



UNIVERSITAT DE BARCELONA

The metabolic aspects of macroH2A histone variants

Iva Guberović

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

The metabolic aspects of macroH2A histone variants

Iva Guberović



PhD Thesis
2021

Universitat de Barcelona
Facultat de Biologia
Programa de Doctorat en Biomedicina

The metabolic aspects of macroH2A histone variants

Memòria presentada per

Iva Guberović

per optar al grau de doctora per la Universitat de Barcelona

Tesi realitzada a

l'Institut de Recerca contra la Leucèmia Josep Carreras

Director



Marcus

Buschbeck

Tutora



Montserrat

Corominas Guiu

Doctorand



Iva

Guberović

Barcelona, juny 2021

Table of Contents

| | | |
|------------|---|-----------|
| 1 | INTRODUCTION | 6 |
| 1.1 | CHROMATIN ORGANIZATION AND EPIGENETIC MODIFICATIONS | 7 |
| 1.2 | HISTONE VARIANTS | 9 |
| 1.3 | THE MACROH2A HISTONE VARIANTS | 11 |
| 1.3.1 | THE DISCOVERY OF MACROH2A FAMILY OF HISTONE VARIANTS | 11 |
| 1.3.2 | THREE ISOFORMS WITH TRIPARTITE STRUCTURE | 12 |
| 1.3.3 | THE CHROMATIN ENVIRONMENT OF MACROH2A HISTONE VARIANTS | 13 |
| 1.4 | THE MACRODOMAIN DISTINGUISHES THE ISOFORM-SPECIFIC FUNCTIONS OF MACROH2A1.1 AND MACROH2A1.2 | 16 |
| 1.4.1 | THE IMPLICATION OF MACROH2A1 HISTONE VARIANTS IN MOUSE PHYSIOLOGY | 18 |
| 1.4.2 | MACROH2A1.1 IS IMPLICATED IN THE NAD ⁺ -PARP1 AXIS | 19 |
| 1.4.3 | THE INSIGHTS INTO THE PHYSIOLOGICAL IMPLICATIONS OF MACROH2A1.2 ISOFORM | 23 |
| 2 | AIMS AND OBJECTIVES..... | 25 |
| 2.1 | MAIN AIMS AND OBJECTIVES..... | 26 |
| 3 | RESULTS | 27 |
| 3.1 | PREAMBLE | 28 |
| 3.2 | RESULTS I: THE SEARCH FOR THE EVOLUTIONARY ORIGIN OF MACROH2A1.1 IMPLICATION IN NAD⁺ METABOLISM | 29 |
| 3.2.1 | MACROH2A APPEARED IN PROTISTS AND REMAINED CONSERVED THROUGH EVOLUTION..... | 30 |
| 3.2.2 | THE STRUCTURAL CONSERVATION AND ADP-RIBOSE BINDING OF THE PROTIST MACROH2A MACRODOMAIN..... | 33 |
| 3.2.3 | TWO RESIDUES CRITICAL FOR THE INDUCED-FIT CLOSING OF THE BINDING POCKET ARE THE KEY EVOLUTIONARILY DIVERGENT FEATURE..... | 39 |
| 3.2.4 | THE SELECTION OF TWO RESIDUES FINE-TUNED ADP-RIBOSE AFFINITY THROUGH EVOLUTION | 43 |
| 3.2.5 | DYNAMIC REGULATION OF <i>CAPSASPORA</i> METABOLISM ACROSS THE LIFE STAGES INCLUDES THE MACROH2A – PARP1 – NAD ⁺ AXIS..... | 49 |
| 3.2.6 | <i>CAPSASPORA</i> MACROH2A IS A STRONGER PARP1 INHIBITOR THAN MOUSE MACROH2A1.1 AND HAS A MORE PRONOUNCED IMPACT ON MITOCHONDRIAL RESPIRATION..... | 55 |
| 3.2.7 | CONTRIBUTIONS..... | 60 |
| 3.3 | RESULTS II: INTERDISCIPLINARY APPROACH IN THE SEARCH FOR THE MACROH2A1.2 LIGAND | 62 |
| 3.3.1 | STRUCTURAL CHARACTERIZATION OF MACROH2A1.2 MACRODOMAIN..... | 63 |
| 3.3.2 | THE TARGETED APPROACH IN THE SEARCH FOR THE LIGAND OF MACROH2A1.2 | 70 |
| 3.3.3 | THE UNTARGETED APPROACH IN THE SEARCH FOR THE LIGAND OF MACROH2A1.2 | 78 |

| | | |
|------------|---|------------|
| 3.3.4 | THE ANALYSIS OF THE AVAILABLE TRANSCRIPTOMICS DATASET IN THE SEARCH FOR THE PHYSIOLOGICAL ROLE OF MACROH2A1.2 | 85 |
| 3.3.5 | THE SEARCH FOR THE PHYSIOLOGICALLY RELEVANT CELL LINE: MACROH2A1.2 IS EXPRESSED IN MOUSE BRAIN NEURONS | 87 |
| 3.3.6 | P19 CELL LINE AS A MODEL FOR STUDYING THE ROLE OF MACROH2A1.2..... | 90 |
| 3.3.7 | THE CHARACTERIZATION OF MACROH2A1.2 KNOCK-OUT MICE..... | 94 |
| 3.3.8 | CONTRIBUTIONS..... | 99 |
| 4 | <u>DISCUSSION.....</u> | 101 |
| 4.1 | DISCUSSION I | 102 |
| 4.1.1 | THE ORIGIN OF A MACRODOMAIN-CONTAINING HISTONE VARIANT..... | 102 |
| 4.1.2 | POSITIVE SELECTION OF FEW KEY RESIDUES DRIVES EVOLUTIONARY CHANGE | 103 |
| 4.1.3 | THE MACROH2A – PARP1 – NAD ⁺ AXIS SUSTAINS CATABOLISM IN NON-PROLIFERATIVE STAGES .. | 104 |
| 4.1.4 | THE NEED FOR THE COMPARTMENTAL REGULATION OF NAD METABOLISM | 105 |
| 4.1.5 | RELEVANCE OF MACROH2A FOR COMPARTMENTAL REGULATION IN THE CONTEXT OF EVOLUTION .. | 106 |
| 4.2 | DISCUSSION II | 108 |
| 4.2.1 | THE SPECIFIC FEATURES OF NON-ADP-RIBOSE BINDING MACRODOMAINS | 108 |
| 4.2.2 | IDENTIFICATION OF PHOSPHOLIPIDS AS PUTATIVE NATURAL LIGANDS OF MACROH2A1.2..... | 110 |
| 4.2.3 | THE IMPLICATION OF PHOSPHOLIPIDS IN THE REGULATION OF MEMBRANE STRUCTURE AND FUNCTION AND POTENTIALLY RELATED FUNCTIONS OF MACROH2A | 111 |
| 4.2.4 | THE ROLE OF PHOSPHOLIPIDS IN GENE REGULATION, CELL SPECIFICATION AND BEHAVIOR AND POSSIBLE LINKS THROUGH DIRECT BINDING TO MACROH2A1.2..... | 113 |
| 4.2.5 | A POTENTIAL LESSON FROM A DOMAIN CO-OCCURRING WITH THE ORPHAN MACRODOMAIN IN GDAP2 117 | |
| 5 | <u>CONCLUSIONS</u> | 119 |
| 5.1 | CONCLUSIONS | 120 |
| 7 | <u>MATERIALS AND METHODS</u> | 122 |
| 7.1 | COMPUTATIONAL ANALYSES | 123 |
| 7.1.1 | MOLECULAR DATA MINING | 123 |
| 7.1.2 | SEQUENCE ALIGNMENTS..... | 123 |
| 7.1.3 | PHYLOGENETIC AND EVOLUTIONARY ANALYSES | 124 |
| 7.1.4 | FOCUSED PHYLOGENETIC ANALYSIS OF SELECTED KEY RESIDUES..... | 125 |
| 7.1.5 | METABOLISM-FOCUSED ANALYSIS OF HIGH-CONTENT DATA FROM <i>CAPSASPORA</i> | 125 |
| 7.1.6 | MOLECULAR DOCKING AND SIMILARITY SEARCH..... | 126 |
| 7.2 | APPROACHES IN BIOPHYSICS AND BIOCHEMISTRY | 127 |
| 7.2.1 | PLASMIDS | 127 |
| 7.2.2 | PROTEIN PRODUCTION AND PURIFICATION | 128 |
| 7.2.3 | SATURATION TRANSFER DIFFERENCE NUCLEAR MAGNETIC RESONANCE (STD-NMR) | 129 |

