueciéndome intelectual y personalmente.

Michael, thanks for being always there. To discuss science or life over a beer (corrector suggests plural) and for being my phD mate from the beginning to the end. I really hope this is the start of a journey! (A life journey).

Arantxa. ¿Que decir? Si este trabajo es medio tuyo, y no me refiero solo al articulo. Gracias por enseñarme tanto, en el laboratorio y fuera de el. Enumerarlo seria una tarea interminable.
Eres un ejemplo de inteligencia, sencillez y amistad verdadera del que espero poder seguir disfrutando con los años.
Carlitos! Cuando te entre la tontuna púber (como nos entra a todos) léete este párrafo y convéncete de que tu madre ¡es la bomba! Una luchadora sin igual, créeme.

Abuelo, eres el espejo en el que quisiera mirarme con los años. Avivaste el fuego de mi curiosidad y alimentaste mi intención de perseguir este camino. Me encantaría haber podido leerte esta tesis antes de tu marcha, pero me queda el pequeño placer de dedicártela. Va por ti!

Abstract

This thesis reports the study of chromatin composition and conformation on the expression of integrated reporters at thousands of genomic locations in the *Drosophila* genome. We have adapted and improved a technology (Thousands of Reporters Integrated in Parallel TRIP) to randomly integrate barcoded reporters allowing us to measure the context effects on transcription at ~80.000 different loci. We have focused on housekeeping promoter-reporters due to their relative autonomy from distal regulatory elements.
Taking advantage of published genome-wide localization maps of chromatin protein and histone marks, together with the three-dimensional genome structure of the *Drosophila* Kc167 cell line, we have been able to computationally extract the features that best predict the expression of the integrated reporters. Centromeric heterochromatin is highly silencing but position effects were also observed along chromosome arms, away from heterochromatin. Chromatin states such as Black and Blue (Polycomb, H3K27me3) were found to be refractory to expression while Green (HP1, H3K9me) was found to be permissive or refractory depending on the location. Yellow (MRG15, H3K4me) chromatin was the most permissive while Red (Brm, H3K4me) could also be repressive or activating depending on the integration point.
Surprisingly we discovered that the housekeeping reporters are maximally expressed when they land on loci engaging in contacts with promoters and terminators of active genes.
The low dependence on enhancers confirms the hypothesis that the requisites for developmental regulation are different that for broadly expressed genes.
Moreover our results bring experimental evidence to the observation that housekeeping genes tend to cluster during evolution along the chromatin fiber, providing an explanation to the spatial contacts among these clusters observed in Hi-C experiments.

Resumen

Esta tesis recoge los resultados del estudio del efecto de la composición y conformación de la cromatina en la expresión génica mediante la integración de miles de reporteros en el genoma de Drosophila. A tal efecto hemos adaptado y mejorado una técnica (Thousands of Reporters Integrated in Parallel, TRIP) permitiendo la integración aleatoria de genes reportero marcados (barcoded) con el fin de medir su expresión dependiente de contexto en ~80.000 loci distintos. Gracias a los numerosos mapas de ocupación a escala genómica de proteínas asociadas a cromatina y marcas de histonas, así como la estructura tridimensional del genoma en la linea celular utilizada Drosophila Kc167, hemos podido extraer las variables que mejor explican la expresión de los genes reportero integrados.
La Heterocromatina pericentromérica demostró su capacidad represora aunque los efectos de posición también pudieron observarse en los brazos cromosómicos, lejos de dicha cromatina.
Estados de la cromatina como Black y Blue (Polycomb, H3K27me3) se mostraron refractarios a la expresión mientras que la de tipo Green (HP1, H3K9me) demostró tener efecto ambivalente en función del lugar de integración. La cromatina Yellow (MRG15, H3K4me) mostró ser la mas permisiva mientras que la de tipo Red (Brm, H3K4me) evidenció el mismo carácter ambivalente en función del punto de integración.
Sorprendentemente descubrimos que los reporteros housekeeping se expresan de forma óptima cuando se integran en loci contactando promotores y terminadores de genes activos.
La escasa dependencia de enhancers confirma la hipótesis según la cual los requisitos para la regulación de genes del desarrollo difieren de los utilizados por genes de expresión ubicua.
Por ultimo nuestros resultados brindan evidencia experimental a la observación de la agrupación de genes housekeeping a lo largo de la fibra de ADN durante la evolución.
De mismo modo aportan una explicación para el elevado numero de contactos que muestran dichas agrupaciones en experimentos Hi-C.

Prologue

The time of starting my phD was a very exciting one for someone interested in gene regulation, i.e understanding how a cell implements “decision making” to generate different cell types, to temporally express genes allowing its development and to respond to intracellular and extracellular cues in order to maintain homeostasis. Several authors had just started to publish the results of the first genome-wide profiling experiments of chromatin proteins, histone-marks and transcription factor occupancy.
The analysis of these results showed that even when using hundreds of factors (chromatin proteins, histone marks, accessibility measures, transcription factor occupancy) few states of chromatin were recovered. This meant that, on the one hand heterochromatin and euchromatin concepts were oversimplifications of a more complex system of gene control. On the other hand the genome-wide segmentation in homogeneous states pointed to general but different mechanisms to achieve this regulation. Moreover half of the genome was devoid of the vast majority of these features showing that the principal mechanism to achieve gene silencing was not understood. When faced with the opportunity of testing the effects of these newly discovered chromatin types with a high-throughput technique avoiding the “locus of interest” approach, I was sold.
I present here the results of the effects of chromatin composition and structure in the expression of ectopically integrated housekeeping genes in Drosophila melanogaster. Adapting and improving the TRIP technology we have been able to assess the context effect on gene expression at unprecedented throughput (~80.000 loci).

Table of contents

1. Genetics: Understanding phenotypic encoding and transmission

The concept of heredity, i.e how phenotypic traits are mixed and transmitted has occupied the mind of philosophers of antiquity, farmers in the quest of obtaining the best crops and animal breeds but also in the general population with the observation that children resemble their parents. In the course of the XXth century not only we have been able to acknowledge that phenotypic information is stored in the form of discrete units, the genes, but we have been able to understand how these genes are materially stored in a long molecule of deoxyribonucleic acid (DNA) separated in chromosomes with a characterized structure.

We have learnt the physical rules allowing this molecule to be copied and transmitted preserving the information and at the same time allowing the generation of different combinations of the genes contained to be transgenerationally transmitted, generating organismal diversity and allowing adaptation.

Finally we have been able to grasp the general picture of the mechanisms responsible for controlling the activities of the proteins encoded by the genes in time and space, gene regulation.

1.1 The gene: discrete unit of phenotypic inheritance

Mendel had already realized in the mid-late XIXth century the existence of discrete inheritable units responsible of the observable phenotypes in his pea plant crossing experiments (Mendel, 1865). Not only his findings refuted the idea of blending inheritance (proposing that an inherited trait was a random mixture of the parental ones) but also foreshadowed the difference between genotype and phenotype.
Largely unnoticed, the ideas of Mendel were rediscovered and

confirmed by the independent work of three botanists:
Hugo De Vries (de Vries, 1900), Carl Correns (Correns, 1924) and Erich von Tschermak (Tschermak, 1900) in the early 1900s. The term "gene" was first used by the Danish geneticist Wilhelm Johannsen (Johannsen, 1905), but it was the research in genetics during the first half of the XX century that materialized the abstract concept of the gene.

1.2 Chromosomal theory of inheritance

Walter Sutton described in 1903 how Mendel's laws of inheritance and dominance were in accordance with the patterns of chromosome segregation observed in his cytological experiments of mitosis and meiosis in Brachystola (Sutton, 1903). Bovery published his studies in sea urchin, showing that chromosomes were needed for its correct development (Boveri, 1904).
Thomas Hunt Morgan and his group brought genetic evidence confirming the Sutton-Boveri theory of "chromosomes as bearers of the hereditary material" with Drosophila studies on sex-inheritance (Morgan, 1915).

Alfred Sturtevant demonstrated that genes were linearly ordered along chromosomes after realizing that the distance among them was proportional to the frequency of recombination, which allowed him to construct the first chromosome maps (Sturtevant, 1913).

1.3 One gene, one enzyme

The first systematic description of gene function came from the work of George Beadle and Edward Tatum in Neurospora. Coupling X-Ray mutagenesis and strictly defined synthetic media, they were able to test their idea: "It is entirely tenable to suppose that these genes which are themselves a part of the system, control or regulate specific reactions in the system either by acting directly as enzymes or by determining the specificities of enzymes"

(Horowitz, Bonner, Mitchell, Tatum, & Beadle, 1945).
In short, they produced mutant spores that were tested for their ability to grow in complete medium but not in minimal medium. They subsequently screened a battery of single compounds supplementing the minimal medium to discover the putative mutated single enzyme. With this experimental approach, they deduced the ordered sequence of biochemical reactions of full metabolic pathways such as the biosynthesis of arginine (Beadle & Tatum, 1941; Sturtevant, 1913).

Although this theory is now known to be an oversimplification, it represents a more detailed approach to study the relationship between mutation effects and phenotype, especially when compared to complex morphological changes such as the ones observed in Drosophila.

1.4 The genes are made of DNA

In 1928 Frederick Griffith had shown that injecting a mouse with heat-inactivated virulent Smooth Pneumococci in conjunction with non-virulent and alive Rough Pneumococci was lethal. Shockingly he isolated live Smooth Pneumococci in the blood of the dead animals.
He concluded that non virulent Rough Pneumococci had been transformed into virulent Smooth strains. He hypothesized that some heat-resistant component responsible for virulence had been transferred to the non-virulent bacteria. This early example of genetic information transfer would prove to be a good system to ascertain the chemical nature of the heredity determinant (Griffith, 1928).

In 1944 Oswald Avery, Colin MacLeod, and Maclyn McCarty indirectly showed that the genes were made of DNA (and not proteins) and that the DNA was the "transforming principle" (Avery, Macleod, & McCarty, 1944). The experiments showed that:

"The data obtained by chemical, enzymatic, and serological analyses together with the results of preliminary studies by electrophoresis, ultracentrifugation, and ultraviolet spectroscopy indicate that, within the limits of the methods, the active fraction contains no demonstrable protein, unbound lipid, or serologically reactive polysaccharide and consists principally, if not solely, of a highly polymerized, viscous form of desoxyribonucleic acid" (Avery et al., 1944).

A more direct evidence that DNA is the genetic material came with the "blender experiment" of Alfred Hershey and Martha Chase. Using radiolabeled Sulfur (present only in proteins) or Nitrogen (present in DNA) they showed that the material injected by bacteriophages during bacterial infection was the Nitrogen-radiolabelled DNA. Moreover it was detected in the viral progeny after lysis of the infected host (Hershey, 1952).

1.5 The double helix and the code

One year before the publication of the results of Hershey and Chase, James Watson and Francis Crick started working on the structure of the DNA. Two years later they published the correct structure of the genetic fiber. Their model relied on two important pieces of information: 1) An X-ray diffraction image (Photograph 51) from Rosalind Franklin and Raymond Gosling (shown by Maurice Wilkins without Franklin's approval or knowledge) and 2) the Erwin Chargaff rule: "the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid" (Chargaff, 1950).

With this information they worked on building a physical model incorporating the structures of the nucleobases, the chemical bond distances and the helical configuration demonstrated by crystallography.
The resulting structure consisted of two antiparallel nucleotide

polymers forming a dextrogyre double helix stabilized by the hydrogen bonds between the nucleobases Adenine and Thymine or Guanine and Cytosine. The base complementarity prompted them to suggest "a possible copying mechanism for the genetic material" (Watson & Crick, 1974).

Having the chemical and structural properties of the molecule sparked the research on DNA physiology.

The years to come would lead to the discovery of the enzymes responsible for the duplication of DNA, the bacterial DNA polymerases I and II of E. Coli (T. Kornberg & Gefter, 1970; Lehman, Bessman, Simms, & Kornberg, 1958; Moses & Richardson, 1970).

Matthew Meselson and Franklin Stahl demonstrated that DNA replication is semiconservative in 1958. Each of the strands of the double helix serves as a template copied by DNA polymerase. This results into two new double DNA helices composed by the old template strand and the newly synthesized one (Meselson & Stahl, 1958).

The genetic code directing protein translation was deciphered biochemically, with the help of improved bacterial whole cell extracts with the ability to produce detectable protein synthesis from supplemented Ribonucleic acids (RNAs) and thanks to chemically synthesized sequence specific templates.

Heinrich Matthaei and Marshall Nirenberg provided the first piece of the puzzle. Using their newly developed bacterial whole cell extract, they showed that a poly-uridine tract was translated into radiolabelled phenylalanine peptides (Nirenberg & Matthaei, 1961).

Severo Ochoa had discovered what was thought to be the bacterial DNA dependent RNA polymerase (Grunberg-Manago, Ortiz, & Ochoa, 1956) (now called polynucleotide phosphorylase). The

enzyme provided Ochoa's laboratory with heterogeneous synthetic RNAs expanding early Nirenberg & Matthaei experiments showing that poly UC produced phenylalanine and serine, and poly UA produced of phenylalanine and tyrosine peptides (Lengyel, Speyer, & Ochoa, 1961; Speyer et al., 1963; Speyer, Lengyel, Basilio, & Ochoa, 1962).

Nirenberg & Matthaei obtained similar synthetic RNAs from Leon Heppel and by the commercial availability of 16 RNA doublets that allowed them to assemble and decipher all of the 64 possible trinucleotides in seven years, by 1965 (Caskey & Leder, 2014).

Har Gobind Khorana confirmed and expanded these results by chemical synthesis of DNAs with desired sequences, and demonstrated the DNA-dependent RNA polymerase incorporation of specific amino acids into proteins in accordance with the triplet code (Nishimura, Jones, & Khorana, 1965).

1.6 mRNA: the information messenger

The existence of an intermediate molecule in the pathway transforming biological information to function was demonstrated in 1961.

It was accepted that genetic information about protein structure was contained in the DNA but also work suggesting that the conversion of this information happened in the cytoplasm of eukaryotic organisms, and was carried by the ribosomes. This localization was experimentally proved in the late 1940s (Caspersson, 1947).

Since proteins were not made directly on the genes themselves, an intermediate molecule had to transfer the information from the DNA to the cytoplasmic ribosomes. The most accepted idea was that each gene encoded a specific RNA which would form a gene specific ribosomal RNA responsible of the synthesis of each

specific polypeptide (the one gene-one ribosome-one protein hypothesis). Experimental evidence questioned this hypothesis. The ribosomal RNA was stable, homogeneous in size and base composition, both characteristics contrary to the observed polypeptide size diversity and variable base composition measured in the DNAs from different bacterial species.

Studies in bacteria suggested that the regulation of protein synthesis acts at the DNA level, with the existence of operon repressors and activators. Francois Jacob and Jacques Monod proposed that ribosomes were non specialized structures that received the genetic information in the form of a chemically uncharacterized "messenger" (Jacob & Monod, 1961).

Experiments with phage-infected bacteria, carried out by Sydney Brenner, Francois Jacob and Matthew Meselson brought evidence for the short-lived messenger RNA hypothesis. Letting bacteria grow in heavy isotope medium labels all ribosomes and RNAs as "heavy". After that, bacteria are infected and immediately transferred to a medium containing light isotopes. Density centrifugation allows the separation of the newly synthesized molecules upon the first minutes of viral replication.
Characterization of the molecules found in the different "heavy" and "light bands" shows that no new ("light") ribosomes are produced immediately after infection but a new ("light") RNA with a relative rapid turnover is synthesized, and it that has base composition corresponding to the phage DNA.
Moreover most if not all protein synthesis was carried out by preexisting ribosomes (Brenner, Jacob, & Meselson, 1961).

A simultaneously published paper from James Watson and collaborators combining pulse labeling and ultracentrifugation showed that non infected E.coli cells also produce short lived soluble RNA with base composition closer to the RNA produced by the infected cells than the ribosomal RNA.

They therefore provided evidence against the idea that the messenger RNA was a property of bacteriophage infection (Gros et al., 1961).

The discovery of the mRNA not only shed light on the mechanism of genetic information transfer but also paved the way for understanding translation and splicing.

2. The logic of gene regulation: bacteria and phages

The foundations of what we now understand as control of gene expression were captured in the studies on bacterial induction by François Jacob and Jacques Monod.
The proposed operon model (Jacob & Monod, 1961) was the first genetic regulatory mechanism to be fully understood proving to capture the essence of gene regulation: to use proteins with specific DNA site affinities to target RNA polymerase and conditionally express genes.

The lactose operon is used by bacteria to express the genes encoding the necessary proteins for the utilization of the disaccharide only when it is present in the culture medium. Lactose induces the expression of the genes for its metabolism.

In absence of lactose or when the prefered carbon source glucose is present, the lactose operon is not transcribed. In this situation, the constitutively expressed repressor LacI binds the operator, a DNA region near the promoter preventing the recruitment of RNA Polymerase and therefore its transcription.

In the presence of lactose (and when glucose is exhausted), E.Coli co-expresses three genes necessary for the disaccharide metabolism: β-galactosidase to hydrolyze lactose, a permease to speed its incorporation from the medium, and a transacetylase of unclear function. LacI acts as a repressor and as a sensor: When bound by lactose it detaches from DNA after an allosteric shift, allowing RNA polymerase to transcribe the polycistronic mRNA.
Other bacterial operon systems were described suggesting the generality of the mechanism (although some used activation instead of repression) (Englesberg, Irr, Power, & Lee, 1965).

An additional mechanism involved in the specific recruitment of

RNA polymerase to DNA was discovered when attempting to reconstitute in vitro transcription systems.

The bacterial RNA polymerase I was independently purified by the biochemists Charles Loe, Audrey Stevens, and Jerard Hurwitz in 1960 (Hurwitz, 2005). Posterior characterization of the enzyme and its activity with in vitro reconstitution experiments opened new questions. When intact double stranded DNA was used, only one strand was transcribed (Hayashi, Hayashi, & Spiegelman, 1963).