| | | |
|------------|--|-------------------|
| 7.2.4 | ISOTHERMAL TITRATION CALORIMETRY..... | 129 |
| 7.2.5 | THERMAL SHIFT ASSAYS | 130 |
| 7.2.6 | CRYSTALLIZATION, DATA COLLECTION AND PROCESSING | 130 |
| 7.2.7 | PARP1 ACTIVITY ASSAY | 132 |
| 7.2.8 | ANTIBODIES..... | 132 |
| 7.3 | METABOLOMICS | 133 |
| 7.3.1 | ADP-RIBOSE PULL-DOWN ASSAY | 133 |
| 7.3.2 | UNTARGETED METABOLOMICS FOR LIGAND DETECTION..... | 134 |
| 7.4 | CELL CULTURE AND GENE EXPRESSION ANALYSES | 137 |
| 7.4.1 | STANDARD CULTURE OF HUMAN AND <i>CAPSASPORA</i> CELLS | 137 |
| 7.4.2 | INDUCTION OF NEURONAL DIFFERENTIATION IN P19 CELLS | 138 |
| 7.4.3 | GENE TRANSDUCTION AND ESTABLISHMENT OF STABLE CELL LINES | 139 |
| 7.4.4 | CELL PROLIFERATION ASSAYS | 139 |
| 7.4.5 | CELL FRACTIONATION, IMMUNOPRECIPITATION AND IMMUNOBLOTTING | 140 |
| 7.4.6 | RNA AND DNA ANALYSIS | 141 |
| 7.4.7 | ANALYSIS OF MITOCHONDRIAL OXIDATIVE PHOSPHORYLATION | 142 |
| 7.5 | EXPERIMENTAL WORK ON MOUSE TISSUES AND IN MICE..... | 142 |
| 7.5.1 | IMMUNOFLUORESCENCE OF MOUSE BRAIN SECTIONS | 142 |
| 7.5.2 | BREEDING AND PHENOTYPING OF MACROH2A1.2 KNOCK-OUT MICE | 143 |
| 7.6 | STATISTICAL ANALYSIS AND FIGURE EDITING | 143 |
| 7.7 | ANTIBODIES TABLE | 144 |
| 7.8 | PRIMERS TABLE | 145 |
| 8 | <u>ANNEX</u> | <u>153</u> |
| 8.1 | LIGPLOT INTERACTION NETWORK | 154 |
| 8.2 | OTHER RESIDUES AT POSITIONS 225 AND 316 DURING THE EVOLUTION OF MACROH2A(1.1)-LIKE MACRODOMAINS | 155 |
| 8.3 | COMPOUNDS TABLE..... | 156 |
| 8.4 | DETAILED GENE ONTOLOGY TABLE | 157 |
| 8.5 | THE MEASUREMENTS OF BODY WEIGHT AND COMPOSITION | 159 |
| 8.6 | INDIRECT CALORIMETRY RESULTS..... | 160 |
| 9 | <u>BIBLIOGRAPHY</u> | <u>161</u> |

SUMMARY

The histone variant macroH2A is the only structural chromatin component containing a macrodomain. In vertebrates, two genes and one event of alternative splicing give rise to three macroH2A proteins that differ in their macrodomains. As histone variants, macroH2A proteins contribute to the protein content of chromatin (Buschbeck & Hake, 2017). On the other hand, the capacity to bind ADP-ribose via its macrodomain is limited to the splice variant macroH2A1.1 (Kustatscher et al., 2005). As a consequence, macroH2A1.1, but not macroH2A1.2 or macroH2A2, binds auto-ADP-ribosylated PARP1 (Timinszky et al., 2009). Since the alternative splicing of the exon 5 affects the binding pocket of macroH2A1.2, as a consequence it cannot bind ADP-ribose (Kustatscher et al., 2005) and it remains an orphan protein.

In the first study presented here, we investigated the evolution of the macrodomain-containing histone variant macroH2A1.1, an integral chromatin component that limits nuclear NAD⁺ consumption by inhibiting PARP1. We found that macroH2A originated in pre-metazoan protists. The crystal structure of the macroH2A macrodomain from the protist *Capsaspora* allowed us to identify highly conserved principles of ligand binding and pinpoint key residue substitutions, selected for during the evolution of the vertebrate stem lineage. Metabolic characterization of the *Capsaspora* life cycle indicated that the metabolic function of macroH2A was associated with non-proliferative stages. Taken together, we provide insight into the evolution of a chromatin element involved in compartmental NAD regulation, relevant for understanding of its metabolism and potential therapeutic applications.

In the second study, we described the structurally relevant elements for ligand binding by the orphan macroH2A isoform, macroH2A1.2. Furthermore, using targeted and untargeted approaches on the verge of *in silico*, *in vitro* and *in cellulo* approaches, we detected phospholipids as the first putative physiological ligands of macroH2A1.2. We further observed a behavioral phenotype in macroH2A1.2 knock-out mice and report for the first time the upregulation of macroH2A1.2 expression in the differentiated cells, more specifically in differentiated neurons. We postulate that macroH2A1.2 might have a binding-pocket related role in the regulation of behavior, similarly to what was observed for PPAR α in hypothalamus, whereby it regulates animal behavior depending on the binding of its phospholipid ligands (Chakravarthy et al., 2007; Roy et al., 2016).

1 Introduction

1.1 Chromatin organization and epigenetic modifications

The DNA in the eukaryotic nucleus is arranged in chromatin, a dynamic nucleoprotein complex. The nucleosome is the basic subunit of chromatin, comprised of a histone octamer containing two copies of canonical histones, namely H2A, H2B, H3 and H4. The histone octamer is wrapped by 145-147 base pairs of DNA (Kornberg, 1977; Luger et al., 1997; McGhee & Felsenfeld, 1980; Zhou et al., 2019), with histone H1 further protecting the linker DNA between the neighboring nucleosomes (Khochbin, 2001). This way of chromatin organization allows for increased DNA compaction and contributes to the stabilization of the chromatin architecture (Hergeth & Schneider, 2015). While the field of genetics studies the DNA and the processes that directly affecting DNA, the epigenetics is concerned with the changes which do not occur on the level of DNA (Elia et al., 2012).

The term “epigenetics” was coined by Conrad Waddington to define “the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being” (Waddington, 2012). As an embryologist, he was mainly interested in the process of organismal development whereby one cell gives rise to the plethora of different cell types with different functions, despite their identical genotype. He further proposed the concept of an “epigenetic landscape” to explain the dynamic decision-making process during cell differentiation in which a cell can adopt different trajectories, depending on the stimuli it receives thus giving rise to different cell types (**Figure 1**) (Goldberg et al., 2007).

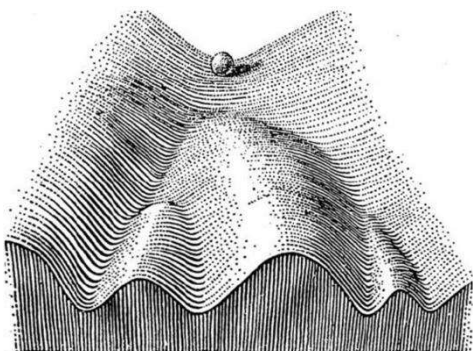


Figure 1. Epigenetic landscape proposed by C. Waddington.

Figure adapted from Goldberg et al., 2007.

Over time, several different definitions of epigenetics arose resulting in many heated discussions in the field (Deans & Maggert, 2015). The definition which we converge upon here is the widely accepted definition of epigenetics as the study of covalent and non-covalent modifications of chromatin which results in the change of gene expression and chromatin structure, and thus affects the epigenetic landscape without affecting the DNA sequence, rather its readout (Goldberg et al., 2007; Murr, 2010). The chromatin epigenetic modifications depend on the intracellular and environmental conditions and result in a differential epigenome (Weinhold, 2006). These modifications can occur on several levels, and include DNA modifications, histone tail post-translational modifications or the incorporation of histone variants (**Figure 2**). Thanks to the efforts in crystallography, cryo-EM, NMR and molecular dynamics it is now well established that the nucleosomes are dynamic platforms which contribute to the structural and functional dynamics of the chromatin regulation (Zhou et al., 2019).

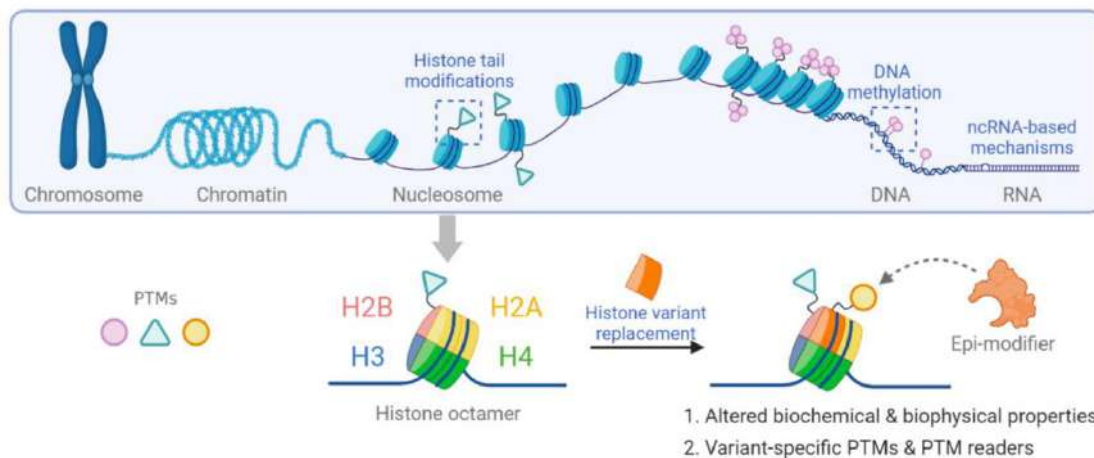


Figure 2. Epigenetic alterations affecting epigenetic landscape.

Different levels of chromatin modifications, representing DNA methylation, histone tail post-translational modifications, exchange of canonical histones for histone variants and ncRNA-based mechanisms all of which result in dynamic chromatin reorganization. Figure taken from Hsu et al., 2021.

1.2 Histone variants

The increasing number of reports show the histone variants play an important role in the regulation of differentiation, development, somatic cell reprogramming and cancer development (Buschbeck & Hake, 2017). The genes encoding for canonical histones are organized in clusters which are expressed during the S-phase of the cell cycle. As such, they are loaded on the DNA during replication and are referred to as the replication-coupled histones. Unlike them, the histone variants are usually present in one or two copies and their expression is regulated independently of the cell cycle, in a context dependent manner. As a result, the histones endow chromatin with unique properties and enable for coordinated and rapid response to the various stimuli (Buschbeck & Hake, 2017; Hurtado-Bagès et al., 2018; Martire & Banaszynski, 2020).

Histone variants of histone H2A are the biggest and most diverse group of histone variants, followed by the group of H3 histone variants. Interestingly, H2B has only testis-specific variants (Buschbeck & Hake, 2017), while the first variant of histone H4 was found only recently (Long et al., 2019) (**Figure 3**). The roles of histone variants in chromatin organization differ significantly. For example, H2A.Z takes part in chromatin relaxation and transcription initiation, while H2A.X incorporates into chromatin upon DNA damage and takes part in the recruitment of the repair machinery (Buschbeck & Hake, 2017; Martire & Banaszynski, 2020). Among all histone variants, macroH2A histone variants differ the most from their canonical counterpart due to the presence of the C-terminal macrodomain (Buschbeck & di Croce, 2010).