Moreover using known templates such as T2 or lambda phage DNA showed that transcription started at specific sites (Geiduschek, Snyder, Colvill, & Sarnat, 1966; Naono & Gros, 1966). The template specificity turned to depend on a protein, the sigma factor, co-purified with RNA polymerase using gentler extraction protocols (Burgess, Travers, Dunn, & Bautz, 1969).

Research on the sigma factor (σ) has revealed that E.Coli contains seven different factors that direct RNA polymerase to different types of genes. For instance σ70 directs polymerase to most "housekeeping" genes, active during growth. Instead σ32 is only expressed in heat shock conditions enabling the transcription of chaperones, proteases and DNA-repair enzymes to lessen heat damage (Burgess et al., 1969; Gross et al., 1998).

Mark Ptashne and collaborators focused their efforts in the study of the genetic switch of the lambda (λ) bacteriophage of E. Coli. This system has provided detailed mechanistic insight into transcription factor (TF) mediated gene regulation and establishment of stable transcriptional states. The switch refers to the non reversible induction of the viral replication cycle from a repressed lysogenic state upon UV radiation exposure.

In the course of the lytic growth the λ phage expresses several proteins that are virtually undetectable in the lysogenic state, a phenomenon similar to β-galactosidase induction. In lysogeny, λ repressor is highly transcribed. Its dimerization and binding to different viral operator regions inhibits the expression of the provirus early genes, the CRO activator and also feeds-back its own transcription. Upon DNA damage (usually UV), the repressor dimers undergo proteolytic cleavage through the bacterial SOS response and lose their affinity for the lysogen. The liberated operator allows RNA polymerase recruitment and Cro transcription. CRO dimers are in turn able to transactivate the bacteriophage genes responsible of the lytic cycle and also feed-back to its own gene.
The establishment of lysogeny upon bacterial infection is a delicate equilibrium between the production of repressor and the production of activator, competing for the same binding sites with different affinities in partially overlapping promoters (M. Ptashne et al., 1980); (Mark Ptashne, 2004).

Gene transcription in bacteria is the result of promoter-specific binding of proteins with the capacity to facilitate or abolish the recruitment of RNA polymerase to the gene. Contrary to metabolic enzymes, whose substrate specificity is encoded in the structure, RNA polymerase is an unspecific enzyme that conditionally transcribes DNA when efficiently recruited to the substrates of interest, the genes to be transcribed (Mark Ptashne & Gann, 2002).

3. From Prokaryote chromosomes to Eukaryotic chromatin

While the logic of transcription is equivalent to bacteria, in eukaryotes the DNA is associated with nucleosomes and chromosome-associated proteins that affect the ability of the DNA sequence to direct transcription. Chromatin context ultimately determines the ability of transcription factors in the process of RNA polymerase recruitment while modulating the efficiency of initiation and elongation of the transcript.

Therefore grasping the complexity of gene regulation in eukaryotes requires the understanding of how chromatin context (in composition and structure) affects and expands the transcription factor-centric regulation in bacteria.

3.1 Heterochromatin and Euchromatin

Chromatin is the the meld of proteins and DNA that constitutes the eukaryotic chromosome. Its role in gene regulation, chromosome compaction, DNA maintenance and replication has been evidenced with our increasing ability to study its complexity. The first level of heterogeneity that was described corresponds to the two morphological types exhibited by the mitotic chromatin fiber when observed by microscopy.

Friedrich Miescher discovered DNA in 1869. Working on the chemical composition of leukocytes nuclei he found proteins and a new phosphorus-rich substance that he called nuclein (Dahm, 2008).
In 1879 the cytogeneticist Walther Flemming developed a new staining technique. He was able to visualize a fibrous scaffold in the nucleus of dividing cells and called it Chromatin, the stainable substance of the nucleus (Paweletz, 2001).

Another cytogeneticist, Emil Heitz, proposed to divide chromosomes in what he coined eu- and heterochromatin. Euchromatin stretches change staining intensity during the cell cycle, they look condensed and darkly stained during cell mitosis but decondense in interphase showing a lighter staining. On the contrary, heterochromatin remains condensed along interphase, looking denser and stably stained (positively heteropycnotic) at all time (Passarge, 1979; Paweletz, 2001).

Heterochromatin was regarded as "genetically inert" by Heitz (Passarge, 1979; Paweletz, 2001). Loss or reduction in heterochromatin content proved largely inconsequential for the organism, an argument that was compatible with the idea that gene activity ceased in the compacted mitotic chromosomes. Consistent results came from genetic studies. The known X-linked genes resided in the euchromatic region of the chromosome but no gene was discovered in the larger heterochromatic region. The Y chromosome (seen as heterochromatic) was found to be dispensable for fly viability (Muller & Painter, 1932).

The "inert" nature of heterochromatin was called into question with the discovery of heterochromatin-residing genes (Hilliker & Sharp, 1988). A better understanding of the relationship between gene activity and chromatin composition had to wait until the understanding of position effects, allowing the study of gene activity after changing their euchromatin context to heterochromatin by translocation.

3.2 Position effect variegation

Position effect variegation (PEV) is the change of a gene activity resulting from a change of its chromosomal location. This phenomenon was the first example of chromatin effect on gene activity. In more than 60 years of study it has helped with the discovery of chromatin proteins and histone modifications associated with gene regulatory functions.

Hermann Muller found position effect variegation while systematically studying the phenotypes of X-ray induced mutations in Drosophila melanogaster. White-mottled eyes appeared in flies with chromosomal translocations changing the chromosomal position of white, a gene necessary for the synthesis of the wild-type red pigment.
The different cells (ommatidia) forming the composite eye of Drosophila showed different degrees of pigmentation. This result was puzzling, some ommatidia had a normal red coloration while others didn't, showing the gene ability to function normally in some cells while behaving as a mutant in others. This was not a mutation in the sense of the other "point mutations" that he described.
He noted: “even when all parts of the chromatin appeared to be represented in the right dosage (though abnormally arranged) the phenotypic result was not always normal.”

Given that individual eye cells are largely self-differentiating, Muller hypothesized that "the chromosome or gene controlling the eye color in this case must be subject to frequent genetic changes during eye development, i.e. must somehow be eversporting, somewhat like the genes for variegated pigmentation in corn and in some other plants". He referred to this type of "mutations" as "eversporting displacements" (Muller, 1930).

The discovery of white variegation attracted geneticists and cytogeneticists, who teamed up to study the relationship between gene regulation and chromatin.

The role of heterochromatin in position effect variegation was recognized in the mid 1930s. Muller concludes from Jack Schultz (Muller, 1930; Schultz, 1936) and his own work: "The mosaically expressed rearrangements always involve a transfer of the affected gene or genes into the neighborhood of a heterochromatic region; that is, the variegation is a kind of position effect peculiar to heterochromatin" (Muller, 1941).

The position effect hypothesis was further supported by accumulating examples of variegating gene rearrangements resulting from the relocation of genes to heterochromatin and the cytological visualization of heterochromatinization of the transposed loci (Hannah, 1951; Muller, 1941).

In his review of the late 1960's, William Baker notes that only minor advances have been achieved in the last two decades in the understanding of position effect variegation. He writes: "Explanations of the reason a gene is not producing its product (or at least, a normal one) in a given region of the variegated tissue are still put in terms of "heterochromatinization" or "compaction," terms that, in reality, expose our ignorance rather than our understanding" (Baker, 1968).

Already in 1967, Janice Spofford published the discovery of the first suppressor of variegation, the gene Su-V. A mutant allele mapping to chromosome 3 restored the wild type red eye in white variegating flies (Anderson et al., 1968; Spofford, 1967).

But in order to understand how mutations affecting heterochromatin composition could alleviate the observed silencing, a mechanistic understanding was needed in the first place. The role of the histones as barriers of DNA accessibility (preventing sequence recognition or transcription) and the organization of the chromatin in nucleosomal fibers with different levels of packaging provided a useful study framework as we will see in the next chapter.

3.3 Histones: From passive repressors to determinants of specificity

A role for the nuclear histones as suppressors of gene activity was proposed as early as 1950 by Stedman and Stedman (Stedman & Stedman, 1950; Watson & Crick, 1974). They demonstrated that the chemical composition of histones was cell-specific, after studying

lymphocyte and erythrocyte histone composition in the fowl.
They speculated: "The physiological functions of the nuclei are presumably due to the genes which they contain; they should, therefore, be identical in all nuclei of a given organism. If, however, it is postulated that nuclei contain some mechanism for the suppression of the activities of particular genes, or groups of genes, and that this mechanism is specific for each cell type, these difficulties disappear.
The demonstration in the work outlined above that some of the basic proteins present in cell nuclei are certainly cell-specific leads to the hypothesis that one of their physiological functions is to act as gene suppressors".

Direct proofs of the inhibitory effect of histones came a decade later with in vitro transcription experiments. Trypsin treatment removing about two-thirds of the total histone increased mRNA production 300% in nuclear suspensions of calf thymus (Allfrey & Mirsky, 1962) while supply of histones reduced mRNA production in proportion to the quantity added (Allfrey, Littau, & Mirsky, 1963).

Finally, DNA complexed with histones was unable to prime pea RNA polymerase. When the same DNA was deproteinized, transcription was efficient (Allfrey et al., 1963; Huang & Bonner, 1962).

The discovery that a fraction of the histones were chemically modified, and that the nature of the modification could be of different types (acetylated, phosphorylated and methylated histones were detected) prompted to study their putative role on gene expression.
Allfrey and collaborators proposed: "histone effects on nuclear RNA metabolism may involve more than a simple inhibition of RNA synthesis, and that more subtle mechanisms may exist which permit both inhibition and reactivation of RNA production at different loci along the chromosome".

Their hypothesis of histone modification was substantiated when isolated calf thymus nuclei were shown to incorporate radiolabeled acetyl and methyl precursors in the presence of puromycin (which blocks translation). This demonstrated that the modifications were post-translational and not a synthesis of specialized histones (Allfrey, Faulkner, & Mirsky, 1964).

The role of modification in allowing transcription was demonstrated when they showed that acetylated histones purified by chromatography were unable to block transcription upon addition of a calf thymus or bacterial polymerase fraction. The amount of transcripts correlated positively with the degree of histone acetylation and they were not due to eviction of the acetylated nucleosomes. They concluded that the complex DNA-acetylated histone is more favorable to transcription than the parent histone template (Allfrey et al., 1964).

Methylation and phosphorylation were also detected in the purified histones but, like acetylation, their role in transcription would not be characterized until the 1990s when the organization of the chromatin in nucleosomes was described.

3.4 The nucleosome and the "beads on a string" model

Elucidating the role of the histones and their modifications in DNA physiology was facilitated by the characterization of the nucleosome as the discrete unit of chromatin organization.
The advances in biochemistry allowed the dissection of the histone composition and their modifications, while crystallography and electron microscopy showed the organization of the chromatin in nucleosomes.

Digestion of chromatin with a nuclear deoxyribonuclease showed discrete bands after electrophoresis. In contrast, when applied to naked DNA a smear was obtained. This suggested that histones were protecting DNA digestion, and that they should be regularly spaced along the DNA (Hewish & Burgoyne, 1973).
In 1974 Roger Kornberg proposed the "beads on a string" model based on findings from several groups. His research proved that histone monomers were organized in discrete globular octamers composed of a dimer of the tetramer H4,H3,H2A and H2B (R. D. Kornberg & Thonmas, 1974).
Repeating units of the octamer were compatible with X-ray diffraction images obtained in the Wilkings laboratory. Moreover he argued that the repeating octamer, interspersed with "free DNA", was bound to 200 base pairs of DNA. Endonuclease restriction of chromatin and further gel separation was shown to give bands in multiples of 200 base pairs (R. D. Kornberg, 1974).
Finally, electron microscopy images of chromatin preparations (Olins & Olins, 1974) showed a "jointy structure" with "nodules" alternating with "thin strands" (R. D. Kornberg, 1974).

In 1975 the term nucleosome was coined. Electron microscopy images showed "a flexible chain of spherical particles (nucleosomes) of about 125 A in diameter, connected by DNA filaments".
High quality electron micrographs showing the "beads on a string" organization were obtained by depleting chromatin of histone H1. Biochemical and image analysis of histone H1 digested and reconstituted nucleosomes of the same samples confirmed the binding to ~200 base pairs of DNA and the octamer composition proposed by Kornberg (Oudet, Gross-Bellard, & Chambon, 1975).

Depleting histone H1 to obtain good quality of chromatin spreads was the first hint into its role in chromatin compaction. Later studies showed that histone H1 is not part of the nucleosome core but binds the DNA-core structure surface contacting the DNA at the exit/entry sites of the octamer particle stabilizing the structure.

Most of the knowledge of its role in establishing high order chromatin structures comes from In vitro reconstitution studies (Robinson & Rhodes, 2006) and the current structural details remain elusive for the difficulty of obtaining a crystal structure.
The viability of H1 knockout studies in different eukaryotic organisms is made difficult by the high number of subtypes, especially in vertebrates where lethality was shown in mice only after a triple knockout.
The more important idea arising from the knockout studies was that instead of a general repressor of transcription, in histone H1 knockout conditions only a subset of genes were misregulated. Histone H1 variants and the role of chemical modifications is an area of active study with promising insights in gene regulation (Hergeth & Schneider, 2015).

X-Ray crystallization of nucleosomes allowed for resolving the nucleosome structure at high resolution (Luger, Mader, Richmond, Sargent, & Richmond, 1997).
The DNA fiber was shown to wrap the disk-shaped particle 1.7 times encompassing 146 base pairs. Importantly it was shown that the N-terminal domains of the histone proteins protruded from the core particle, suggesting their greater accessibility and making them the ideal candidates to the chemical modifications discovered by Allfrey (Khorasanizadeh, 2004).

4. Eukaryotic gene regulation is implemented in chromatin

Laboratories working on different biological phenomena provided accumulating examples of the preeminent role of chromatin in eukaryotic gene regulation.
As a result, transcriptional control is now viewed as the result of the coordinated effects of sequence specific transcription factors together with histone modifications and nucleosome remodelling or exchange.
Those reactions are orchestrated by a plethora of chromatin proteins organized in macromolecular complexes with functional but complementary diversity.
The rapid pace of discovery in the chromatin field was propelled by the widespread incorporation of molecular biology techniques in the 1980s.
Notable examples encompass: recombinant DNA, rapid nucleic acid sequencing, separation and immunodetection of DNA, RNA and proteins. Years later the polymerase chain reaction further facilitated the selective amplification of genes to be characterized and to specifically mutate DNA using degenerated primers.

Given the high number of species-specific detailed mechanisms associating chromatin modification and composition with gene regulation we will discuss the ones that are better established in our organism of study, Drosophila melanogaster.

4.1 Histone acetylation: alleviating the nucleosome barrier

Building on the pioneering studies of Allfrey, several groups directed their efforts to understand the role of acetylation after the confirmation that nucleosomal DNA was refractory to transcription in vitro.

In vitro transcription of reconstituted chromatin with variable amounts of nucleosomes functionally confirmed that nucleosomes inhibit transcription initiation. The inhibition was shown to be dose-dependent: only when one third of the physiological amount of nucleosomes was added was transcription possible.
The authors argued that this low amount of nucleosomes diminished the likelihood of histone promoter occupancy therefore allowing RNA polymerase recruitment and transcription (Knezetic & Luse, 1986).

Different lines of research in several organisms brought further evidence that histone acetylation affects transcription in vivo.

Michael Grunstein and collaborators demonstrated the involvement of histone tails in transcription using S. cerevisiae mutants.
Yeast is particularly amenable to histone studies for it has only a single copy of each histone gene, greatly simplifying mutagenesis screens. They discovered that deletions of histone H4 N-terminal tails, or mutation of their conserved acetylation sites reduced the expression of the GAL4 gene by a factor 20, and by a factor 5 in the case of the gene PHO5 (Durrin, Mann, Kayne, & Grunstein, 1991).

With the advent of Chromatin immunoprecipitation (ChIP), a clear relationship between gene expression and histone acetylation was discovered. Histone hyperacetylation was shown to be correlated to gene activity, while silent regions showed hypoacetylation in different organisms (Kouzarides, 2007; Weintraub & Groudine, 1976).

Tetrahymena contains 2 types of functionally differentiated nuclei: a transcriptionally active macronucleus and a silent germline micronucleus. ChIP and immunofluorescence analysis demonstrated that the active macronuclei contain high levels of acetylation in opposition to the silent germline micronuclei (Durrin et al., 1991; Lin, Leone, Cook, & Allis, 1989).

D. melanogaster lacks chromosome X inactivation in females, expressing both chromosomes. In order to obtain the same expression levels, fly males double the transcription output of their single X chromosome in a sophisticated process called dosage compensation mediated by the male-specific lethal (MSL) complex (Conrad & Akhtar, 2012).

A specific acetylated isoform of histone H4, H4K16ac was detected in the hyperactive male X chromosome but was absent from autosomes or female X chromosomes (Turner, Birley, & Lavender, 1992).
In contrast the inactivated X chromosome of mammals was shown to be largely hypoacetylated except in the regions known to contain active genes (Jeppesen & Turner, 1993).
The relationship between histone acetylation and transcription was strengthened when histone acetyltransferases and deacetylases were found to form protein complexes with transcriptional coactivators and repressors.

Histone acetyltransferase A (HAT A) was discovered in the hyperacetylated macronuclei of Tetrahymena by David Allis and his collaborators. Other histone acetyltransferases (HATs) had been previously discovered but the correlation between acetylation and gene activity was mechanistically unexplained (Turner, 1991).
Cloning, sequencing and alignment of HAT A revealed a high degree of homology with a yeast transcriptional adaptor Gcn5p, conserved in flies and humans.
The yeast protein is able to bind acidic activators such as VP16 and was proven necessary for the correct activation of a subset of S. cerevisiae genes. The group of David Allis showed that Gcn5p also presented acetyltransferase activity (Brownell et al., 1996; Turner, 1991).
Functionally relating a transcriptional coativator, binding a diversity of sequence specific transcriptional activators and exhibiting histone acetylation capacity, supported more directly the role of this histone modification in gene activation.