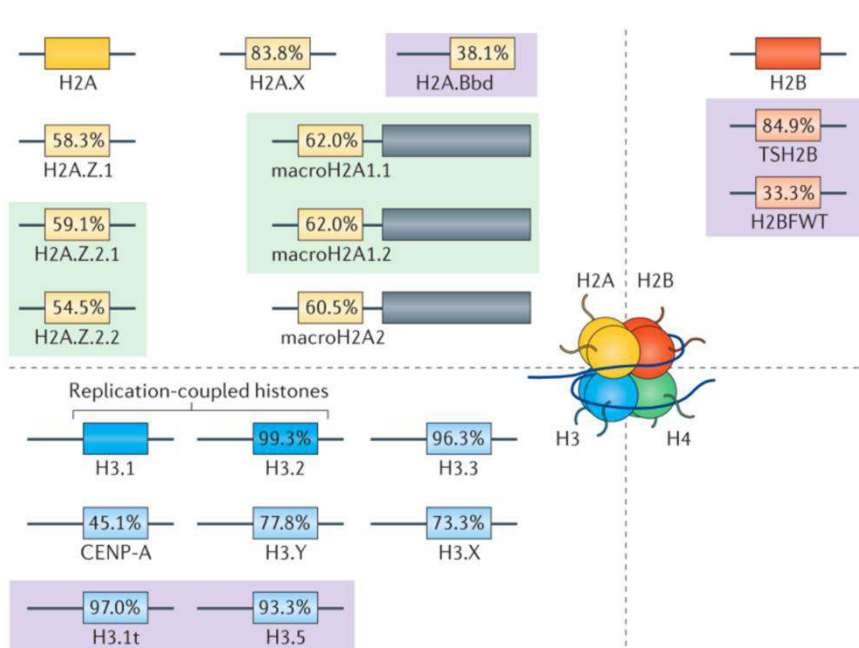


Figure 3. A schematic representing the panel of known histone variants.

A schematic representation of the known variants of canonical histones, grouped and color-coded according to the canonical histone of the family (H2A variants in yellow, H2B in red, H3 in blue). Testis-specific variants are denoted by the purple square. A recently discovered H4 histone variant (Long et al., 2019) is missing from the figure. Figure is taken from Buschbeck & Hake, 2017.

1.3 The macroH2A histone variants

1.3.1 The discovery of macroH2A family of histone variants

MacroH2A histone variants were discovered in the rat liver nucleosomes, as non-core histone proteins of 42 kDa which remain present in the mononucleosome after the sedimentation with 0.5M NaCl. They got their name as the “macro” variants of H2A due to the presence of a big globular domain at the C-terminus (Pehrson & Fried, 1992). For a very long time, macroH2A histone variants were thought to be vertebrate-specific histone variants, but recently the macroH2A encoding gene was also found in non-vertebrates and a member of holozoans (Rivera-Casas et al., 2016).

The levels of macroH2A proteins are the lowest in the early embryonic development (Nashun et al., 2010) and their expression increases with the differentiation of embryonic and skin stem cells (Creppe, Janich, et al., 2012). The functional studies in cells depleted of macroH2A proteins indicate they have tumor suppressive functions (Kapoor et al., 2010), inhibit somatic cell reprogramming (Pasque et al., 2011) and promote cell differentiation and development (Buschbeck et al., 2009; Creppe, Janich, et al., 2012). Taken together, these studies suggest the role of macroH2A histone variants in stabilizing the differentiated cell state (Buschbeck & Hake, 2017).

1.3.2 Three isoforms with tripartite structure

In vertebrates, two genes encode for three macroH2A proteins. The mRNA product of *MACROH2A1* gene undergoes mutually exclusive alternative splicing of exon 5, thus giving rise to either macroH2A1.1 or macroH2A1.2 isoform. MacroH2A2 isoform is a product of *MACROH2A2* gene (**Figure 4A**) (Chadwick & Willard, 2001; Costanzi & Pehrson, 2001; Pehrson & Fried, 1992). All three macroH2A isoforms have a particular tripartite structure with an H2A histone-fold at the N-terminus which enables macroH2A incorporation into chromatin. Histone fold is followed by the linker domain which is an unstructured domain enriched in positively charged amino acids lysine and arginine. The linker domain places the C-terminal globular macrodomain outside of the nucleosome. Although the macroH2A-containing nucleosomes have high similarity to the canonical H2A-nucleosomes, the positioning and properties of the linker and the macrodomain significantly affect the biophysical properties of the nucleosome (**Figure 4B**) (Chakravarthy et al., 2005).

MacroH2A histone variants are the only structural chromatin component containing a macrodomain, and thus, the macrodomain is considered to be the hallmark of macroH2A histone variants. Macrodomains are ancient globular protein modules that have emerged as key players in NAD⁺-dependent ADP-ribose signaling (Rack et al., 2016). They bind ADP-ribose as a free molecule, as an oligomer, or when covalently bound to proteins as a post-translational modification (Karras et al., 2005; Singh et al., 2017; Timinszky et al., 2009a), while in some cases they have hydrolysing activity (Jankevicius et al., 2013; Rosenthal et al., 2013). However, the capacity to bind ADP-ribose via its macrodomain is limited to the splice variant macroH2A1.1, which consequently interacts with poly-ADP-ribose polymerase 1 (PARP1), the main ADP-ribose producer in the nucleus (Schreiber et al., 2006). Since the alternative splicing of the exon 5 affects the binding pocket of macroH2A1.2, as a consequence it cannot bind ADP-ribose (Kustatscher et al., 2005). Similarly, the macrodomain of macroH2A2 has a significantly different binding pocket which cannot accommodate ADP-ribose neither and its ligand is currently unknown (Kozłowski

et al., 2018). As such, macroH2A1.2 and macroH2A2 macrodomains remain orphan proteins (**Figure 4C**).

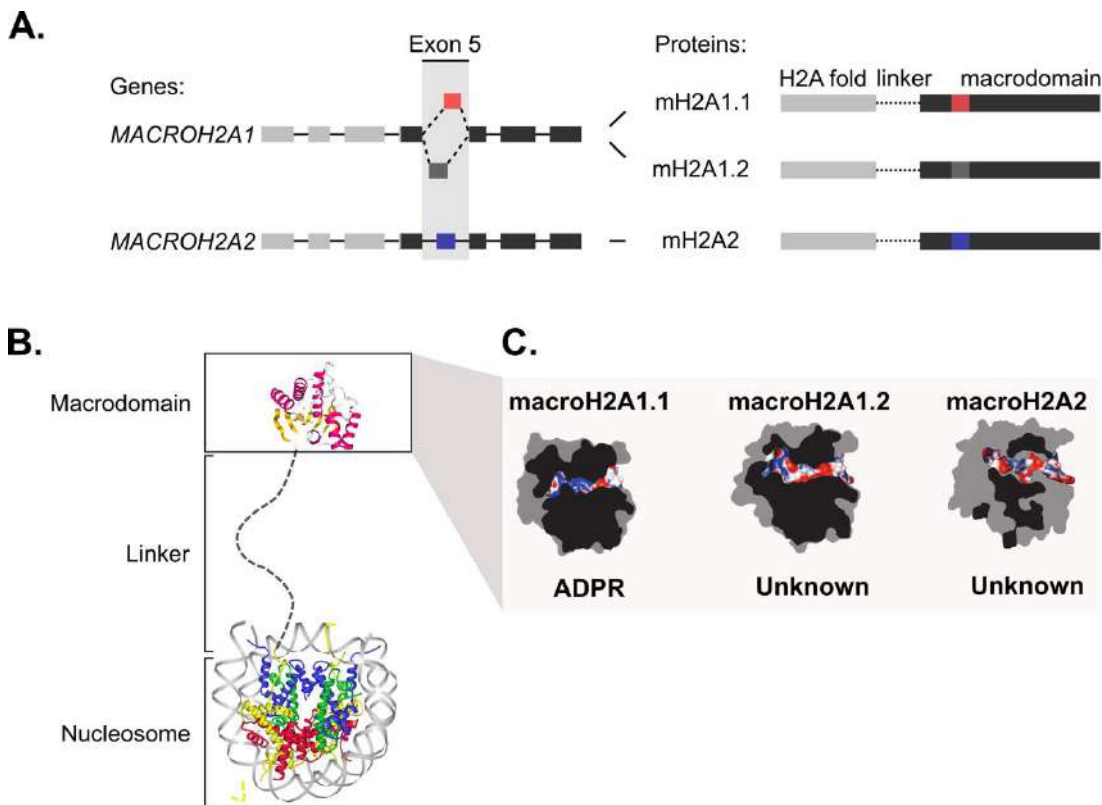


Figure 4. Two genes and alternative splicing event give rise to three macroH2A isoforms. (A) Two genes, *MACROH2A1* and *MACROH2A2*, encode for macroH2A histone variants. The exon 5 of macroH2A1 mRNA undergoes an event of exclusive alternative splicing resulting in macroH2A1.1 and macroH2A1.2 isoforms which differ in the ~30 residues in the macrodomain. (B) MacroH2A incorporation into the nucleosome significantly affects the biophysics of the nucleosome. (C) MacroH2A histone variants are the only structural components of chromatin which contain macrodomains. The macrodomains of the three isoforms differ significantly in the structure and the ability to bind ligands. Only macroH2A1.1 has a known ligand, ADP-ribose, while the other two isoforms remain orphans and their ligands are unknown.

1.3.3 The chromatin environment of macroH2A histone variants

The early research on macroH2A histone variants showed it is significantly enriched on the inactive X chromosome in female cells (Costanzi & Pehrson, 1998). Additionally, macroH2A was shown to be enriched in heterochromatin marked by H3K9me3 and H3K27me3 post-translational modifications (Buschbeck et al., 2009; Gamble et al., 2010; Gaspar-Maia et al., 2013). As such, it takes part in the organization of nuclear architecture and cells depleted of macroH2A suffer a global loss of heterochromatin and nuclear organization

(**Figure 5A**) (Douet et al., 2017; Fu et al., 2015). Furthermore, the linker region of the macroH2A has been shown to promote the chromatin compaction in a manner similar to histone H1 *in vitro* (Chakravarthy et al., 2012; Muthurajan et al., 2011). Recently, our group reinforced these results by confirming the implication of the linker region of all three macroH2A isoforms in the regulation of chromatin plasticity *in cellulo* (Kozlowski et al., 2018a). Hence, the linker region seems to mediate chromatin organization and plasticity, making it a general role of all macroH2A histone variants.

Due to its enrichment in the transcriptionally silent regions of heterochromatin, macroH2A histone variants were for a long time regarded as transcriptional repressors (Angelov et al., 2003; Chang et al., 2008; Doyen et al., 2006; Mermoud et al., 1999). Additionally, they are implicated in reducing the transcriptional variability and ensuring for the robust expression of target genes (Lavigne et al., 2015).

Although it was historically thought that all macroH2A isoforms incorporate at the same chromatin sites, accumulating evidence indicates much more complex and context-dependent roles of macroH2A histone variants. On the one hand, a highly similar ChIP-seq profile was observed between macroH2A1.2 and macroH2A2 in the HepG2 hepatoblastoma cells (**Figure 5B**) (Douet et al., 2017) which do not express macroH2A1.1 (H. Chen et al., 2014). As a consequence, it was assumed that the different macroH2A isoforms contribute to a different chromatin environment via the special properties arising from their macrodomains. For example, macroH2A1.1 was shown to be implicated in the regulation of gene expression in a binding pocket-dependent manner and relative to the ADP-ribosylation status in the nucleus (H. Chen et al., 2015; Creppe, Posavec, et al., 2012; Ouararhni et al., 2006).

However, several recent studies underline the differential chromatin incorporation of different variants (**Figure 5C**) (H. Chen et al., 2014; J. M. Kim et al., 2018; Jinman Kim et al., 2018; Ruiz & Gamble, 2018) coupled to a differential regulation of gene expression (**Figure 5D**) (Dardenne et al., 2012; Hurtado-Bagès, Posavec Marjanovic, et al., 2020; J. M. Kim et al., 2018; Jinman Kim et al., 2018; Simonet et al., 2020).

The data on the chaperons of macroH2A is scarce, and only very recently several studies indicated the possible chaperons of macroH2A histone variants. Two studies from the same group showed the involvement of LSH in the deposition of macroH2A to chromatin (Ni et al., 2020; Xu et al., 2021), whereas FACT and ATRX are reported to be implicated in their eviction (Kim et al., 2019; Sun et al., 2018). However, there are currently no data which could indicate putative isoform-specific chaperons.