The same year, the Stuart Schreiber laboratory reported the characterization of a mammalian histone deacetylase homolog of the yeast protein Rpd3. Using an affinity purification matrix with the irreversible histone deacetylase inhibitor Trapoxin, they eluted HD1 (for histone deacetylase 1). The homolog yeast protein Rpd3 was previously identified in a mutant screen of transcriptional repressors reinforcing the link between acetylation status and gene regulation (Taunton, Hassig, & Schreiber, 1996).

Those complementary and simultaneous findings established the early model for histone acetylation.
Transcription activating factors will recruit HATs acetylating the histones at the promoter, thereby overcoming the nucleosomal barrier. In the case of gene repression the converse will take place, transcriptional repressors will target histone deacetylases (HDACs) to keep the gene in a compacted and inaccessible state much like the chromatin structure observed at pericentromeric chromatin, the yeast mating type loci and telomeres, which will be discussed in detail later.

Today numerous HATs have been characterized and classified in different families according to their homology.
The better characterized families are GNAT (GCN5 related acetyltransferase), the MYST family and CBP/p300. More importantly, they have been proven to be associated with macromolecular complexes involved in gene activation in vivo, therefore suggesting a mechanism for their substrate specificity and expanding the correlative observations of gene activity and acetylation status.

The molecular mechanism that underlies the positive effect of acetylation in gene regulation is not fully understood but is proposed to mediate an increase in accessibility of transcription factors, chromatin proteins and the transcriptional machinery.

One example is the ability to modify the chromatin-DNA binding strength thereby increasing protein accessibility, as evidenced by higher DNAse I activity in hyperacetylated histones (Sealy & Chalkley, 1978). Another example is the increased ability of transcription factor binding to acetylated nucleosomal arrays in vitro (Vettese-Dadey, 1996).
This effect is attributed to histones having lost the electrostatic capacity to tightly bind DNA, forming a more "relaxed" structure where DNA is more exposed (Shogren-Knaak et al., 2006).

This accessibility hypothesis was already suggested in 1975, in experiments of DNAse I digestion of the globin genes in chicken erythrocytes. Embryos express a specific ^{3} -globin that is silenced at the end of development to express adult ^{3} -globin. The DNAse I sensitivity at the fetal ^{3} -globin gene from embryonic erythrocytes was higher than the sensitivity of the adult-specific ^{3} -globin. Conversely when using adult erythrocytes the most DNAse I sensitive ^{3} -globin sequence corresponded to the adult specific ^{3} -globin gene. (Kouzarides, 2007; Weintraub & Groudine, 1976).

The characterization of the bromodomain that recognizes acetylated lysine histone residues in several chromatin proteins, HATs (e.g Gcn5), nucleosome remodelers (e.g Brahma), and components of the basal transcriptional machinery (e.g TFIID complex) suggests that acetylation can also work as a signal for targeted recruitment.

Deacetylation is carried out by HDACs, members of three families according to homology and cofactor requirement. Like HATs, they form part of multiprotein complexes that target them to the appropriate genome locations to silence gene expression. In this model acetylation can target proteins to further activate or repress gene expression, depending on the protein that contains the bromodomains (Kouzarides, 2007; Weintraub & Groudine, 1976).

4.2 Nucleosome remodeling

Together with histone acetylation, chromosome remodeling contributes to increase the accessibility of regulatory proteins to the nucleosome-occluded DNA. Moreover it facilitates RNA polymerase elongation during transcription, playing a central role in proper gene expression.

The first discovered and most characterized chromatin remodeler is the yeast protein complex switch/sucrose-nonfermentable (SWI/SNF). The swi(1-5) proteins were discovered in genetic mutation screens for transcriptional activators affecting the switching mating type and other genes.
Meanwhile, the SNF proteins were discovered by mutations causing defects in transcription of SUC2, a glucose-repressible gene that encodes an invertase. Realising that the gene SWI2 was identical to SNF2 (now called SWI/SNF) fostered the discovery that several genes from those families were interdependent.
SWI/SNF mutants were recovered by secondary mutations affecting histones and other chromatin components which prompted the study of the chromatin structure at the regulated genes.
Micrococcal nuclease (MNase) digestion confirmed that SWI/SNF mutants had an altered chromatin structure at the SUC2 promoter gene (Hirschhorn, Brown, Clark, & Winston, 1992).

The complex was purified in yeast and mammals where it proved to disrupt nucleosome structure, enhance the binding of transcription factors, facilitate digestion with nucleases and generate DNAse I hypersensitive sites (DHS) in nucleosomal arrays in vitro (Becker & Workman, 2013).

Several homologs of SWI/SNF were subsequently purified in eukaryotes in which one of the hallmarks is the presence of bromodomains connecting this remodelers with histone acetylation.

In Yeast, another adenosine triphosphate (ATP)-dependent remodeler complex was described, “Remodels Structure of Chromatin” (RSC), which was shown to be more abundant and essential for mitotic growth (Cairns et al., 1996).
In Drosophila, the trithorax protein Brahma was discovered as a suppressor of polycomb mutants and was shown to be homolog of SWI/SNF when cloned and sequenced (Cairns et al., 1996; Tamkun et al., 1992).
In mammals, Brahma homologs are found as a subunits of multiple SWI/SNF-type remodeling complexes. For instance at least 30 SWI/SNF related remodelling ATPases are present in humans (Becker & Workman, 2013).

Cell-free functional assays of ATP-dependent nucleosome remodelling allowed the discovery of “Imitation Switch” (ISWI) in Drosophila embryos (de la Cruz, Lois, Sánchez-Molina, & Martínez-Balbás, 2005; Tsukiyama & Wu, 1995), a protein that had already been related to SWI/SNF for its partial homology.

The ISWI ATPase subunit, founding member of the subfamily, is present in several remodeler complexes such as the Nucleosome Remodelling Factor (NURF), the ATP-dependent nucleosome assembly and remodeling factor and the chromatin accessibility complex. This subfamily is characterized by the presence of a SANT histone interaction domain (Becker & Workman, 2013; de la Cruz et al., 2005; Tsukiyama & Wu, 1995).

The INO80 subfamily was discovered by homology search between Drosophila ISWI and Saccharomyces genome database. Of the three unknown hits, one was previously described in a mutational screen of the inositol pathway. INO80 was shown to be conserved in both Drosophila and Humans. The macromolecular complex was purified from yeast and confirmed in vitro to possess ATP dependent remodeling activity comparable to NURF and to facilitate transcription of a chromatin template (Shen, Mizuguchi, Hamiche, & Wu, 2000).

A third subfamily has chromodomains linking remodeling of nucleosomes with histone H3 lysine methylation. The Chromodomain helicase DNA-binding (CHD) subfamily was discovered while characterizing a transactivating protein of certain immunoglobulin promoters in mouse.
Sequencing of the protein revealed homology with SWI/SNF, the presence of a chromodomain and a DNA binding domain, which is absent from other chromatin remodelers (Delmas, Stokes, & Perry, 1993).
Members of the CHD subfamily are present from yeast to Humans in numerous remodelling complexes encompassing three families with different roles in chromatin biology (Delmas et al., 1993; Marfella & Imbalzano, 2007).

The role of chromatin remodeling complexes in gene regulation operates both at the initiation and transcript elongation. They are recruited to target genes via protein-protein interactions with DNA binding transcription factors, and via interactions with acetylated histones and methylated H3K4 through the bromodomains.
In yeast many transcription factors engage in direct contacts with SWI/SNF; in Drosophila NURF is recruited by the YY1 transcription factor and in mammals interactions with steroid nuclear receptors, pRB, BRCA-1, c-Myc and MLL have been documented in the context of cancer (Hargreaves & Crabtree, 2011).

A role for chromatin remodelers in repression has also been proposed. Mutation of chromatin remodelers in Yeast shows up and down-regulation of genes, suggesting that remodelers are also involved in repression (Hargreaves & Crabtree, 2011). For example the yeast ISW2 complex is targeted to meiotic genes by Ume6 where it contributes to their repression (Clapier & Cairns, 2009; Hargreaves & Crabtree, 2011).

The prevalent hypothesis is that remodeling complexes disrupt the nucleosomes at the target promoters thereby allowing the binding of transcriptional coactivators/repressors and the transcriptional machinery. For instance promoter nucleosome eviction has been demonstrated in the activation of the PHO5 gene in yeast after acetylation by the SAGA complex (Reinke & Hörz, 2003).

In elongation they are thought to serve in disrupting nucleosomes in front of elongating RNA polymerase and reassembling nucleosomes in its wake.
RSC and SWI/SNF assists Pol II in overcoming the nucleosomal barrier during elongation in vitro (Carey, Li, & Workman, 2006). In Drosophila, mutants of SWI2/SNF2 and CHD families of ATP-dependent chromatin-remodelers, diminished the amount of elongating RNA polymerase II but didn't prevent its recruitment to promoters (Srinivasan et al., 2005).

As of today the physical remodelling mechanism is still a matter of debate, but is thought to be attributable to the DNA translocation ability common to all the ATP-dependent remodelers (Clapier, Iwasa, Cairns, & Peterson, 2017).

4.3 Heterochromatin: Lessons on silencing

With the advent of more advanced molecular biology techniques, it was possible to advance the understanding of the phenomenon of position effect variegation. The role of heterochromatin silencing could be then studied beyond cytological terms and led to the discovery of one of the most understood mechanisms of chromatin-mediated gene silencing at constitutive heterochromatin. Three systems proved instrumental to dissect the proteins and mechanisms responsible for heterochromatin silencing:

the mating type loci and the telomeres of S. cerevisiae, the RNAi mediated heterochromatin silencing discovered in fission yeast and the studies on position effect variegation in centromeric heterochromatin of D. Melanogaster.

To mate correctly, yeast cells activate one of the two genes encoding the a or α mating factor. To express only one of them, they maintain one repressed copy of each mating type gene (HMLα and HMRa) in their genome. Induced recombination will insert either of the genes at expressor MAT locus leading to their activation and to the acquisition of a mating type. Both the repressed mating type genes exhibit hypoacetylation when compared to the same mating gene after activation through recombination with the MAT locus (Braunstein, Rose, Holmes, Allis, & Broach, 1993).
When both genes are derepressed, yeast are sterile, a phenotype which was used to characterize the proteins involved in silencing.

Several mutant proteins were shown to cause derepression of the mating loci: the "Silent Information Regulators" or SIRs genes. Mutations in SIR2, SIR3 or SIR4 completely derepressed HMLα and HMRa while SIR1 produced an a type yeast unable to mate. Also shown to be involved in the maintenance of the silenced state were the origin of replication complex ORC, Rap1 and the N-terminal domains of histone H3 and H4 (Grunstein & Gasser, 2013).
Relocating genes near the telomeres of S. cerevisiae showed that yeast heterochromatin was capable of inducing position effect variegation much like in Drosophila centromeres (Gottschling, Aparicio, Billington, & Zakian, 1990).
Subsequently it was shown that the same SIR proteins together with Rap1 could act as suppressors of variegation, showing that the same proteins repressed transcription at the mating type loci and at the telomeres, suggesting a common mechanism (Aparicio, Billington, & Gottschling, 1991).

Although immunoprecipitation techniques showed that SIR proteins interacted physically in heterochromatin, SIR1, SIR3 and SIR4 are only conserved among S. cerevisiae and other budding yeast, defeating comparisons based on homology in other organisms.

On the contrary SIR2 is well conserved from bacteria to humans (Blander & Guarente, 2004) and its overexpression showed histone hypoacetylation activity in yeast (Braunstein et al., 1993), a hallmark of silent chromatin. Molecular characterization demonstrated that SIR2 is an H4 lysine 16 deacetylase, paving the way to our current understanding of the role of deacetylation in heterochromatin silencing (Imai, Armstrong, Kaeberlein, & Guarente, 2000).

The 'initiation and spreading model' (Hecht, Laroche, Strahl-Bolsinger, Gasser, & Grunstein, 1995) proposes that Rap1 recognizes and binds specific DNA sequences at the telomeric TG rich repeats. It recruits SIR4 which in turn recruits SIR3 and SIR2, the histone deacetylase. As a result, histone tails from H4 and H3 are deacetylated, producing SIR3/SIR4 binding sites and allowing the nucleation the SIR2-3-4 complex to spread to adjacent nucleosomes.
This results in the folding of the telomeres in an low accessibility higher-order structure (Grunstein & Gasser, 2013).
A similar model is proposed for heterochromatin silencing at mating type loci, but the initiation is not repeat-driven. Instead SIR4 is recruited by Rap1/Abf1 or by ORC and SIR1 (Grunstein & Gasser, 2013).
While the system was useful to understand the role of decetylation in gene repression, the lack of conservation of important players in the heterochromatin formation and maintenance pathway stressed the need to study other model organisms to understand heterochromatin silencing.

Studies in S. pombe provided a mechanism of initiation and maintenance of heterochromatin silencing that relies on the interference RNA (RNAi) machinery. The RNAi pathway was initially discovered for its ability to cause post-transcriptional silencing (PTGS) in Caenorhabditis elegans germline (Fire et al., 1998).

RNA interference is used in all organisms from yeast to humans to mediate post-transcriptional silencing by cytosolic mRNA degradation. Double stranded RNAs (dsRNAs) and pre-microRNAs are processed by the Dicer family of endonucleases generating small (~22 nucleotides), single stranded small interference RNAs (siRNAs) that are loaded to the Argonaute family of proteins in the RNA-induced silencing complex (RISC) responsable of post transcriptional gene silencing (PTGS) of target mRNAs.

Mechanistically Argonaute proteins are capable of base-pair recognition between loaded siRNA and target mRNA resulting in its the degradation (Martienssen & Moazed, 2015).

In the context of heterochromatin formation and maintenance, the pathway is nuclear instead of cytosolic and is referred to as RNA-mediated transcriptional silencing (TGS).
Heterochromatin silencing is mediated by the RNA-induced transcriptional silencing (RITS) complex, responsible for the sequence-directed recruitment of protein repressors such as Clr4, the yeast H3K9 methylase.

S. pombe contains a unique central core region (Cnt) enriched in repeated sequences. This region is flanked by two families of repeats called “innermost” and “outermost”.
It is the transcription of those repeated regions that provides the non-coding RNAs that will be processed by Dicer to generate siRNAs thought to target the RITS complex together with the heterochromatin repressing proteins that mediate and maintain silencing (Allshire & Ekwall, 2015).

Deletions of Argonaute, Dicer or RdRP caused the loss of centromeric H3K9 methylation and were accompanied by transcriptional de-repression of transgenes integrated at the centromere. Moreover, accumulation of complementary transcripts from centromeric repeats was detected (T. A. Volpe et al., 2002). This suggested that the non-coding repeat transcripts are processed and incorporated to Argonaute, in accordance with the sequencing results of small RNA library in which › 22-nt RNAs that mapped exclusively to centromeric repeat regions were detected (Reinhart & Bartel, 2002; T. Volpe et al., 2003).

It was the characterization of the heterochromatin-associated chromodomain containing protein Chp1 that led to the discovery and purification of the RITS complex and allowed mechanistic studies. The composition of the complex showed that in addition to Chp1, it contains the yeast Ago1 (encoding Argonaute), Tas3 GW domain protein (of unknown function), and dicer-processed centromeric siRNAs.
The identity of siRNAs coprecipitated with the complex was confirmed by Northern blot using siRNAs complementary probes previously identified as centromeric repeats (Reinhart & Bartel, 2002; T. Volpe et al., 2003).
Furthermore, the centromeric localization of the RITS complex was confirmed by ChIP, a localization that was lost in a Dicer mutant without siRNA associated to RITS. Taken together these results indicated that targeting of the RITS was dependent on centromeric siRNAs (Verdel et al., 2004).

The mechanism by which siRNAs target the RISC complex is debated between base pairing between siRNAs with a partially unwound repeat region or by complementarity to non-coding transcripts arising from the repeats.

The key connection to heterochromatin silencing is the ability of the RITS complex to recruit Clr4 to catalyze the targeted methylation of H3K9. The methylated histones will then be recognized and bound by the heterochromatin protein 1 (HP1) yeast homologs Swi6 and Chp2 via their chromodomain, stabilizing the RNAi complex and recruiting the SHREC (Sugiyama et al., 2007) complex containing the HDAC Clr3 and the chromatin remodeler Mit1.
The activity of the complex is responsible for shutting down transcription from the repeats and for establishing the repressive chromatin structure (Martienssen & Moazed, 2015). The involvement of the RITS complex in heterochromatin silencing has provided insight on a family of proteins that is conserved in organisms such as Drosophila and mammals, allowing the discovery of mechanisms for silencing transposon remnants and for viral defense.

While interference RNA has a clear role in post-transcriptional silencing, especially in the germline, its role as a transcription regulator of non-transposon genes at the level of chromatin is still unclear in Drosophila and mammals (Martienssen & Moazed, 2015).

D. melanogaster, the original organism where PEV was discovered and studied, has two advantages to study heterochromatin gene silencing and the majority of the proteins involved are conserved in mammals.The first advantage is technical: the development of P-element derived transposable reporters allowed to integrate the white gene in different positions, speeding the generation and selection of variegating fly lines. The second is the the high resolution of immuno-localization studies thanks to the giant polytene chromosomes which have permitted the genome localization of chromatin proteins and histone modifiers involved in heterochromatic silencing (Elgin & Reuter, 2013).

The more recent techniques based on genome-wide chromatin immunoprecipitation (ChIP-Chip, ChIP-seq) have confirmed the microscopic observations of the different chromosomal composition between hetero- and euchromatin, with much better resolution (Filion, van Bemmel, Braunschweig, et al., 2010; Kharchenko et al., 2011a).

Mutational screens of proteins enhancing or suppressing the variegating phenotype (E(var) and Su(var) alleles) allowed the characterization of the first components linking chromatin proteins and histone modifications to heterochromatic silencing. With many yet unidentified alleles, the variegation modifiers continue to provide candidates to complete our understanding of chromatin silencing.

The key discovery was that HP1, a protein characterized for its preferential binding to heterochromatin, acted as a dominant suppressor of variegation (and therefore was also known as Su(var)2-5). This observation suggested that heterochromatin-specific proteins play a central role in gene suppression at the centromeres (Eissenberg, Morris, Reuter, & Hartnett, 1992).

Another suppressor of variegation colocalizing with HP1 in polytene preparations, Su(var)3-9, was purified and characterized. The encoded protein contains a Su(var)3-9, E(z), Trithorax (SET) domain homologous to the developmental regulators Enhancer of Zeste and Trithorax (Eissenberg et al., 1992; Tschiersch et al., 1994).
The methyltransferase activity of Su(var)3-9 was demonstrated later in the human homolog SUV39H1, and proved to be specific for the lysine 9 of the histone H3 N-terminal tail (Rea et al., 2000). The functional role of this methylation in heterochromatin silencing was supported by the dosage dependence of Su(var)3-9 in silencing and the enhanced effect obtained with an enzymatic hyperactive mutant of the protein (Elgin & Reuter, 2013).

Accumulating studies on the function and interaction of those factors has led to the model in which HP1 (Su(var)2-5) is recruited to centromeric chromatin in a macromolecular complex that includes Su(var)3-7 (a zinc-finger protein with DNA binding capability), Su(var)3-9 which catalyzes H3K9me3 and HDAC1, a previously recognized histone deacetylase (Rpd3 in yeast) that acts also as suppressor of variegation and therefore contributes to heterochromatin repression (Elgin & Reuter, 2013).

Another suppressor of variegation catalyzes the methylation of histone H4 at lysine 20, a mark characteristic of centromeric chromatin which is also reduced in HP1 mutants . Su(var)4-20, which has been shown to be very stably associated with heterochromatin, would help HP1 recruitment in addition to the methylated H3K9 nucleosomes (Schotta et al., 2004).

The role of transcriptional RNAi silencing in heterochromatin formation is not fully established in Drosophila. Mutations in components of the RNAi pathway such as piwi, aubergine (the piwi-subfamily Argonaute homolog) and spn-E act as suppressors of variegation of white transgenes both at the centromeres and at the fourth chromosome. The effect is accompanied with reductions in H3K9 methylation and mislocalization of HP1 suggesting that components of the RNAi pathway are needed to silence in a HP1-H3K9 dependent manner (Pal-Bhadra et al., 2004).

While the P-element Induced WImpy testis (PIWI) pathway for PTGS transposon silencing in the germline is well documented in flies, their activity doesn't seem to be found in somatic cells. A mutation of spn-E was found to cause transposon upregulation in the germline but not in somatic tissues from carcass or heads (Klenov et al., 2007).

A more recent study proposed that heterochromatin is established in the early embryo in a PIWI dependent manner but inherited mitotically in the somatic tissues via HP1 binding to already deposited H3K9me and further recruitment of Su(var)3-9 without the participation of piRNAs (Gu & Elgin, 2013).

A mechanism for de novo heterochromatin establishment in Drosophila has recently been proposed. Initial monomethylation of histone H3K9 is carried out by Prdm3 and Prdm16, two cytoplasmic HKMTs, and the monomethylated nucleosomes are then incorporated to repressive chromosomal regions and further methylated by Su(var)3-9 in the case of the centromeres.

The recruitment of the H3K9 methyltransferase will in this context be aided by two transcriptional repressors (Pax3 and Pax9). The centromeric major satellite repeats contain binding sites for those repressor proteins explaining the DNA-directed recruitment (Dambacher, Hahn, & Schotta, 2013).

H3K9me3 is observed at telomeres and at the fourth chromosome of Drosophila, but perturbation of HP1 does not affect variegation of reporters integrated at the telomeres. Moreover, mutations affecting reporter expression on the fourth chromosome point to the egg methyltransferase (homolog to human SETDB1) but not to Su(var)3-9.

This suggests some conservation but also some divergence in the molecular players responsible for heterochromatin silencing at different genomic locations. At the same time, the high number of variegation modifiers detected in recent screens (~500) suggests that molecular characterization of novel players may contribute to a deeper understanding of the different actors involved in heterochromatin silencing (Elgin & Reuter, 2013).

4.4 Repression and Activation: insights from Development

Development is characterized by the progressive acquisition of cell identities resulting from the selective regulation of specific gene sets in time and space encoded both by maternally deposited RNAs and zygote transcription.
Early studies aiming at discovering the molecular determinants responsible of transcriptional program maintenance unraveled several chromatin proteins and histone modifications key to understanding eukaryotic gene regulation.

The study of D. melanogaster development evidenced the importance of the tissue-specific and time-restricted expression of HOX genes.
This family of transcription factors contains a homeodomain that recognizes specific DNA sequences (called homeboxes) in the promoters of genes responsible for giving an identity to the body segment where they are transcribed. HOX genes are master regulators of several development genes, a fact that was proven by the diverse phenotypes caused by their mutation or misregulation.

Classical examples include the appearance of legs in the place of antennae, additional thoracic segments or additional sex combs in the second and third leg pairs. Genetic studies of those phenotypes led to the discovery of highly conserved protein complexes that orchestrate the repression and sustained activation of identity gene sets via chromatin remodeling and chromatin modification.

In the case of gene repression, the Polycomb group of proteins (PcG) assemble in multiprotein complexes called Polycomb Repressive Complexes 1 and 2 (PRC1 and PRC2), which mediate the maintenance of repression initiated by the HOX genes.

PRC2 is conserved in all eumetazoans. Biochemical analysis of its composition revealed four core proteins: “Enhancer of Zeste” E(z), “Extra Sexcombs” esc, Caf1-55 and “Suppressor of Zeste” Su(z)12. Moreover some of the complexes additionally contained the previously described histone deacetylase Hdac1 (homolog to yeast Rpd3). The sequence of E(z) was shown to contain a SET domain (Jones & Gelbart, 1993), prompting biochemical purification and reconstitution to test its methylation ability.

The protein was shown to methylate H3K27 in vitro using nucleosome arrays or free histones. H3K9 was also methylated, although to a lesser extent.

Moreover E(z) mutants with abnormal HOX silencing were rescued by recombinant E(z) expression during development.
This result linked HOX repression with H3K27 methylation (Müller et al., 2002).
The accessory protein Hdac1, found in some PRC2 complexes, is an histone deacetylase already involved in yeast silencing (Grossniklaus & Paro, 2014; Fischle, 2003).

The PRC1 complex is conserved from flies to mammals, and it also contains four core proteins: Polycomb (Pc), Polyhomeotic (one of the orthologs ph-p or ph-d), Posterior sex combs (Psc) and Sex combs extra (Sce). Polycomb is a chromodomain-containing protein. As a heterozygote mutation, it causes the appearance of additional sex combs in the second and third pairs of legs, an effect already hypothesized to result from the inability to repress the bithorax gene complex in abdominal segments in the late 1970s (Lewis, 1978; Müller et al., 2002).

Characterization of its protein sequence revealed the presence of a chromodomain, suggesting its functional parallelism with HP1 and the mitotically heritable silencing shown in heterochromatin PEV (Lewis, 1978; Müller et al., 2002; Paro & Hogness, 1991).

H3K27me3 and Polycomb colocalization experiments in polytene chromosomes and preferential binding of Pc chromodomain to H3K27me in vitro suggest a recruitment mechanism of PRC1 to methylated H3K27, the product of the PRC2 protein E(z), although this mechanism is believed to be insufficient for targeting the PRC1 complex (Grossniklaus & Paro, 2014; Fischle, 2003).

In Drosophila, PRC1 and PRC2 recruitment is also dependent on DNA binding affinity to Polycomb Response Elements (PREs). This cis regulatory modules are thought to loop to the promoters to be silenced by PcG proteins and their role was shown by their ability to silence transgenes (Grossniklaus & Paro, 2014; Fischle, 2003).

The role of DNA sequences responsible for PRCs targeting in mammals and plants is currently unclear and remains an active area of study.
This suggests the existence of distinct mechanisms of silencing and recruitment (David Allis, Caparros, Jenuwein, & Reinberg, 2015).

Instead of altering the properties of nucleosomes, as proposed for acetylation, the hypothesis for polycomb-mediated silencing points to H3K27me3 nucleosomes as specific binding platforms for the chromodomain of PRC1, specifically recognizing the methylated mark and targeting the enzymatic activities responsible for silencing transcription (Grossniklaus & Paro, 2014; Fischle, 2003).
This is demonstrated by experiments in which increasing the H3K27me3 levels in human cells enhances PRC1 binding (Lee et al., 2007).
Moreover Drosophila mutants impeding methylation in H3K27 were shown to phenocopy Polycomb mutants at PcG target genes, leading to their expression. This evidences the necessity of H3K27 histone methylation for silencing (Pengelly, Copur, Jäckle, Herzig, & Müller, 2013).

In vitro reconstitution of the mammalian PRC1 complex was found to inhibit chromatin remodeling by SWI/SNF and to restrict access of RNA Polymerase II to nucleosomal arrays (King, Francis, & Kingston, 2002; Shao et al., 1999), suggesting that repression could be due in part to a nucleosome high order structure at the targeted loci.
The Polycomb-mediated repression mechanism is thought to be responsible for facultative heterochromatin silencing, although genome-wide histone modification and chromatin protein profiling has shown that the majority of the silenced genes are not bound by either H3K27me3 or PcG proteins (Filion, van Bemmel, Braunschweig, et al., 2010; Kharchenko et al., 2011a).
Polycomb repression is therefore only partially responsible for gene silencing via facultative heterochromatin establishment.

Another link between Polycomb and gene repression came from the study of X inactivation in mammals, whereby the majority of genes in one chromosome is silenced coincident with cell differentiation. The mechanism of inactivation is complex and involves modification of histone tails, incorporation or exclusion of histone variants, DNA methylation of some CpG islands, and reorganization of higher-order chromatin structure. ChIP studies showed that H3K27me3 and H2AK119ub1 are enriched at non-overlapping regions across the inactive X (Chadwick, 2003).
The recruitment of PCR2 responsible for the methylation of H3K27 has shown to be attributable to the “X-inactive specific transcript” (Xist) non-coding RNA, the principal effector in the inactivation process. It was suggested that PRC2 binds the A-repeat element of Xist RNA coating the inactive chromosome (Zhao, Sun, Erwin, Song, & Lee, 2008).

Nevertheless it is a current matter of debate whether this recruitment mechanism is sufficient, since deletion of the repeat does not abrogate PRC2 localization (Mira-Bontenbal & Gribnau, 2016). The recruitment of PRC1 to H3K27me3 and its silencing would follow the normal chromodomain-dependent recruitment mentioned above (David Allis et al., 2015).

In mirroring opposition to PcG, proteins of the trithorax group (TxG) are responsible of the maintenance of activated states of HOX genes. They were genetically identified as suppressors of PcG phenotypes or by mutations that phenocopy HOX gene loss of function (Kennison, 1995).

Molecular characterization of its members has revealed a diverse family of effector proteins with different biochemical activities associated with gene activation. Conserved members of the group from yeast to humans strengthened the connexion between transcription and chromatin modification.

Mutations in the brahma (brm) gene compensated the derepression phenotypes observed in Polycomb mutants. Cloning and characterization of brahma showed domain conservation with the yeast chromatin remodeler SWI2/SNF2 linking brahma to gene activation (Tamkun et al., 1992).
Moreover brahma was later shown to contain a bromodomain which binds acetylated histones, connecting remodeling with acetylation.
Two other TrxG genes encode chromatin remodelers kismet (kis) and moira (mor) (Kingston & Tamkun, 2014).
TxG proteins containing a SET domain contributed to link histone methylation with gene activation.

The products of trithorax (trx) and “Absent, small or homeotic discs 2” (ash2) are the most studied examples. The trx gene and its homologs from yeast (SET1) to Humans (MLL) methylate H3K4, which can be mono-, di- and trimethylated and correlates with gene activation.

In HOX gene clusters this mark was shown to be distributed broadly along the locus and to correlate with the activity of the underlying specific HOX gene. In genome-wide localization studies this mark is observed at active enhancers and promoters in Drosophila (Filion, van Bemmel, Braunschweig, et al., 2010; Kharchenko et al., 2011a)and Humans (Ernst et al., 2011).

TrG proteins act in complexes that have multiple functions in addition to transcription, making their isolated study difficult. Moreover mutation of lysine 4 in histone tails also prevents acetylation at the same residue, making it impossible to disentangle the effects.

One of those complexes in Drosophila contains the trx protein and the broad histone acetylase nejire (nej, homolog of human CBP/P300) linking both modifications with gene activity (Kingston & Tamkun, 2014). In the case of ash2, the lysine methylation pattern is less specific, and it is able to methylate H3K4, H3K9, H4K20, and H3K36 in vitro. In vivo, methylated H3K36 is broadly observed in the body of active genes. The mechanism of action is less understood than trx, but it has also been found to interact with several histone acetylases (Kingston & Tamkun, 2014).

In yeast, the Drosophila trx and ash2 ancestral homologues associate in a protein complex, Set1/COMPASS, the first H3K4 methylase identified and characterized as being able to catalyze mono-, di- and trimethylation of H3K4, suggesting that the active correlation of this histone mark is not limited to developmental processes and HOX gene regulation (Shilatifard, 2012).

It was observed that Set1/COMPASS travels with the transcribing polymerase. Yeast screens showed the RNA polymerase II C-terminal domain associated kinases (Ctks), involved in the transition to elongation, recruited the Set1/COMPASS to the transcribing polymerase.

The higher time spent of the polymerase in the proximal promoter and in the proximal pause release region of the gene would explain the higher degree of H3K4me3 methylation compared to the rest of the gene, more rapidly traversed by RNA polymerase upon transition to the elongating form of RNA Polymerase II (Wood et al., 2007).

Like the repressive mark methylated H3K27, H3K4me3 is thought to exert its action by recruiting other proteins. In Drosophila the trithorax related (trr) H3K4 methylase was shown to immuno-colocalize with ecdysone receptors (EcR) on polytene chromosomes. The trr protein was shown to bind EcR in an ecdysone-dependent manner and cause H3K4 trimethylation at ecdysone-inducible promoters. Mutations in the trr SET domain, abrogating H3K4 methylation were accompanied by defects in the developing eye (Sedkov et al., 2003).

Similarly, in human cells the MLL2 complex acts as a coactivator of estrogen response. Estrogen receptor alpha (ERa) recruits MLL2 in a ligand-dependent manner, which is necessary for the correct induction at estrogen regulated promoter genes (Mo, Rao, & Zhu, 2006; Sedkov et al., 2003).

Despite multiple examples of correlations between gene activity and the methylation levels of H3K4 (Benayoun et al., 2014), an instructive role for this mark in transcription is still debated. Studies in different organisms, targeting different states of the methylating mark yield conflicting results in terms of expression changes upon loss of H3K4me1/2/3 (Howe, Fischl, Murray, & Mellor, 2017).

The same is true for ash2-deposited H3K36me3 that has been associated to transcription activation in promoters, to prevent aberrant elongation products when found in the body of active genes and suggested to mediate repression in yeast (Wagner & Carpenter, 2012).

4.5 Histone code and genome-wide chromatin

In the early 2000s the accumulating examples of histone mark combinations and their association with transcriptional regulation increased exponentially.
This has resulted in the general agreement that histone modifications play a central role in gene regulation
(view Table 1 for histone modifications and their associated roles in gene regulation).

The phylogenetic conservation of the modifiers, histone marks and the presence of bromo- and chromodomains contributed to the idea that histone modification constituted a code.

The “histone code” postulates that different combinations of histone modifications result in alterations of the chromatin higher-order structure leading to silencing or activation, thereby “expanding the information content of the genome past DNA code” (Jenuwein & Allis, 2001).

In this fashion the old concept of heterochromatin equating silencing and euchromatin equating activity was actualized with histone marks associated with silent regions equating silencing and histone marks associated with active regions equating activation.

Standard Name	Human	D. melanogaster	S. cerevisiae	S. pombe	Substrate Specificity	Function
KDM1	LSD1/BHC110	Su(var)3-3	SpLsd1/Swm1/Saf110		H3K4me1/2, H3K9me1/2	Transcription activation and repression, heterochromatin formation
KDM2	Jhd1	H3K36me1/2				Transcription elongation
KDM3A	JHDM2a	H3K9me1/2				Androgen receptor gene activation, spermatogenesis
KDM4	Rph1	H3K9/K36me2/3				Transcription elongation
KDM4A	JMJD2A/JHDM3A	H3K9/K36me2/3				Transcription repression, genome integrity
KDM4B	JMJD2B	H3K9/H3K36me2/3				Heterochromatin formation
KDM5B	JARID1B/PLU-1	H3K4me1/2/3				Transcription repression
KDM6A	UTX	H3K27me2/3				Transcription activation
KDM6B	JMJD3	H3K27me2/3				Transcription activation
KAT2	dGCN5/PCAF	Gcn5	Gcn5		H3 (9, 14, 18, 23, 36)/H2B; yHtzl (14)	Transcription activation, DNA repair
KAT2A	hGCN5				H3 (9, 14, 18)/H2B	Transcription activation
KAT2B	PCAF				H3 (9, 14, 18)/H2B	Transcription activation
KAT3	dCBP/NEJ				H4 (5, 8); H3 (14, 18)	Transcription activation, DNA repair
KAT3A	CBP				H2A (5); H2B (12, 15)	Transcription activation
KAT3B	P300				H2A (5); H2B (12, 15)	Transcription activation
KAT4	TAF1	dTAF1	Taf1	Taf1	H3 > H4	Transcription activation
KAT5	TIP60/PLIP	dTIP60	Esa1	Mst1	H4 (5, 8, 12, 16); H2A (4,5,7,9, 13,14,15)	Transcription activation, DNA repair
KAT6	(CG18894)	Sas3	(Mst2)		H3 (14, 23)	Transcription activation and elongation, DNA replication
KAT6A	MOZ/MYST3	ENOK			H3 (14)	Transcription activation
KAT6B	MORF/MYST4				H3 (14)	Transcription activation
KAT7	HBO1/MYST2	CHM	(Mst2)		H4 (5, 8, 12) > H3	Transcription, DNA replication
KAT11	Rtt109				H3 (56)	Genome stability, transcription elongation
KAT12	TFIIIC90				H3 (9, 14, 18)	Pol III transcription
KAT13A	SRC1				H3/H4	Transcription activation
KAT13B	ACTR				H3/H4	Transcription activation
KAT13C	P160				H3/H4	Transcription activation
KAT13D	CLOCK				H3/H4	Transcription activation
KMT1	Su(Var)3-9	Clr4			H3K9	Heterochromatin formation/silencing
KMT1A	SUV39H1				H3K9	Heterochromatin formation/silencing
KMT1B	SUV39H2				H3K9	Heterochromatin formation/silencing
KMT1C	G9a				H3K9	Heterochromatin formation/silencing
KMT1D	EuHMTase/GLP				H3K9	Heterochromatin formation/silencing
KMT1E	ESET/SETDB1				H3K9	Transcription repression
KMT2	Set1	Set1			H3K4	Transcription activation
KMT2A	MLL1	Trx			H3K4	Transcription activation
KMT2B	MLL2	Trx			H3K4	Transcription activation
KMT2C	MLL3	Trr			H3K4	Transcription activation
KMT2D	MLL4	Trr			H3K4	Transcription activation
KMT2E	MLL5				H3K4	Transcription activation
KMT2F	hSET1A				H3K4	Transcription activation
KMT2G	hSET1B				H3K4	Transcription activation
KMT2H	ASH1	Ash1			H3K4	Transcription activation
KMT3	Set2	Set2			H3K36	Transcription activation
KMT3A	SET2				H3K36	Transcription activation
KMT3C	SYMD2				H3K36	Transcription activation
KMT4	DOT1L	Dot1			H3K79	Transcription activation
KMT5A	Pr-SET7/8	PR-set7			H4K20	Transcription repression
KMT6	EZH2	E(Z)			H3K27	Polycomb silencing
KMT8	RIZ1				H3K9	Transcription repression

Table 1. Histone lysine modifications associated with gene regulation. Standard names: KDM: Histone lysine Demethylases, KAT: Histone lysine Acetyltransferases and KMT: histone lysine Methyltransferases. Adapted from Allis et al. 2007.