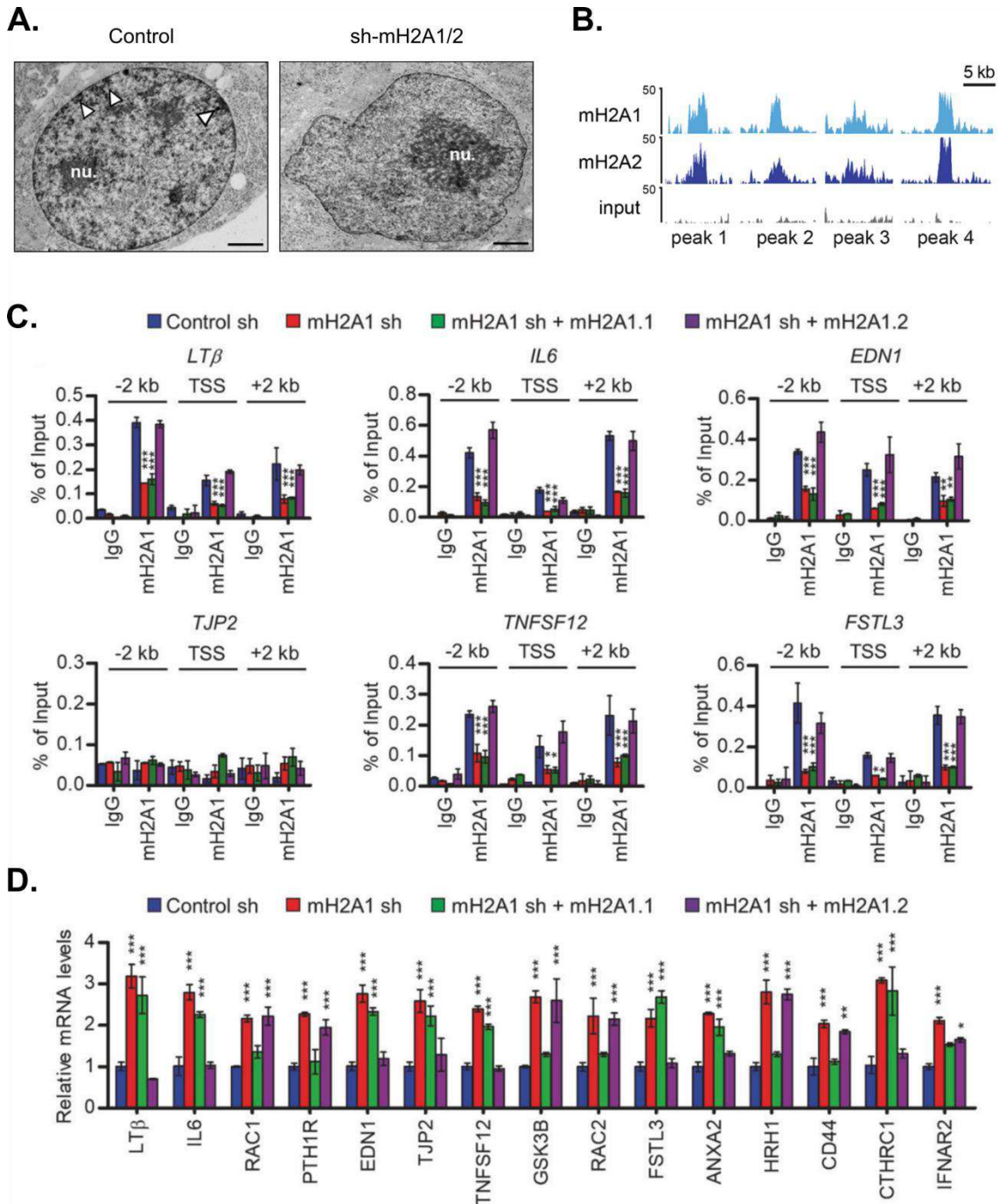


Figure 5. The incorporation macroH2A in chromatin and its implication in the organization of nuclear architecture. (A) The transmission electron microscopy images of HepG2 control and cells depleted of macroH2A1 and macroH2A2 histone variants (sh-mH2A1/2). White arrowheads indicate dark contrast staining corresponding to dense heterochromatin and 'nu.' marks the nucleolus. Scale bars: 1 μ m. **(B)** The ChIP-seq experiment from HepG2 cells representing four examples of macroH2A peak enrichments. **(C)** ChIP-qPCR experiment in DU145 cells mock-depleted (Control sh) or macroH2A1-depleted (mH2A1 sh) DU145 cells using mH2A1 antibody. To rescue the effects of macroH2A1 knockdown, macroH2A1-depleted cells were transfected with shRNA-resistant macroH2A1.1 or macroH2A1.2. **(D)** Expression levels macroH2A1-depleted DU145 cells complemented with macroH2A1.1 or macroH2A1.2. Figures (A) and (B) are from Douet et al., 2017; figures (C) and (D) are from Kim et al., 2018.

1.4 The macrodomain distinguishes the isoform-specific functions of macroH2A1.1 and macroH2A1.2

The macroH2A1.1 and macroH2A1.2 isoforms differ only in the binding pocket of their macrodomains, whereas the histone and the linker folds remain the same (Kustatscher et al., 2005). However, their different expression profile, chromatin incorporation and physiological implications indicate the macrodomain-, or the binding pocket-, specific roles of the two isoforms.

The cells characterized by high proliferation rate, such as proliferating stem cells or cancer cells, are characterized by the high expression of macroH2A1.2 isoform (Chen et al., 2014; Creppe, Posavec, et al., 2012). Interestingly, the splicing switch was observed during the cell differentiation resulting in a predominant expression of macroH2A1.1 in differentiated cells (Posavec-Marjanović et al., 2017; Judith C. Sporn & Jung, 2012).

The tissue expression analysis of macroH2A1.1, macroH2A1.2 and macroH2A2 isoforms in a panel of different mouse tissues shows a complex expression profile of different isoforms. MacroH2A2 is detected at significant levels in heart, intestine, colon, kidney, prostate, lung and brain. When we focus on the alternatively spliced isoforms, we observe they have differential expression across mouse tissues. While macroH2A1.1 is detected in the majority of the tissues evaluated, macroH2A1.2 has a more particular expression profile and is detected at significant levels in colon, pancreas, liver, kidney, lung and brain (**Figure 6**) (Kozłowski et al., 2018).

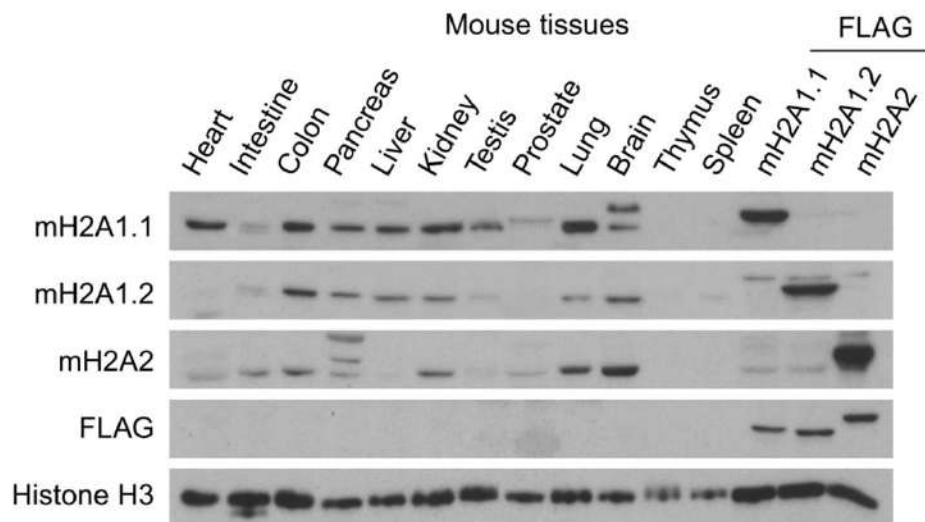


Figure 6. The tissue expression analysis of macroH2A histone variants. The expression of macroH2A1.1, macroH2A1.2 and macroH2A2 histone variants in adult mouse tissues. Samples of FLAG-tagged isoforms in HepG2 cells are included to aim comparison. Figure taken from Kozłowski et al., 2018.