As we have discussed above, the role of most of these modifications is not completely understood and it appears to be context-dependent. This might be due to our incomplete characterization of the proteins (and their combinations) capable to transform this histone modifications into functions relevant to gene regulation. The variable functions associated with each chromatin mark combination and the debate around their causative effect in expression has led some scientist to question if they constitute a code extending the DNA sequence (Henikoff & Shilatifard, 2011; Wagner & Carpenter, 2012).

Molecular genetics studies have contributed several examples confirming that heterochromatin is "compact" and generally silent and euchromatin is "relaxed" and principally active. Genome-wide studies have nevertheless reported notable exceptions to this dichotomy, showing that gene-specific relationships with the chromatin states (combinations of histone-marks, acetylation levels and chromatin associated proteins) capture with greater accuracy the complexity of chromatin effects in gene regulation. (Ernst et al., 2011; Filion, van Bemmel, Braunschweig, et al., 2010; Kharchenko et al., 2011a).
Perhaps the most shocking example is the finding that the vast majority of silent genes lie in a chromatin state devoid of known histone marks, chromatin proteins and appears to be refractory to transcription factor binding. Additionally two other different chromatin types have been found to be associated with silencing involving distinct chromatin composition, a situation that is also seen in euchromatin.
Housekeeping and regulated genes seem to utilize different chromatin proteins and histone marks to control their activity.

In the next chapter we will summarize the findings of genome-wide segmentations in chromatin states conducted in D. melanogaster, C. elegans and human cells.

5. Chromatin states

In the early 2010s several laboratories undertook the genome-wide characterization of chromatin composition. The studies were carried in Drosophila, C. Elegans, and Humans.
Using ChIP-chip, ChIP-seq and DamID, they provided genome-wide location maps for the principal histone marks and hundreds of chromatin associated proteins. The first surprising result was that given the theoretical combinatorial complexity, only few combinations of proteins and histone marks (a chromatin state) were reported in all the studies.

Consistent with the idea that co-regulated genes often cluster along the chromosome, the genomes were found to be organized in domains (Hurst, Pál, & Lercher, 2004; Thévenin, Ein-Dor, Ozery-Flato, & Shamir, 2014) with chromatin states encompassing several genes but also subdividing gene regions into promoters, introns and gene bodies and other regulatory elements such as enhancers.
Using hidden Markov models to segment the chromosomes in different consistent protein and histone mark combinations, the genome of Drosophila was found to contain only five (Filion, van Bemmel, Braunschweig, et al., 2010) or nine (Kharchenko et al., 2011a) states, depending on the study.
In the the case of humans, a first fine-grained classification found 51 states (Ernst & Kellis, 2011) which were compressed down to only seven states (ENCODE Project Consortium, 2012).

Even if there are much fewer states than theoretically possible, the question is which are the equivalents of hetero- and euchromatin.

Constitutive heterochromatin with the expected characteristics, i.e. H3K9me3, HP1 and SU(Var)3-9 was detected in all the studies (ENCODE Project Consortium, 2012; Filion, van Bemmel, Braunschweig, et al., 2010; Gerstein et al., 2010; Kharchenko et al., 2011a) concentrated in pericentromeric areas.

Its role as a general transcriptional repressor was questioned when 45% of the Drosophila pericentromeric genes, covered by this chromatin state showed active expression.
It is tempting to speculate that some constitutive heterochromatic genes have evolved a mechanism to escape heterochromatin silencing (Yasuhara & Wakimoto, 2006).

In the case of facultative heterochromatin, another major surprise emerged. Domains characterized by H3K27me3, the polycomb proteins assembling PRC1, PRC2 and associated cofactors, were also captured by all the models in all organisms. They silence developmentally regulated genes, as anticipated by the study of the role of Polycomb and H3K27me3 (Grossniklaus & Paro, 2014).

Surprisingly, the majority of silent genes present in any studied organism was not contained in those polycomb chromatin states. Instead, the majority of silent genes sit in a chromatin state that is devoid of most chromatin protein or histone mark tested. This chromatin state was given different names in the different organisms or studies "Black chromatin" (Filion, van Bemmel, Braunschweig, et al., 2010), "silent domains" in C. elegans (Gerstein et al., 2010), "low-activity regions" (ENCODE Project Consortium, 2012) or "quiescent domains" (Hoffman et al., 2013).

The only feature commonly reported was a resistance to DNAse I digestion suggesting a compacted state. Which are the effectors and maintainers of this compaction is currently unknown. Moreover those regions frequently contact the nuclear envelope in what is called Lamina Associated Domains (LADs), a chromosomal compartment previously associated to gene silencing, although the silencing mechanism is still an open area of research (van Steensel & Belmont, 2017) suggesting that the principal mechanism for gene silencing in higher eukaryotes is not fully understood.

Euchromatin and its key components H3K4me3, H3K4me1, H3K9ac along with HATs (ENCODE Project Consortium, 2012; Filion, van Bemmel, Braunschweig, et al., 2010; Gerstein et al., 2010; Kharchenko et al., 2011a) were also captured by all models.

Here the interesting result was that in all organisms there are two differentiated types of euchromatin corresponding to housekeeping genes and regulated genes.
It was previously known that different promoter characteristics such a as nucleosome positioning, DNA motif composition and patterns of transcription initiation differed among these gene categories, (Hoskins et al., 2011; Lenhard, Sandelin, & Carninci, 2012).

Moreover their structure is also different. In Drosophila, housekeeping genes are short (typically less than 2 kb) while regulated genes are longer (typically more than 3 kb) and generally present a large first intron (Filion, van Bemmel, Ulrich, et al., 2010a; Kharchenko et al., 2011a; modENCODE Consortium et al., 2010a) . In the five chromatin types classification (Filion, van Bemmel, Ulrich, et al., 2010b), the distinction is captured by the Yellow and Red chromatin types.
Being in a different class means that both gene types have different chromatin composition. Housekeeping genes are covered in Yellow chromatin and are bound by MRG15, a chromodomain protein that binds the H3K36me3 histone mark present in the gene body. Regulated genes are covered in Red chromatin and are bound by the vast majority of the assayed proteins. The modENCODE classification has an equivalent of red chromatin called “state 3” (Kharchenko et al., 2011b; modENCODE Consortium et al., 2010b).
The state 3 chromatin covers the long intronic regions of active regulated genes, it is enriched in chromatin remodelers such as Brahma, corresponding with the detected high nucleosome turnover and high DNAse hypersensitivity. The chromatin of state 3 carries histone marks usually found on enhancers, such as H3K27ac, H3K4me1, H3K18ac and H3K36me1.

An additional observation, which was not appreciated before this large scale studies is the existence of regions along the genome bound by the majority of transcription factors or chromatin proteins tested.
These regions were called “HOTs” for Highly Occupied Target (ENCODE Project Consortium, 2012; Ernst & Kellis, 2011; Filion, van Bemmel, Braunschweig, et al., 2010; Kharchenko et al., 2011a).
The analysis of 21 functionally heterogeneous transcription factors binding patterns during development evidenced the existence of these regions (MacArthur et al., 2009), in which many of the transcription factors bound in the absence of their characteristic binding motif.
It was only after the genome-wide studies that this phenomenon was known to be prevalent in worms, flies and humans. Aside from the possibility of an artefact which would be very unlikely since HOTs are detected by ChIP and DamID, the functional relevance of these protein aggregates is disputed.
It was shown that HOTs can work as cell type-specific enhancers during Drosophila development by ectopic integration (Kvon, Stampfel, Yáñez-Cuna, Dickson, & Stark, 2012).

Another study analyzed the binding patterns of 65 TFs and 19 cofactors during Drosophila development finding that HOTs overlap housekeeping genes and enhancers of developmentally regulated genes (Kvon et al., 2012; Slattery et al., 2014).
In contrast when analyzing only the top 1% occupied targets in humans and C. elegans (Drosophila studies included the top 5%) no enhancer signatures were found but CpG-rich promoters of broadly expressed genes (Chen et al., 2014).

The phenomenon of non DNA-directed binding of transcription factors has been speculated to result from protein-protein interactions and highly accessible chromatin coupled to high nuclear concentration of TFs (Gerstein et al., 2010); (MacArthur et al., 2009).

It is now accepted that the separation between hetero- and euchromatin is a simplification. The richer complexity of chromatin arising from genome-wide studies suggests that different genes with different functions, expression patterns and chromosomal context have evolved different mechanisms to ensure their proper regulation.

This chromatin state concept has added detail to our view of chromatin organization in relation to gene activity, but it has also opened additional questions and defied the generality of the histone code.
The finding that the majority of silent genes are devoid of known proteins and histone marks and the presence of marks associated with silent heterochromatin in active regions are puzzling.

Moreover the discovery of HOTs as loci massively occupied by transcription factors lacking their recognition sequences further complicates the understanding of the eukaryotic logic of transcription.

6. Three-dimensional genome folding

The development of chromosome conformation capture techniques (called 3C, 4C, 5C and Hi-C) has allowed to interrogate the spatial contact frequency between chromatin fragments in the nucleus. In its genome-wide version, Hi-C, the frequencies of contact between all genome loci are interrogated in a single experiment.
The principle of the technique is based on the fact that chromosome fragments that lie closely in space will preferentially be ligated after a formaldehyde fixation followed by a genome digestion and religation in diluted conditions. Sequencing will quantify the relative amount of contacts between any two chromosomal contacts genome wide (Lieberman-Aiden et al., 2009).

While this is a recent and active area of research some general observations seem relevant to gene regulation and chromatin composition.
The first study was carried in a human lymphoblastoid cell line and achieved megabase scale resolution. A plaid pattern in the normalized contact matrix evidenced the existence of two types of chromatin domains (arbitrarily referred to as A and B) who interacted preferentially. Contacts between A compartments or among B compartments were much more frequent than among A-B compartments.
The authors compared the relative enrichment of several features within each compartment type and found an enrichment of genes, mRNA expression, DNAse I accessibility and H3K36 trimethylation in the A vs B compartment.
A similar study took advantage of the smaller genome size of Drosophila to obtain better genome resolution.

The result appeared as a linear partition of the genome in well demarcated self-interacting domains of size ranging from 10 to 500 kb (Sexton et al., 2012) separated by regions of low frequency of contacts. This highly interacting domains were later called

topologically associating domains (TADs) and were shown to be present in Drosophila mouse and humans (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012).
Moreover those TADs were shown to correspond to homogeneous chromatin states in Drosophila (Sexton et al., 2012). Recent studies are attempting to make sense of the functional implications of this genome organization for chromatin biology in several organisms and biological phenomenons (Pombo & Dillon, 2015).

In Drosophila, A/B compartments are less clear than in mammals, nevertheless TADs and chromatin states seem to have a tight correspondence. While silenced genes (from Blue Polycomb and Black states) and regulated active genes from Red state are found inside TADs with homogeneous composition, housekeeping genes of the Yellow type accumulate in inter-TAD regions with short range contacts. Moreover this non-TAD like regions containing active housekeeping genes appear to contact among successive inter-TADs, a fact that has suggested to create active compartments in the nuclear space and even drive the establishment of the TAD / inter-TAD / TAD structure (Eagen, Hartl, & Kornberg, 2015; Sexton et al., 2012; Ulianov et al., 2016).

Indeed the research paper presented in this thesis contributes to the idea that spatial clustering of housekeeping promoters might be a way to compartmentalize and recycle transcriptional activators (Corrales et al., 2017).

7. Methods: Thousands of Reporters in Parallel (TRIP)

Taking advantage of the detailed knowledge of protein composition genome-wide in Drosophila Kc167 cells and the maturity of multiplex sequencing technology, we used a massively parallel approach to test position effects. Much in the vein of position effect variegation in Drosophila, we wanted to ask how the chromatin composition and structure affect the transcription of a reporter gene integrated at thousands different ectopic loci.
In order to achieve this throughput we combined reporter library barcoding, random integration via sleeping beauty transposase and Illumina sequencing in an improved adaptation of TRIP (Akhtar et al., 2014).

To obtain a transcriptional readout that does not depend on the promoter or gene sequence of the reporter we decided to generate barcoded copies of a gene reporter with an active Drosophila promoter driving GFP expression inside a transposable cassette. In a barcoding PCR, each reporter plasmid will be tagged by a random stretch of 21 nucleotides (the barcode) upstream of the polyadenylation signal of the gene, generating unique reporters in each PCR cycle.
In this way each reporter will contain a unique sequence that will allow to map the genomic location of the insertion and to quantify the number of transcripts it generates by sequencing. Moreover we took advantage of Gibson cloning to re-circularize the plasmids avoiding the use of restriction enzymes that may cut in the barcodes and limit our promoter choices. A reporter library with at least 1 million different colonies is constructed to minimize the amount of barcode collisions, i.e. different reporters having the same barcode.

To insert the library of reporters homogeneously along the Drosophila genome, cultures of Drosophila Kc167 cells were co-electroporated with both the library and an expression plasmid containing Sleeping Beauty transposase under a heat-shock inducible promoter. The electroporation introduces approximately 1-5 reporters per cell, which minimizes the effect of the integrations in the cell physiology while targeting thousands of locations (in different cells) in a single experiment.

After electroporation the cells are transferred from 25C to 37C, which is the temperature causing a heat shock leading in the expression of the Sleeping beauty transposase and its optimal transposition activity. Six high temperature expositions of 2 hours in 3 days have proven to give the best integration survival ratio in our hands. The transposase cuts the transposon-specific sequences flanking each reporter gene and pastes it to a random region of the Drosophila genome (Figure 1a).

A control experiment without transposase inserts the reporters transiently. Their GFP fluorescence allows monitoring the dilution of the non integrated plasmid in order to stop the experiment when only integrated plasmids remain. This minimizes the molecular amplification of unintegrated plasmids in successive steps, which are uninformative and are present at high number at the start of the experiment.

When transient plasmids are diluted from the population, we gather the cells and we extract the genomic DNA and the RNA.
The genomic DNA serves two purposes: on the one hand to map the insertions to their genomic location and on the other to correct the measure of expression by the number of copies of each integrated reporter (variable due to cell doublings after integration). RNA is extracted, mRNA selected and GFP mRNAs selectively reverse transcribed.

Since the barcode is present before the polyadenylation signal of the GFP open reading frame, the number of transcripts corresponding to each integrated reporter can be quantified by RNA-seq.

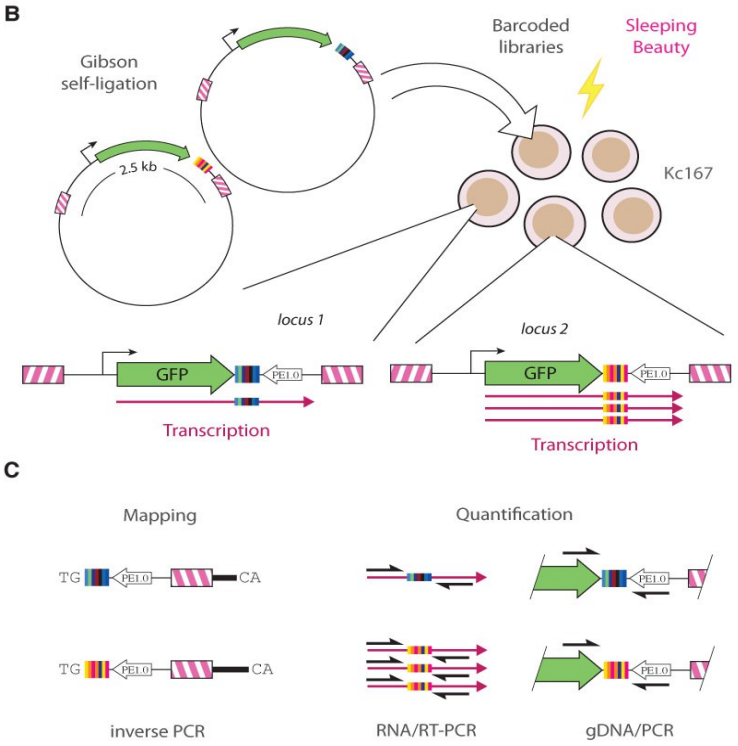

Figure 1. Overview of the TRiP experiment. *a. The reporter plasmids show the transposon sequences recognized by Sleeping Beauty (pink boxes) and the barcode after the GFP coding sequence (rainbow boxes), the different colors in the barcodes represent their different sequences. Schematized is the co-electroporation of the barcoded libraries to the Kc167 population, and the putative result of the integration in a non permissive transcription locus (locus 1) and to a permissive transcription locus (Locus 2) resulting in different amount of barcoded mRNAs from each integration. b. Mapping represents the NlaIII restriction fragments from two different integrations showing the different lengths of the genomic DNA (gDNA) to the right of the integration and the cohesive ends allowing to circularize the barcode-gDNA fragment by ligation possibiliting the inverse PCR and the further sequencing. Quantification shows the location of the primers used to amplify and sequence a low expressed integration (top) and a highly expressed one (bottom) both in gDNA and RNA to obtain corrected measures of expression.*

In order to map the integrations to their genomic loci we digest the genomic DNA with NlaIII, a restriction enzyme that will cut the genome at its recognition sequence CATG, which occurs on average each 347 bp, and inside our reporter that contains this restriction sequence immediately upstream of the barcode.

For a better understanding of a putative fragment resulting from the genomic restriction of the reporters, refer to panel (Figure 1b). The fragment contains the barcode on the left side and a piece of the Drosophila genome adjacent to the integration on the right side.
A ligation of this fragments followed by an inverse PCR will result in molecules that can be pair-end sequenced and will give the barcode in one read and the genomic sequence in the complementary read, which will be afterwards mapped against the reference genome to obtain the location. In summary, after the mapping procedure we obtain a table containing all the barcodes associated with their genomic locations.
Further details relating to the experimental conditions and the bioinformatic pipeline implementing the quality controls can be read in the Methods section of the attached paper.

To obtain the expression level associated with each barcode and thus with a uniquely mapped genomic location, we PCR amplify in the same conditions the cDNAs resulting from the reverse transcription and the genomic DNA. After sequencing the number of barcodes obtained from the extracted RNA, will be corrected by the number of times the barcode was present in the genomic DNA to have a normalized estimation of the RNA (Figure 1b, Quantification).

In order to process and analyze the data, we devised a computational pipeline combining in-house scripts and software developed in the lab to control and correct for the sequencing errors in the barcodes and R scripts for the statistical analysis and figure generation. The pipeline is open source and can be consulted at https://hub.docker.com/r/histonemark/tripeline/.

Each experiment contains a particular combination of promoter-GFP reporter and we present here the results obtained from four different TRIPs containing each a Drosophila promoter selected with the only condition of being active in a transient transfection. Together, these experiments allowed us to test the chromatin effects at › 85,000 loci. Figure 2 shows the expression results along the chromosomes.

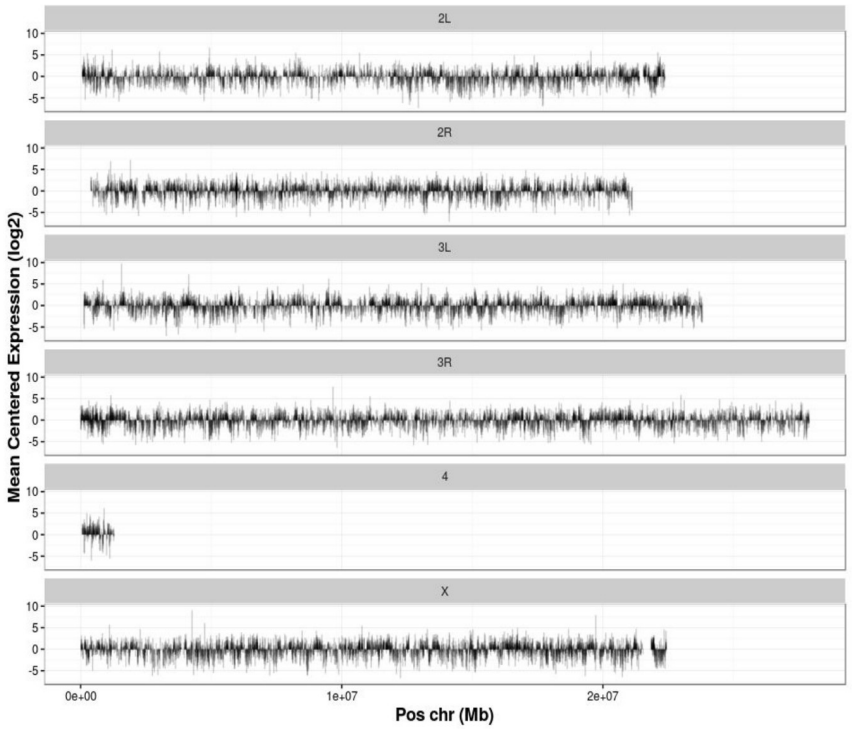

Figure 2. ***Relative expression along chromosome arms***. *Each row represents a Drosophila chromosome in linear coordinates (x-axis). The height of each bar represents the log2 mean centered expression of a singular integrated reporter. The alternance of domains of high and low expression along the chromosome are visible as darker aggregations of consistently expressed reporters.*

8. Research publication: Clustering of Drosophila housekeeping promoters facilitates their expression

8.1 Aims of the study

In order to understand the effects of chromatin composition and structure in the expression of a fixed transcription unit, we conducted 5 TRIP experiments.

To minimize the effects of promoter dependence we chose 4 random housekeeping promoters with the sole requisite of being active when transiently electroporated in Drosophila Kc167 cells.

An additional TRIP experiment was conducted with a promoterless library in order to control for reporter expression not arising from the integrated promoter.

The aim was then to relate the expression levels of the same integrated reporter with the characteristics of the receiving loci, in terms of epigenetic marks, chromatin composition, nucleosome accessibility and contact frequency with other elements in the nucleus.

Corrales M, Rosado A, Cortini R, van Arensbergen J, van Steensel B, Filion GJ. Clustering of Drosophila housekeeping promoters facilitates their expression. Genome Res. 2017 Jul;27(7):1153–61. DOI: 10.1101/gr.211433.116

9. Discussion and perspectives

One of the most important caveats in the field of chromatin and gene regulation is that the majority of results are generalizations of correlative evidence. With the advent of genome-wide studies it has become easier to test the combinatorial degree of nuclease accessibility, histone marks, chromatin proteins or transcription factors. These combinations are then correlated with the underlying transcriptional activity in order to identify associations between factors and activity. The patterns that emerge can then be tested with more direct molecular techniques.

The difficulty is to carry these direct molecular studies at different loci, with heterogeneous complexes in which all the proteins might have compounded effects. Moreover it is difficult to tease apart sequence from chromatin effects: a particular regulatory sequence may have context requirements or result in different outcomes depending on the chromatin proteins or marks that bind it. In this regard integrating the same reporter sequence at ectopic genomic locations excludes sequence effects (provided that one controls the effect of the integration on the composition and chromatin structure) and allows the inspection of multiple context combinations on the same sequence.

With a method such as TRIP (Akhtar et al., 2014), expression of the inserted reporters with respect to the composition of the landing locus is closer to causality than endogenous correlations, since the sequence is fixed and the reporters are ectopically relocated.

In order to simplify the extracted conclusions, we have used housekeeping promoters which are known to be shorter and to be less dependent on regulatory elements such as distal enhancers (Zabidi et al., 2015). Having observed consistent effects in four different promoters with diverse endogenous activities lowers the possibility of sequence-specific effects.

Using the extensive amount of chromatin profiles (Filion, van Bemmel, Braunschweig, et al., 2010; modENCODE Consortium et al., 2010a); (Kharchenko et al., 2011b; modENCODE Consortium et al., 2010b) and chromosome conformation data (L. Li et al., 2015) in our cell line we have been able to study their particular contribution to the expression of our integrated reporters.

The first interesting result was the observation that the reporters expression profile along chromosome arms was organized in successive domains of permissive or refractory to transcription. This confirms the observation that co-regulated genes cluster together in chromosomes (Hurst et al., 2004; Vinogradov, 2004) and is compatible with the putative three-dimensional genome organization in A and B compartments in vertebrates or homogeneous epigenetic three-dimensional domains (Sexton et al., 2012). It also shows that chromatin effects happen away from centromeres and telomeres.

Reporters landing in Black chromatin (the unmarked and unoccupied state) have the lowest expression levels suggesting that this type of chromatin has an active role in gene silencing. It is also true that this chromatin type shows a huge expression variance which might indicate a more heterogeneous composition than anticipated.

The only feature associated with this chromatin type is the tendency to contact the nuclear lamina (Filion, van Bemmel, Braunschweig, et al., 2010), but not all black domains are lamina-associated, and lamina association as a silencing mechanism is not fully understood, especially in *Drosophila* where the levels of H3K9me at the lamina are depleted (Filion, van Bemmel, Braunschweig, et al., 2010).

It could also be possible that the close proximity to the lamina brings those loci away from the active genes (which tend to be localized towards the cell interior and associated in the so called "transcription factories" (Razin et al., 2011) and therefore disfavors contacts with active promoters and terminators, but again the mechanism underlying the tethering to the lamina are proposed to rely on HP1 and H3K9 methylation (van Steensel & Belmont, 2017), both depleted in *Drosophila* Black chromatin (Filion, van Bemmel, Braunschweig, et al., 2010).
It is also shocking that other chromatin states that establish "closed" silencing chromatin structures rely on active mechanisms for their establishment and maintenance, resulting in our ability to detect HP1, PRC complexes, histone deacetylases and several histone marks associated with the activity of the chromatin state.

Integrations in Blue chromatin, characterized by PcG proteins and H3K27me3, had a median expression close to Black chromatin confirming the role of polycomb mediated silencing.

The case of Green chromatin, typically covered by HP1 and H3K9me2/3 is more interesting. The Green chromatin state is found at pericentromeric regions, the fourth chromosome and scattered along chromosome arms. Reporters integrated in this chromatin type show a higher level of expression than expected for this chromatin composition, normally associated with silencing.
While the integrations in centromeric heterochromatin are severely silenced confirming the centromeres as silencing regions, this is not the case in chromosome arms or the 4th chromosome.

In the modENCODE chromatin classification, state 8 captures these regions along the chromosome arms with Su(var)3-9, HP1a and H3K9me2/3 (calling them heterochromatin-like) (Kharchenko et al., 2011b; modENCODE Consortium et al., 2010b).

This state shows low levels of RNA polymerase II, but our integrations are mainly active in those regions. For instance out of 216 integrations landing in green chromatin along the 2L chromosome arm, 153 are expressed while 63 are repressed. This evidences the fact that heterochromatin composition at the centromeres and chromosome arms has different silencing capacity for our housekeeping reporters.

Moreover, the majority of integrated reporters in chromosome 4 are active, out of 236 integrations 182 show high expression and 54 low. This chromosome is considered heterochromatic by cytological and chromatin composition standards (green chromatin state), although endogenous expression is detected from several loci. In one study the authors found that HP1a repressed transcription on chromosome 4 via eggless (the drosophila homolog of SETDB1) H3K9 methylase. The enzyme targeted preferentially non-ubiquitously expressed genes and therefore genes containing a different core promoter structure than our integrated reporters (Lundberg, Stenberg, & Larsson, 2013).

Additionally a chromosome 4 specific histone acetyltransferase, painting of four (Pof) is known to bind active endogenous transcripts from the chromosome increasing the expression from the locus (Johansson, Stenberg, Allgardsson, & Larsson, 2012). Taken together these results might explain why our integrations are active on chromosome 4.

This suggests to be cautious when generalizing chromatin effects, several combinations of proteins can have evolved to give different outputs in different contexts, resulting in the inability to extrapolate chromatin effects without knowing all the players involved (such as eggless and Pof in this case) and a greater mechanistic understanding of their individual contributions in heterogeneous complexes.

Moreover there are active genes in heterochromatin that rely on the typical chromatin proteins for their expression suggesting that complex relationships between regulatory sequences and proteins result in different outcomes (Dimitri, Corradini, Rossi, & Vernì, 2005).

Yellow chromatin is the endogenous chromatin type of the promoters used in the experiment, the active euchromatin composition characteristic of housekeeping genes. Unsurprisingly, the reporters integrated in this chromatin environment have the highest median expression and the lowest variance.

Does an active Yellow gene remain active when transplanted to the Red environment?
The question arises naturally from the observed separation of euchromatin in two different types which may be functionally equivalent since both share many of the associated active marks and proteins. Our results provide a test for this question.
The median expression of the reporters integrated in Red is more similar to Green than to Yellow.
This suggests that the mechanisms fulfilling optimal expression for housekeeping genes are different from the ones activating regulated genes. This result is largely explained by the finding of paradoxical domains and the conclusions of a study demonstrating that housekeeping promoters do not respond to developmental enhancers (Zabidi et al., 2015).

Reporters integrated in paradoxical domains are silent. Paradoxical regions have all the characteristics of euchromatin: they are enhancer-rich (both housekeeping and developmental), accessible, covered in active marks and chromatin remodelers. However, they do not contact promoters and terminators of other active genes.

Most of these paradoxical domains occur in the body of long genes, generally depleted of contacts with other active gene regulatory regions, despite being active themselves. For those active long genes, the regions that contact other active regions are only their promoters and terminators, with the body of the gene excluded from these contacts.

This observation suggests that the limiting factor for expression of the reporters might be assembled and released in the contacting promoters and terminators of active genes, a situation reminiscent of the described "transcription factories" (Arnold et al., 2017; Rieder, Trajanoski, & McNally, 2012), spatial regions of the nucleus where active promoters colocalize. This effect is confirmed in the computational simulations included in the research paper attached.

To test this idea, we compared the ability of hundreds of chromatin proteins, histone marks, enhancer contacts and chromatin state segmentations to predict integrated reporter transcription against the single predictors: the amount of contacts with active promoters or the amount of contacts with active terminators. Surprisingly, the predictive power of this metric alone was almost twice the one of all the chromatin features. This suggests that housekeeping genes take advantage of contacts with other active regions to maximize their expression. This finding is in complete agreement with two observations.

First, in *Drosophila*, the majority of housekeeping genes are arranged in tight clusters of several genes along chromosome arms. The positive effect of this clustering was revealed in a study in which inserting randomly fragmented sequences in a housekeeping promoter driven reporter was tested for their enhancer capacity. Genomic sequences arising from 5' UTRs and proximal promoters were enhancers of housekeeping promoters but failed to do so when developmental promoters were used (Zabidi et al., 2015).

Terminator sequences were not transactivating housekeeping promoters in the plasmid. Whereas promoters can recruit transcriptional activators, isolated terminators in a plasmid are not regions of polymerase or elongation factor release, since they are not transcribed. This suggests that recruitment of transcriptional activators by the transcriptional machinery is a likely mechanism to enhance the expression of the reporter gene.

The second observation came from three dimensional chromatin conformation studies in *Drosophila*. The distribution of contacts showed a pattern of chromosome-wide alternation of topologically associated domains separated by small domains. These inter-TAD domains invariably contained the mentioned clusters of housekeeping genes, and they showed a tendency to interact among themselves in space at the exclusion of the TADs. This demonstrates that not only housekeeping genes are together in two dimensional clusters but they also interact in the nuclear space. This observation even prompted the hypothesis that the spatial segregation of active and inactive chromosomal regions detected by Hi-C (equivalent to mammalian A/B compartments) is driven by the contacts among clusters of housekeeping genes (Rowley et al., 2017; Ulianov et al., 2016) .

The spatial organization of housekeeping genes is a selected feature, as demonstrated by the higher conservation of housekeeping consecutive genes pairs between human and mouse than expected by chance. Moreover, the number of chromosomal breakages inside the clusters are significantly smaller than outside those clusters (Lercher, Urrutia, & Hurst, 2002; Singer, Lloyd, Huminiecki, & Wolfe, 2005).

What are the selected benefits of clustering is as of today a matter of speculation. Two non mutually exclusive hypotheses can be proposed. First, the two-dimensional and three-dimensional spatial segregation of constantly active regions could perhaps help reduce the interference with the more delicate regulation of spatiotemporally regulated genes, generally isolated and controlled by combinations of distal enhancers (Lercher et al., 2002). Second, the observation that transcribing a gene facilitates the transcription of the others nearby could also represent an advantage to bring housekeeping genes together (Akhtar et al., 2014; Ebisuya, Yamamoto, Nakajima, & Nishida, 2008).

Of course for an organism to develop and maintain homeostasis it has to conditionally express genes spatiotemporally (in response to a certain stimulus, restricted to a particular time in development or tissue). Those genes with a spatiotemporal regulation rely on more sophisticated mechanisms to control their behaviour, as they can be activated from different enhancers and with different combinations of transcription factors. Moreover, contrary to housekeeping genes, regulated genes show a low transcription level when ectopically introduced in a plasmid, evidencing the need for additional regulatory inputs to be expressed at correct levels. In *Drosophila* a recent study has shown that core promoter motifs determine the ability to respond to a developmental enhancer (Zabidi et al., 2015) suggesting that other promoter architectures might impose different context requirements.

Repeating our study with different promoter architectures will yield position effects that could potentially unravel the chromatin and structural context requirements for the correct regulation of regulated genes. In particular it might shed light on the question of the enhancer-promoter compatibility determinants. Another recent study demonstrated that enhancer responsiveness of different sequences could differ by several orders of magnitude, and different levels of responsiveness were associated with genes of different functions (Arnold et al., 2017).

This suggests that the context dependency plays an important role at promoters but also at enhancers, potentially determining their coordinated function.

Moreover it is dubious that core promoter sequences would be the only ones controlling the ability of a gene to be expressed in a certain context. Promoters are conserved beyond the core sequence motifs and a clear correlation between activity in each chromatin type and promoter sequence is not evident (personal unpublished observation).

It will therefore be very interesting to increase the number of TRIP experiments with different promoter architectures and combinations thereof.