The splicing affects the stretch of approximately 30 residues within the macrodomain binding pocket, rendering it impossible for macroH2A1.2 to bind ADP-ribose (Kustatscher et al., 2005). However, it is still not entirely clear how the splicing is regulated. The splicing factors MBNL1 and QKI seem to be implicated in the preferential expression of macroH2A1.1 over macroH2A1.2 (Novikov et al., 2011; Judith C. Sporn & Jung, 2012), whereas RNA helicases Ddx 5 and 17 result in preferential expression of macroH2A1.2 (Dardenne et al., 2012). Interestingly, increased QKI expression in gastric and prostate cancers are related to limited proliferation and increased differentiation, at least in part via increasing the expression of macroH2A1.1 (Li et al., 2016; Vieira-Silva et al., 2019).

Unfortunately, majority of the studies until now did not distinguish between macroH2A1.1 and macroH2A1.2 isoforms, rather they studied the roles of the total macroH2A1. However, the few studies distinguishing between the two isoforms reported a conclusive tumor suppressive role of macroH2A1.1 with its expression being inversely correlated with cell proliferation (Li et al., 2016; Novikov et al., 2011; J. C. Sporn et al., 2009; Judith C. Sporn & Jung, 2012). On the other hand, the role of macroH2A1.2 seems to be largely context dependent

and is reported to be correlated with both cancer promotion and suppression (Dardenne et al., 2012; Kapoor et al., 2010; discussed in Cantariño et al., 2013). Nonetheless, there is no conclusive data showing that macroH2A1.2 directly promotes cancer phenotype, and that this is not the consequence of the loss of macroH2A1.1 expression. Interestingly, macroH2A1.1 specifically inhibits PARP1 activity via ADP-ribose binding pocket (Kozłowski et al., 2018a; Posavec-Marjanović et al., 2017; Ruiz et al., 2020), whereas the PARP1 inhibitors are successfully used in cancer therapy (Slade, 2020). Thus, it remains to be seen if the loss of macroH2A1.1 or the increase of macroH2A1.2 expression contribute to observed cancer promoting phenotype of macroH2A1.2.

1.4.1 The implication of macroH2A1 histone variants in mouse physiology

The physiological roles of macroH2A isoforms were studied in the mice with systemic loss of macroH2A proteins. These studies reported that the mice with the deletion of *MACROH2A1* gene are viable and fertile, and do not show an increased sensitivity to ionizing radiation (Boulard et al., 2010), whereas the mice with the deletion of both *MACROH2A1* and *MACROH2A2* genes exhibit growth defects reflected in the smaller size of the animals (Pehrson et al., 2014). Apart from the smaller size of the animals, majority of the other phenotypes were related to the impaired metabolic function.

In two studies, the *MACROH2A1*^{-/-} mice coming from a different genetic background showed a deregulation of the expression of genes involved in lipid metabolism in the liver (Boulard et al., 2010; Changolkar et al., 2007). Furthermore, the female mice presented with lipid steatosis with 50% penetrance of the phenotype (Boulard et al., 2010). The *in cellulo* data from liver cancer cells show that the overexpression of macroH2A1.1 results in the lower accumulation of lipids (Pazienza et al., 2014), whereas the overexpression of macroH2A1.2 in mice results in the reduction of adiposity, thus promoting leanness (Pazienza et al., 2016).

Interestingly, although *MACROH2A1*^{-/-} mice on chow diet show decreased glucose tolerance (Changolkar et al., 2010), a study of the *MACROH2A1*^{-/-} mice

on high fat diet showed they present with an increased energy expenditure, lower adiposity and increased leanness (Sheedfar et al., 2015). As such, the obtained data from the mice depleted of macroH2A proteins indicate a complex implication of macroH2A isoforms in the regulation of metabolism. The future studies focusing on isoform-specific macroH2A deletion in a tissue-specific manner might shed more light on the exact roles and mechanisms of these proteins in metabolism.

1.4.2 MacroH2A1.1 is implicated in the NAD⁺–PARP1 axis

1.4.2.1 NAD metabolism – the NAD⁺ salvage pathway

NAD metabolism plays an essential role in all domains of life (Covarrubias et al., 2021; Rajman et al., 2018). NAD functions as a redox cofactor and a signaling molecule. As a redox cofactor in catabolic reactions, it enables ATP production in mitochondria (Xiao et al., 2018). As a signaling molecule in its oxidized form, NAD⁺ donates ADP-ribose moieties to enzymes such as PARPs, sirtuins and CD38 family of hydrolases (Palazzo et al., 2019). After donating the ADP-ribose moiety, NAD can be replenished by de novo synthesis pathways primarily parting from tryptophan, or via one of the NAD⁺ salvage pathways (Canto et al., 2015). The major NAD⁺ regeneration pathway is an intracellular NAD salvage pathway of nicotinamide (NAM) which remains after cleavage of ADP-ribose (**Figure 7**). The rate-limiting activity of NAM phosphoribosyltransferase (NAMPT) converts NAM into NAM mononucleotide (NMN). This reaction primarily occurs in the cytosol, and the resulting NMN is used by one of the three compartment-specific NMN adenylyltransferase (NMNAT) enzymes that convert NMN into NAD⁺ in a highly localized manner in the nucleus, mitochondria, cytosol and Golgi (Berger et al., 2005). Reducing NAD⁺ biogenesis by genetic knock-out of NAMPT leads to dedifferentiation in Schwann cells (Sasaki et al., 2018) and degeneration in muscle (Frederick et al., 2016). The alterations in NAD biosynthesis in humans are a known cause of various diseases, and thus associate the depletion of NAD⁺ pools with the deterioration of normal physiological processes during aging (Covarrubias et al., 2021; Rajman et al., 2018).