Bacteria exhibit a simple yet elegant mechanism to deal with the task of coordinating regulation. The proteins involved in the same effector pathway are frequently organized in co-transcribed regions, the operons, that are simultaneously regulated. As discussed in the introduction the system relies on transcriptional activators capable of recruiting RNA polymerase II to the target promoter. Repressors, in contrast, occlude the recognition sequences at promoters impeding RNA polymerase assembly. Ultimately the fate of transcription of a bacterial gene is determined by the concentration and activity of transcriptional activators *versus* repressors, and their combinations.

In eukaryotes, nucleosome packaging of the genetic fiber prevents transcription factor sequence recognition and enzymatic access to DNA and the genes are individual entities. Transcription factors and polymerases rely on accessory mechanisms that disrupt the stability or position of nucleosomes to gain access to the underlying sequence.

The question that follows this reasoning is how the molecular machinery responsible for granting access to functional elements are themselves recruited to the target regulatory regions in the first place.

The consensus points to special transcription factors able to recognize their binding sequences in nucleosomal conformation. Those transcription factors are called pioneer factors and are able to bind partial motifs of their target sequences at the nucleosome surface (Soufi et al., 2015). A combination of three pioneer factors (Oct4, Sox2 and Klf4) and the transcription factor c-myc were shown to be capable of reprogramming human and mouse somatic fibroblasts to induced pluripotent stem cells *in vitro*, an artificial reversion of the cell fate evidencing the preponderant role that transcription factors and DNA sequence have in the regulation of expression programs (Takahashi & Yamanaka, 2006).

These factors would be responsible for the establishment of transitory chromatin competent states by recruiting chromatin remodelers and histone variant incorporation (Z. Li et al., 2012). These competent states will result in repression or activation by the action of different combinations of sequence-specific transcription factors, polymerase II recruitment and other activators or repressors present in the specific cell type at the appropriate time. As is the case with the previously discussed HOX genes in *Drosophila*, transcriptional states initiated by transcription factor binding are maintained in absence of the initial stimulus via the establishment of chromatin states by Polycomb-mediated silencing and Trithorax activation. Surprisingly, a recent study showed that transcription of genes temporally regulated during fly and worm development occurs in the absence of active histone modifications (Pérez-Lluch et al., 2015) casting doubts on the instructive role of the activating marks in developmental control.

Therefore, dynamic changes in chromatin states need to be accounted for to fully understand the roles of chromatin on gene regulation. This is especially evident in the case of development and in the response to external and internal stimuli.

Moreover, an interdependent network of hundreds of proteins respond to internal and extracellular signals involved in dynamic chemical reactions, ultimately resulting in the quantitative expression and regulation of genes. Studying in isolation any of the components of the network is difficult, because those proteins, histone marks and nucleosome organizations are frequently organized in complexes that show abundant chemical cross-talk among them and with other components of the regulatory network, many of them redundant in their functions. The mutation approach, while needed, is not guaranteed to disentangle additive effects contributed by each of the members of the "chromatin system".

TRIP results are a snapshot of this system in a particular cell type. It has allowed us to test the effect of the recently described chromatin states on housekeeping promoters at an unprecedented resolution. We were able to appreciate the particular regulation of housekeeping genes, in which the effect of enhancers seems to be less preponderant than regulated genes. The realization that the integrated reporters are maximally expressed when contacting active promoters and terminators allowed us to formulate a tentative hypothesis of why housekeeping genes naturally show a tendency to cluster along the chromosome and in the nuclear space. We anticipate that using different promoter architectures, TRIP may reveal determinants of promoter-enhancer compatibility, a question of high importance to understand how more strictly regulated genes achieve the spatiotemporal specific patterns of expression and move one step forward in the quest of understanding the details surrounding the logic of transcription in eukaryotes.

10. References

Akhtar, W., Pindyurin, A. V., de Jong, J., Pagie, L., Ten Hoeve, J., Berns, A., … van Lohuizen, M. (2014). Using TRIP for genome-wide position effect analysis in cultured cells. Nature Protocols, 9(6), 1255–1281. https://doi.org/10.1038/nprot.2014.072

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of rna synthesis. Proceedings of the National Academy of Sciences of the United States of America, 51, 786–794. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14172992

Allfrey, V. G., & Mirsky, A. E. (1962). Evidence for the complete DNA-dependence of RNA synthesis in isolated thymus nuclei. Proceedings of the National Academy of Sciences of the United States of America, 48, 1590–1596. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/13860598

Allshire, R. C., & Ekwall, K. (2015). Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. Cold Spring Harbor Perspectives in Biology, 7(7), a018770. https://doi.org/10.1101/cshperspect.a018770

Anderson, J. M., Erickson, R. C., Munro, D. A., Spofford, W. R., West, G. C., & Evenden, F. G. (1968). Report of Committee on Conservation, 1967. The Auk, 85(1), 117–126. https://doi.org/10.2307/4083629

Aparicio, O. M., Billington, B. L., & Gottschling, D. E. (1991). Modifiers of position effect are shared between telomeric and silent mating-type loci in S. cerevisiae. Cell, 66(6), 1279–1287. https://doi.org/10.1016/0092-8674(91)90049-5

Arnold, C. D., Zabidi, M. A., Pagani, M., Rath, M., Schernhuber, K., Kazmar, T., & Stark, A. (2017). Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. Nature Biotechnology, 35(2), 136–144. https://doi.org/10.1038/nbt.3739

Avery, O. T., Macleod, C. M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. The Journal of Experimental Medicine, 79(2), 137–158. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19871359

Baker, W. K. (1968). Position-effect variegation. Advances in Genetics, 14, 133–169. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4884780

Beadle, G. W., & Tatum, E. L. (1941). Genetic Control of Biochemical Reactions in Neurospora. Proceedings of the National Academy of Sciences, 27(11), 499–506. https://doi.org/10.1073/pnas.27.11.499

Becker, P. B., & Workman, J. L. (2013). Nucleosome Remodeling and Epigenetics. Cold Spring Harbor Perspectives in Biology, 5(9), a017905–a017905. https://doi.org/10.1101/cshperspect.a017905
Benayoun, B. A., Pollina, E. A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E. D., … Brunet, A. (2014). H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell, 158(3), 673–688. https://doi.org/10.1016/j.cell.2014.06.027

Blander, G., & Guarente, L. (2004). The Sir2 Family of Protein Deacetylases. Annual Review of Biochemistry, 73(1), 417–435. https://doi.org/10.1146/annurev.biochem.73.011303.073651

Boveri, T. (1904). Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns. Von Theodor Boveri. https://doi.org/10.5962/bhl.title.28064

Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D., & Broach, J. R. (1993). Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. Genes & Development, 7(4), 592–604. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8458576

Brenner, S., Jacob, F., & Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature, 190, 576–581. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20446365

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., & Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell, 84(6), 843–851. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8601308

Burgess, R. R., Travers, A. A., Dunn, J. J., & Bautz, E. K. F. (1969). Factor Stimulating Transcription by RNA Polymerase. Nature, 221(5175), 43–46. https://doi.org/10.1038/221043a0

Cairns, B. R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., … Kornberg, R. D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell, 87(7), 1249–1260. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8980231

Carey, M., Li, B., & Workman, J. L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. Molecular Cell, 24(3), 481–487. https://doi.org/10.1016/j.molcel.2006.09.012

Caskey, C. T., & Leder, P. (2014). The RNA code: Nature's Rosetta Stone. Proceedings of the National Academy of Sciences, 111(16), 5758–5759. https://doi.org/10.1073/pnas.1404819111

Caspersson, T. (1947). The relations between nucleic acid and protein synthesis. Symposia of the Society for Experimental Biology, (1), 127–151. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20257022

Chadwick, B. P. (2003). Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. Human Molecular Genetics, 12(17), 2167–2178. https://doi.org/10.1093/hmg/ddg229

Chargaff, E. (1950). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia, 6(6), 201–209. https://doi.org/10.1007/bf02173653

Chen, R. A.-. J., Stempor, P., Down, T. A., Zeiser, E., Feuer, S. K., & Ahringer, J. (2014). Extreme HOT regions are CpG-dense promoters in C. elegans and humans. Genome Research, 24(7), 1138–1146. https://doi.org/10.1101/gr.161992.113

Clapier, C. R., & Cairns, B. R. (2009). The biology of chromatin remodeling complexes. Annual Review of Biochemistry, 78, 273–304. https://doi.org/10.1146/annurev.biochem.77.062706.153223

Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nature Reviews. Molecular Cell Biology, 18(7), 407–422. https://doi.org/10.1038/nrm.2017.26

Conrad, T., & Akhtar, A. (2012). Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nature Reviews. Genetics, 13(2), 123–134. https://doi.org/10.1038/nrg3124

Corrales, M., Rosado, A., Cortini, R., van Arensbergen, J., van Steensel, B., & Filion, G. J. (2017). Clustering of Drosophila housekeeping promoters facilitates their expression. Genome Research, 27(7), 1153–1161. https://doi.org/10.1101/gr.211433.116

Correns, C. E. (1924). G. Mendel's Regel über das Verhalten der Nachkommenschaft der Rassenbastarde. In Gesammelte Abhandlungen zur Vererbungswissenschaft aus Periodischen Schriften 1899–1924 (pp. 9–18). https://doi.org/10.1007/978-3-642-52587-2_2

Dahm, R. (2008). Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. Human Genetics, 122(6), 565–581. https://doi.org/10.1007/s00439-007-0433-0

Dambacher, S., Hahn, M., & Schotta, G. (2013). The compact view on heterochromatin. Cell Cycle , 12(18), 2925–2926. https://doi.org/10.4161/cc.26179

David Allis, C., Caparros, M.-L., Jenuwein, T., & Reinberg, D. (2015). Epigenetics. Cold Spring Harbor Laboratory Press. Retrieved from https://market.android.com/details?id=book-ftginwEACAAJ

de la Cruz, X., Lois, S., Sánchez-Molina, S., & Martínez-Balbás, M. A. (2005). Do protein motifs read the histone code? BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology, 27(2), 164–175. https://doi.org/10.1002/bies.20176

Delmas, V., Stokes, D. G., & Perry, R. P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. Proceedings of the National Academy of Sciences of the United States of America, 90(6), 2414–2418. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8460153

de Vries, H. (1900). Sur la loi de disjonction des hybrides. Retrieved from https://books.google.com/books/about/Sur_la_loi_de_disjonction_des_hybrides.html?hl=&id=noyvHAAACAAJ

Dimitri, P., Corradini, N., Rossi, F., & Vernì, F. (2005). The paradox of functional heterochromatin. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology, 27(1), 29–41. https://doi.org/10.1002/bies.20158

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., …

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

Durrin, L. K., Mann, R. K., Kayne, P. S., & Grunstein, M. (1991). Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. Cell, 65(6), 1023–1031. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2044150

Eagen, K. P., Hartl, T. A., & Kornberg, R. D. (2015). Stable Chromosome Condensation Revealed by Chromosome Conformation Capture. Cell, 163(4), 934–946. https://doi.org/10.1016/j.cell.2015.10.026

Ebisuya, M., Yamamoto, T., Nakajima, M., & Nishida, E. (2008). Ripples from neighbouring transcription. Nature Cell Biology, 10(9), 1106–1113. https://doi.org/10.1038/ncb1771

Eissenberg, J. C., Morris, G. D., Reuter, G., & Hartnett, T. (1992). The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. Genetics, 131(2), 345–352. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1644277

Elgin, S. C. R., & Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harbor Perspectives in Biology, 5(8), a017780. https://doi.org/10.1101/cshperspect.a017780

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57–74. https://doi.org/10.1038/nature11247

Englesberg, E., Irr, J., Power, J., & Lee, N. (1965). Positive control of enzyme synthesis by gene C in the L-arabinose system. Journal of Bacteriology, 90(4), 946–957. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/5321403

Ernst, J., & Kellis, M. (2011). Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome. In Lecture Notes in Computer Science (pp. 53–53). https://doi.org/10.1007/978-3-642-20036-6_6

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., … Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature, 473(7345), 43–49. https://doi.org/10.1038/nature09906

Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., … van Steensel, B. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell, 143(2), 212–224. https://doi.org/10.1016/j.cell.2010.09.009

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature, 391(6669), 806–811. https://doi.org/10.1038/35888

Fischle, W. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes & Development, 17(15), 1870–1881.

Geiduschek, E. P., Snyder, L., Colvill, A. J., & Sarnat, M. (1966). Selective synthesis of T-even bacteriophage early messenger in vitro. Journal of Molecular Biology, 19(2), 541–547. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/5338857

Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., … Waterston, R. H. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science, 330(6012), 1775–1787. https://doi.org/10.1126/science.1196914

Gottschling, D. E., Aparicio, O. M., Billington, B. L., & Zakian, V. A. (1990). Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription. Cell, 63(4), 751–762. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2225075

Griffith, F. (1928). The Significance of Pneumococcal Types. The Journal of Hygiene, 27(02), 113–159. https://doi.org/10.1017/s0022172400031879

Gros, F., Hiatt, H., Gilbert, W., Kurland, C. G., Risebrough, R. W., & Watson, J. D. (1961). Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli. Nature, 190, 581–585. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/13708983

Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., & Young, B. (1998). The Functional and Regulatory Roles of Sigma Factors in Transcription. Cold Spring Harbor Symposia on Quantitative Biology, 63(0), 141–156. https://doi.org/10.1101/sqb.1998.63.141

Grossniklaus, U., & Paro, R. (2014). Transcriptional Silencing by Polycomb-Group Proteins. Cold Spring Harbor Perspectives in Biology, 6(11), a019331–a019331. https://doi.org/10.1101/cshperspect.a019331

Grunberg-Manago, M., Ortiz, P. J., & Ochoa, S. (1956). Enzymic synthesis of polynucleotides I. polynucleotide phosphorylase of Azotobacter vinelandii. Biochimica et Biophysica Acta, 20, 269–285. https://doi.org/10.1016/0006-3002(56)90286-4

Grunstein, M., & Gasser, S. M. (2013). Epigenetics in Saccharomyces cerevisiae. Cold Spring Harbor Perspectives in Biology, 5(7), a017491–a017491. https://doi.org/10.1101/cshperspect.a017491

Gu, T., & Elgin, S. C. R. (2013). Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in Drosophila. PLoS Genetics, 9(9), e1003780. https://doi.org/10.1371/journal.pgen.1003780

Hannah, A. (1951). Localization and Function of Heterochromatin in Drosophila Melanogaster. In Advances in Genetics (pp. 87–125). https://doi.org/10.1016/s0065-2660(08)60232-1

Hargreaves, D. C., & Crabtree, G. R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Research, 21(3), 396–420. https://doi.org/10.1038/cr.2011.32

Hayashi, M., Hayashi, M. N., & Spiegelman, S. (1963). Restriction of in vivo genetic transcription to one of the complementary strands of DNA. Proceedings of the National Academy of Sciences of the United States of America, 50, 664–672. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14077495

Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M., & Grunstein, M. (1995). Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. Cell, 80(4), 583–592. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7867066

Henikoff, S., & Shilatifard, A. (2011). Histone modification: cause or cog? Trends in Genetics: TIG, 27(10), 389–396. https://doi.org/10.1016/j.tig.2011.06.006

Hergeth, S. P., & Schneider, R. (2015). The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Reports, 16(11), 1439–1453. https://doi.org/10.15252/embr.201540749
Hershey, A. D. (1952). INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. The Journal of General Physiology, 36(1), 39–56. https://doi.org/10.1085/jgp.36.1.39

Hewish, D. R., & Burgoyne, L. A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. Biochemical and Biophysical Research Communications, 52(2), 504–510. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4711166

Hilliker, A. J., & Sharp, C. B. (1988). New Perspectives on the Genetics and Molecular Biology of Constitutive Heterochromatin. In Stadler Genetics Symposia Series (pp. 91–115). https://doi.org/10.1007/978-1-4613-1037-2_4

Hirschhorn, J. N., Brown, S. A., Clark, C. D., & Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes & Development, 6(12A), 2288–2298. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1459453

Hoffman, M. M., Ernst, J., Wilder, S. P., Kundaje, A., Harris, R. S., Libbrecht, M., … Noble, W. S. (2013). Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Research, 41(2), 827–841. https://doi.org/10.1093/nar/gks1284

Horowitz, N. H., Bonner, D., Mitchell, H. K., Tatum, E. L., & Beadle, G. W. (1945). Genic Control of Biochemical Reactions in Neurospora. The American Naturalist, 79(783), 304–317. https://doi.org/10.1086/281267

Hoskins, R. A., Landolin, J. M., Brown, J. B., Sandler, J. E., Takahashi, H., Lassmann, T., … Celniker, S. E. (2011). Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Research, 21(2), 182–192. https://doi.org/10.1101/gr.112466.110

Howe, F. S., Fischl, H., Murray, S. C., & Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology, 39(1), 1–12. https://doi.org/10.1002/bies.201600095

Huang, R.-C. C., & Bonner, J. (1962). Histone, a suppressor of chromosomal RNA synthesis. Proceedings of the National Academy of Sciences, 48(7), 1216–1222. https://doi.org/10.1073/pnas.48.7.1216

Hurst, L. D., Pál, C., & Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene order. Nature Reviews. Genetics, 5(4), 299–310. https://doi.org/10.1038/nrg1319

Hurwitz, J. (2005). The Discovery of RNA Polymerase. The Journal of Biological Chemistry, 280(52), 42477–42485. https://doi.org/10.1074/jbc.x500006200

Imai, S., Armstrong, C. M., Kaeberlein, M., & Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature, 403(6771), 795–800. https://doi.org/10.1038/35001622

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology, 3(3), 318–356. https://doi.org/10.1016/s0022-2836(61)80072-7

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. Science, 293(5532), 1074–1080. https://doi.org/10.1126/science.1063127

Jeppesen, P., & Turner, B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell, 74(2), 281–289. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8343956

Johannsen, W. (1905). Arvelighedslærens elementer: forelæsninger holdte ved Københavns universitet. Retrieved from https://books.google.com/books/about/Arvelighedsl%C3%A6rens_elementer.html?hl=&id=1FI3AQAAMAAJ

Johansson, A.-M., Stenberg, P., Allgardsson, A., & Larsson, J. (2012). POF regulates the expression of genes on the fourth chromosome in Drosophila melanogaster by binding to nascent RNA. Molecular and Cellular Biology, 32(11), 2121–2134. https://doi.org/10.1128/MCB.06622-11

Jones, R. S., & Gelbart, W. M. (1993). The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. Molecular and Cellular Biology, 13(10), 6357–6366. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8413234

Kennison, J. A. (1995). The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. Annual Review of Genetics, 29, 289–303. https://doi.org/10.1146/annurev.ge.29.120195.001445

Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., … Park, P. J. (2011a). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature, 471(7339), 480–485. https://doi.org/10.1038/nature09725

Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., … Park, P. J. (2011b). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature, 471(7339), 480–485. https://doi.org/10.1038/nature09725

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell, 116(2), 259–272. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14744436

King, I. F. G., Francis, N. J., & Kingston, R. E. (2002). Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. Molecular and Cellular Biology, 22(22), 7919–7928. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12391159

Kingston, R. E., & Tamkun, J. W. (2014). Transcriptional regulation by trithorax-group proteins. Cold Spring Harbor Perspectives in Biology, 6(10), a019349. https://doi.org/10.1101/cshperspect.a019349

Klenov, M. S., Lavrov, S. A., Stolyarenko, A. D., Ryazansky, S. S., Aravin, A. A., Tuschl, T., & Gvozdev, V. A. (2007). Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. Nucleic Acids Research, 35(16), 5430–5438. https://doi.org/10.1093/nar/gkm576

Knezetic, J. A., & Luse, D. S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell, 45(1), 95–104. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3955658

Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science, 184(4139), 868–871. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4825889

Kornberg, R. D., & Thonmas, J. O. (1974). Chromatin Structure: Oligomers of the Histones. Science, 184(4139), 865–868. https://doi.org/10.1126/science.184.4139.865

Kornberg, T., & Gefter, M. L. (1970). DNA synthesis in cell-free extracts of a DNA polymerase-defective mutant. Biochemical and Biophysical Research Communications, 40(6), 1348–1355. https://doi.org/10.1016/0006-291x(70)90014-8

Kouzarides, T. (2007). Chromatin modifications and their function. Cell, 128(4), 693–705. https://doi.org/10.1016/j.cell.2007.02.005

Kvon, E. Z., Stampfel, G., Yáñez-Cuna, J. O., Dickson, B. J., & Stark, A. (2012). HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. Genes & Development, 26(9), 908–913. https://doi.org/10.1101/gad.188052.112

Lee, M. G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., … Shiekhattar, R. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science, 318(5849), 447–450. https://doi.org/10.1126/science.1149042

Lehman, I. R., Bessman, M. J., Simms, E. S., & Kornberg, A. (1958). Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. The Journal of Biological Chemistry, 233(1), 163–170. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/13563462

Lengyel, P., Speyer, J. F., & Ochoa, S. (1961). Synthetic polynucleotides and the amino acid code. Proceedings of the National Academy of Sciences, 47(12), 1936–1942. https://doi.org/10.1073/pnas.47.12.1936

Lenhard, B., Sandelin, A., & Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature Reviews. Genetics, 13(4), 233–245. https://doi.org/10.1038/nrg3163

Lercher, M. J., Urrutia, A. O., & Hurst, L. D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nature Genetics, 31(2), 180–183. https://doi.org/10.1038/ng887

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. Nature, 276(5688), 565–570. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/103000

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … 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

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C.-T., … Corces, V. G. (2015). Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. Molecular Cell, 58(2), 216–231. https://doi.org/10.1016/j.molcel.2015.02.023

Lin, R., Leone, J. W., Cook, R. G., & Allis, C. D. (1989). Antibodies specific to acetylated histones document the existence of deposition- and transcription-related histone acetylation in Tetrahymena. The Journal of Cell Biology, 108(5), 1577–1588. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2654136

Li, Z., Gadue, P., Chen, K., Jiao, Y., Tuteja, G., Schug, J., … Kaestner, K. H. (2012). Foxa2 and H2A.Z Mediate Nucleosome Depletion during Embryonic Stem Cell Differentiation. Cell, 151(7), 1608–1616. https://doi.org/10.1016/j.cell.2012.11.018

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. Nature, 251–260.

Lundberg, L. E., Stenberg, P., & Larsson, J. (2013). HP1a, Su(var)3-9, SETDB1 and POF stimulate or repress gene expression depending on genomic position, gene length and expression pattern in Drosophila melanogaster. Nucleic Acids Research, 41(8), 4481–4494. https://doi.org/10.1093/nar/gkt158

MacArthur, S., Li, X.-Y., Li, J., Brown, J. B., Chu, H. C., Zeng, L., … Eisen, M. B. (2009). Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biology, 10(7), R80. https://doi.org/10.1186/gb-2009-10-7-r80

Marfella, C. G. A., & Imbalzano, A. N. (2007). The Chd family of chromatin remodelers. Mutation Research, 618(1-2), 30–40. https://doi.org/10.1016/j.mrfmmm.2006.07.012

Martienssen, R., & Moazed, D. (2015). RNAi and heterochromatin assembly. Cold Spring Harbor Perspectives in Biology, 7(8), a019323. https://doi.org/10.1101/cshperspect.a019323

Mendel, J. G. (1865). Versuche über Plflanzenhybriden. Verhandlungen Des Naturforschenden Vereines in Brünn, IV, 3–47.

Meselson, M., & Stahl, F. W. (1958). The replication of DNA in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 44(7), 671–682. https://doi.org/10.1073/pnas.44.7.671

Mira-Bontenbal, H., & Gribnau, J. (2016). New Xist-Interacting Proteins in X-Chromosome Inactivation. Current Biology: CB, 26(8), R338–42. https://doi.org/10.1016/j.cub.2016.03.022

modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., … Kellis, M. (2010a). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science, 330(6012), 1787–1797. https://doi.org/10.1126/science.1198374

Morgan, T. H. (1915). The mechanism of Mendelian heredity, by T. H. Morgan [et al.]. https://doi.org/10.5962/bhl.title.6001

Mo, R., Rao, S. M., & Zhu, Y.-J. (2006). Identification of the MLL2 Complex as a Coactivator for Estrogen Receptor α. The Journal of Biological Chemistry, 281(23), 15714–15720. https://doi.org/10.1074/jbc.m513245200

Moses, R. E., & Richardson, C. C. (1970). A new DNA polymerase activity of Escherichia coli. I. Purification and properties of the activity present in E. coli polAl. Biochemical and Biophysical Research Communications, 41(6), 1557–1564. https://doi.org/10.1016/0006-291x(70)90565-6

Muller, H. J. (1930). Types of visible variations induced by X-rays inDrosophila. Journal of Genetics, 22(3), 299–334. https://doi.org/10.1007/bf02984195

Muller, H. J. (1941). INDUCED MUTATIONS IN DROSOPHILA. Cold Spring Harbor Symposia on Quantitative Biology, 9(0), 151–167. https://doi.org/10.1101/sqb.1941.009.01.019
Muller, H. J., & Painter, T. S. (1932). The differentiation of the sex chromosomes ofDrosophila into genetically active and inert regions. Zeitschrift Fur Induktive Abstammungs- Und Vererbungslehre, 62(1), 316–365. https://doi.org/10.1007/bf01948610

Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., … Simon, J. A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell, 111(2), 197–208. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12408864

Naono, S., & Gros, F. (1966). Control and selectivity of lambda DNA transcription in lysogenic bacteria. Cold Spring Harbor Symposia on Quantitative Biology, 31, 363–375. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4866387

Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proceedings of the National Academy of Sciences, 47(10), 1588–1602. https://doi.org/10.1073/pnas.47.10.1588

Nishimura, S., Jones, D. S., & Khorana, H. G. (1965). Studies on polynucleotides. 48. The in vitro synthesis of a co-polypeptide containing two amino acids in alternating sequence dependent upon a DNA-like polymer containing two nucleotides in alternating sequence. Journal of Molecular Biology, 13(1), 302–324. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/5323614

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., … 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
Olins, A. L., & Olins, D. E. (1974). Spheroid chromatin units (v bodies). Science, 183(4122), 330–332. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4128918

Oudet, P., Gross-Bellard, M., & Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell, 4(4), 281–300. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1122558

Pal-Bhadra, M., Leibovitch, B. A., Gandhi, S. G., Chikka, M. R., Rao, M., Bhadra, U., … Elgin, S. C. R. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science, 303(5658), 669–672. https://doi.org/10.1126/science.1092653

Paro, R., & Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proceedings of the National Academy of Sciences of the United States of America, 88(1), 263–267. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1898775

Passarge, E. (1979). Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. American Journal of Human Genetics, 31(2), 106–115. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/377956

Paweletz, N. (2001). Walther Flemming: pioneer of mitosis research. Nature Reviews. Molecular Cell Biology, 2(1), 72–75. https://doi.org/10.1038/35048077

Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A., & Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. Science, 339(6120), 698–699. https://doi.org/10.1126/science.1231382

Pérez-Lluch, S., Blanco, E., Tilgner, H., Curado, J., Ruiz-Romero, M., Corominas, M., & Guigó, R. (2015). Absence of canonical marks of active chromatin in developmentally regulated genes. Nature Genetics, 47(10), 1158–1167. https://doi.org/10.1038/ng.3381

Pombo, A., & Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. Nature Reviews. Molecular Cell Biology, 16(4), 245–257. https://doi.org/10.1038/nrm3965

Ptashne, M. (2004). A Genetic Switch: Phage Lambda Revisited. CSHL Press. Retrieved from https://books.google.com/books/about/A_Genetic_Switch.html?hl=&id=hhrnA-t-sMkC

Ptashne, M., & Gann, A. (2002). Genes & Signals. CSHL Press. Retrieved from https://market.android.com/details?id=book-IvMQRwA3bEAC

Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., … Sauer, R. T. (1980). How the λ repressor and cro work. Cell, 19(1), 1–11. https://doi.org/10.1016/0092-8674(80)90383-9

Razin, S. V., Gavrilov, A. A., Pichugin, A., Lipinski, M., Iarovaia, O. V., & Vassetzky, Y. S. (2011). Transcription factories in the context of the nuclear and genome organization. Nucleic Acids Research, 39(21), 9085–9092. https://doi.org/10.1093/nar/gkr683

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., … Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature, 406(6796), 593–599. https://doi.org/10.1038/35020506
Reinhart, B. J., & Bartel, D. P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science, 297(5588), 1831. https://doi.org/10.1126/science.1077183

Reinke, H., & Hörz, W. (2003). Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. Molecular Cell, 11(6), 1599–1607. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12820972

Rieder, D., Trajanoski, Z., & McNally, J. G. (2012). Transcription factories. Frontiers in Genetics, 3, 221. https://doi.org/10.3389/fgene.2012.00221

Robinson, P. J. J., & Rhodes, D. (2006). Structure of the "30 nm" chromatin fibre: a key role for the linker histone. Current Opinion in Structural Biology, 16(3), 336–343. https://doi.org/10.1016/j.sbi.2006.05.007

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., … Corces, V. G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Molecular Cell, 67(5), 837–852.e7. https://doi.org/10.1016/j.molcel.2017.07.022

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., … Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes & Development, 18(11), 1251–1262. https://doi.org/10.1101/gad.300704

Schultz, J. (1936). Variegation in Drosophila and the Inert Chromosome Regions. Proceedings of the National Academy of Sciences of the United States of America, 22(1), 27–33. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16588038

Sealy, L., & Chalkley, R. (1978). DNA associated with hyperacetylated histone is preferentially digested by DNase I. Nucleic Acids Research, 5(6), 1863–1876. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/673837

Sedkov, Y., Cho, E., Petruk, S., Cherbas, L., Smith, S. T., Jones, R. S., … Mazo, A. (2003). Methylation at lysine 4 of histone H3 in ecdysone-dependent development of Drosophila. Nature, 426(6962), 78–83. https://doi.org/10.1038/nature02080

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., … Cavalli, G. (2012). 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

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C. T., Bender, W., & Kingston, R. E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell, 98(1), 37–46. https://doi.org/10.1016/S0092-8674(00)80604-2

Shen, X., Mizuguchi, G., Hamiche, A., & Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. Nature, 406(6795), 541–544. https://doi.org/10.1038/35020123

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annual Review of Biochemistry, 81, 65–95. https://doi.org/10.1146/annurev-biochem-051710-134100

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R., & Peterson, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science, 311(5762), 844–847. https://doi.org/10.1126/science.1124000

Singer, G. A. C., Lloyd, A. T., Huminiecki, L. B., & Wolfe, K. H. (2005). Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. Molecular Biology and Evolution, 22(3), 767–775. https://doi.org/10.1093/molbev/msi062

Slattery, M., Ma, L., Spokony, R. F., Arthur, R. K., Kheradpour, P., Kundaje, A., … White, K. P. (2014). Diverse patterns of genomic targeting by transcriptional regulators in Drosophila melanogaster. Genome Research, 24(7), 1224–1235. https://doi.org/10.1101/gr.168807.113

Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M., & Zaret, K. S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell, 161(3), 555–568. https://doi.org/10.1016/j.cell.2015.03.017

Speyer, J. F., Lengyel, P., Basilio, C., & Ochoa, S. (1962). Synthetic polynucleotides and the amino acid code, II. Proceedings of the National Academy of Sciences, 48(1), 63–68. https://doi.org/10.1073/pnas.48.1.63

Speyer, J. F., Lengyel, P., Basilio, C., Wahba, A. J., Gardner, R. S., & Ochoa, S. (1963). Synthetic Polynucleotides and the Amino Acid Code. Cold Spring Harbor Symposia on Quantitative Biology, 28(0), 559–567. https://doi.org/10.1101/sqb.1963.028.01.075

Spofford, J. B. (1967). Single-locus modification of position-effect variegation in Drosophila melanogaster. I. White variegation. Genetics, 57(4), 751–766. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/6082620

Srinivasan, S., Armstrong, J. A., Deuring, R., Dahlsveen, I. K., McNeill, H., & Tamkun, J. W. (2005). The Drosophila trithorax group protein Kismet facilitates an early step in transcriptional elongation by RNA Polymerase II. Development , 132(7), 1623–1635. https://doi.org/10.1242/dev.01713

Stedman, E., & Stedman, E. (1950). Cell Specificity of Histones. Nature, 166(4227), 780–781. https://doi.org/10.1038/166780a0

Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. The Journal of Experimental Zoology, 14(1), 43–59. https://doi.org/10.1002/jez.1400140104

Sugiyama, T., Cam, H. P., Sugiyama, R., Noma, K.-I., Zofall, M., Kobayashi, R., & Grewal, S. I. S. (2007). SHREC, an effector complex for heterochromatic transcriptional silencing. Cell, 128(3), 491–504. https://doi.org/10.1016/j.cell.2006.12.035

Sutton, W. S. (1903). THE CHROMOSOMES IN HEREDITY. The Biological Bulletin, 4(5), 231–250. https://doi.org/10.2307/1535741

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell, 126(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., & Kennison, J. A. (1992). brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2SWI2. Cell, 68(3), 561–572. https://doi.org/10.1016/0092-8674(92)90191-e

Taunton, J., Hassig, C. A., & Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science, 272(5260), 408–411. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8602529

Thévenin, A., Ein-Dor, L., Ozery-Flato, M., & Shamir, R. (2014). Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. Nucleic Acids Research, 42(15), 9854–9861. https://doi.org/10.1093/nar/gku667

Tschermak, E. (1900). Ueber künstliche Kreuzung bei Pisum sativum.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., & Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. The EMBO Journal, 13(16), 3822–3831. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7915232

Tsukiyama, T., & Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell, 83(6), 1011–1020. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8521501

Turner, B. M. (1991). Histone acetylation and control of gene expression. Journal of Cell Science, 99 (Pt 1), 13–20. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1757496

Turner, B. M., Birley, A. J., & Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. Cell, 69(2), 375–384. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1568251

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., … Razin, S. V. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Research, 26(1), 70–84. https://doi.org/10.1101/gr.196006.115

van Steensel, B., & Belmont, A. S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell, 169(5), 780–791. https://doi.org/10.1016/j.cell.2017.04.022

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I. S., & Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science, 303(5658), 672–676. https://doi.org/10.1126/science.1093686

Vettese-Dadey. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. The EMBO Journal, 15(10), 2508–2518. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC450183/

Vinogradov, A. E. (2004). Compactness of human housekeeping genes: selection for economy or genomic design? Trends in Genetics: TIG, 20(5), 248–253. https://doi.org/10.1016/j.tig.2004.03.006

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. S., & Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science, 297(5588), 1833–1837. https://doi.org/10.1126/science.1074973

Volpe, T., Schramke, V., Hamilton, G. L., White, S. A., Teng, G., Martienssen, R. A., & Allshire, R. C. (2003). RNA interference is required for normal centromere function in fission yeast. Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology, 11(2), 137–146. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12733640

Wagner, E. J., & Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. Nature Reviews. Molecular Cell Biology, 13(2), 115–126. https://doi.org/10.1038/nrm3274

Watson, J. D., & Crick, F. H. C. (1974). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature, 248(5451), 765–765. https://doi.org/10.1038/248765a0

Weintraub, H., & Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science, 193(4256), 848–856. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/948749

Wood, A., Shukla, A., Schneider, J., Lee, J. S., Stanton, J. D., Dzuiba, T., … Shilatifard, A. (2007). Ctk complex-mediated regulation of histone methylation by COMPASS. Molecular and Cellular Biology, 27(2), 709–720. https://doi.org/10.1128/MCB.01627-06

Yasuhara, J. C., & Wakimoto, B. T. (2006). Oxymoron no more: the expanding world of heterochromatic genes. Trends in Genetics: TIG, 22(6), 330–338. https://doi.org/10.1016/j.tig.2006.04.008

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature, 518(7540), 556–559. https://doi.org/10.1038/nature13994

Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J., & Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science, 322(5902), 750–756. https://doi.org/10.1126/science.1163045