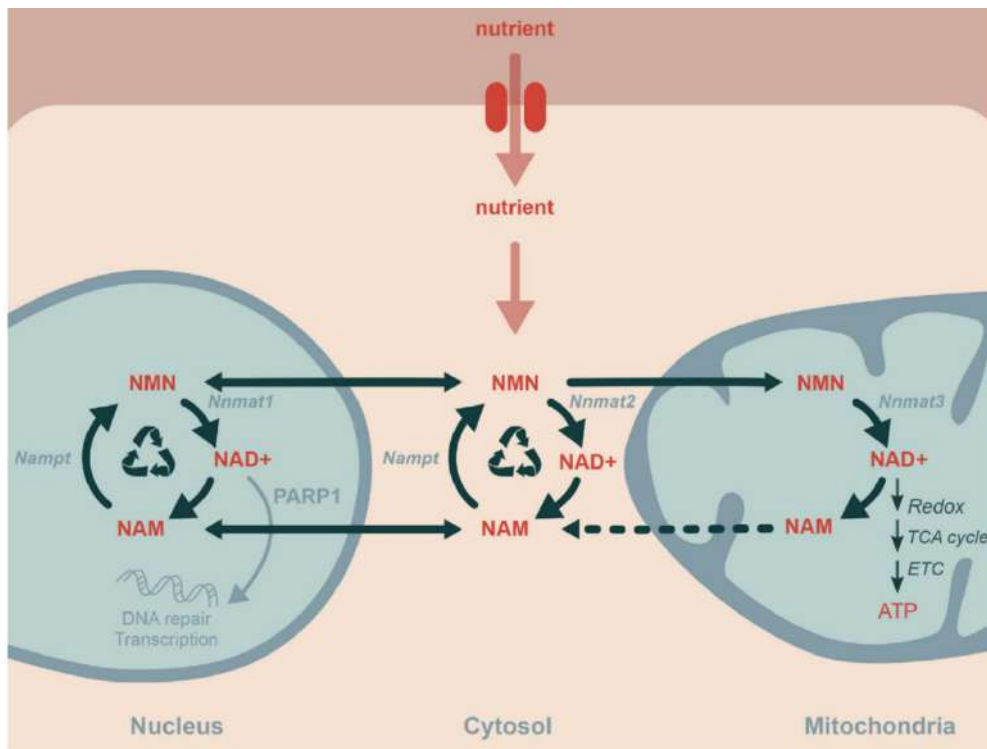


Figure 7. The compartmentalization of NAD metabolism. The major NAD⁺ regeneration pathway is an intracellular NAD salvage pathway of NAM which remains after cleavage of ADP-ribose. The recently discovered NAD⁺ import by SLC25A51 is not represented in this figure. Figure taken from (Hurtado-Bagès, Knobloch, et al., 2020).

1.4.2.2 The compartmentalization of NAD metabolism

NAD metabolism is highly compartmentalized, as corroborated by the compartmental distribution of both the NAD biosynthetic enzymes and the NAD⁺ dependent effector proteins (**Figure 7**) (Cambronne & Kraus, 2020; Strømmland et al., 2019). However, a major challenge for the field is to understand how is NAD metabolism regulated on the compartmental level, and to dissect the relevance of compartmental regulation for health and disease (Cambronne & Kraus, 2020). In particular, the balance between NAD⁺ consumption in the nucleus and its availability for cytosolic and mitochondrial redox reactions is essential for energy homeostasis (Canto et al., 2015; Strømmland et al., 2019). The earliest identified mechanisms of compartmental regulation are detected by monitoring the local changes in the NAD salvage pathway. Stress-induced transport of NAMPT into the nucleus increases the availability of NAD⁺ for nuclear enzymes (Grolla et al., 2020), while the upregulation of the cytosolic enzyme NMNAT2 has the opposite effect and reduces nuclear NAD⁺ availability (Ryu et al., 2018). Furthermore, the

mechanism for mitochondrial NAD⁺ import has just been discovered (Girardi et al., 2020; Kory et al., 2020; Luongo et al., 2020), providing another interesting control point for the regulation of NAD compartmentalization.

1.4.2.3 Nuclear ADP-ribose and PARP1

As a post-translational modification, ADP-ribosylation is implicated in a plethora of different cell processes including regulation of gene transcription, DNA damage repair, aging, cell proliferation, differentiation and death, or the regulation of metabolism (reviewed in Gibson & Kraus, 2012; Gupte et al., 2017; Hurtado-Bagès, Knobloch, et al., 2020). The main enzymes implicated in ADP-ribosylation of proteins are ADP-ribosyltransferases which include PARPs and sirtuins.

The PARP family of ADP-ribosyltransferases catalyzes the transfer of ADP-ribose units from NAD⁺ to the target proteins resulting in their mono- or poly-ADPR-ribosylation, i.e. MARYlation or PARYlation, respectively (Hurtado-Bagès, Knobloch, et al., 2020). The most abundant and best characterized member of PARP family is PARP1, the major NAD⁺ consuming enzyme in the nucleus. It is best known for its function as a sensor in the DNA damage response (Altmeyer & Hottiger, 2009). In response to DNA damage, PARP1 rearranges to reach its active conformation. When active, it first modifies itself in the process of auto-PARYlation, followed by the PARYlation of its target proteins. In this way PARP1 creates a hub of PAR signals and enables the recruitment of PAR binding DNA repair proteins (Hurtado-Bagès, Knobloch, et al., 2020).

The inhibition of PARP1 activity in the nucleus results in increased global NAD⁺ levels in both cultured cells and mice, indicating that PARP1 also consumes NAD⁺ under basal physiological conditions (Bai et al., 2011; Pirinen et al., 2014; Posavec-Marjanović et al., 2017). Conversely, nuclear NAD⁺ levels control PARP1 activity, whereas the decreased levels of nuclear NAD⁺ promote adipogenic differentiation by relieving the transcription factor c/EBP β from PARP1 mediated inhibition (Luo et al., 2017; Ryu et al., 2018).

1.4.2.4 MacroH2A1.1–PARP1 axis and the mitochondrial respiration

In the search for the physiological role of macroH2A1.1 with respect to its ADP-ribose binding pocket, the early reports looked into the possible interactions of macroH2A1.1 and PARP1, as the main source of ADP-ribose in the nucleus. Indeed, it was confirmed that macroH2A1.1 and PARP1 interact in the ADP-ribose-mediated manner, and consequently regulate the stress response in a transcription-dependent manner (Chen et al., 2014; Nusinow et al., 2007; Ouararhni et al., 2006; Timinszky et al., 2009; reviewed in Hurtado-Bagès et al., 2018).

In the more recent research efforts, a picture is emerging in which nuclear NAD⁺ consumption by PARP1 in differentiated cells needs to be kept low to maintain NAD-dependent functions in other compartments (Hurtado-Bagès, Knobloch, et al., 2020). During myogenic differentiation this is achieved by the transcriptional downregulation of PARP1 expression (Oláh et al., 2015) and the simultaneous upregulation of macroH2A1.1 as its endogenous nuclear inhibitor (Posavec-Marjanović et al., 2017). In particular, when more abundant than PARP1, macroH2A1.1 can blunt PARP1 activity and inhibit PARP1-dependent processes (Kozłowski et al., 2018; Ouararhni et al., 2006). In the differentiated muscle cells, macroH2A1.1 reduces nuclear NAD⁺ consumption by PARP1, and thus indirectly promotes mitochondrial respiration and ATP production (**Figure 8**) (Posavec-Marjanović et al., 2017). Interestingly, the macroH2A1.1-depleted myotubes show only a subtle change in the gene expression, with less than 50 genes differentially expressed by more than twofold (Posavec-Marjanović et al., 2017). Thus, the increased mitochondrial respiration in the presence of macroH2A1.1 in myotubes cannot be explained by the change in the gene expression. The extensive exploration of the implication of macroH2A1.1 in the NAD⁺ salvage pathway and the regulation of mitochondrial respiratory capacity showed that the role of macroH2A1.1 in mitochondrial respiration is dependent on its ability to bind ADP-ribose and inhibit PARP1 (**Figure 8**) (Posavec-Marjanović et al., 2017). This implicates macroH2A1.1 in the physiological function via both the metabolite sensing and the regulation of gene expression, and future research will need to discern whether these two things are coupled or de-coupled.

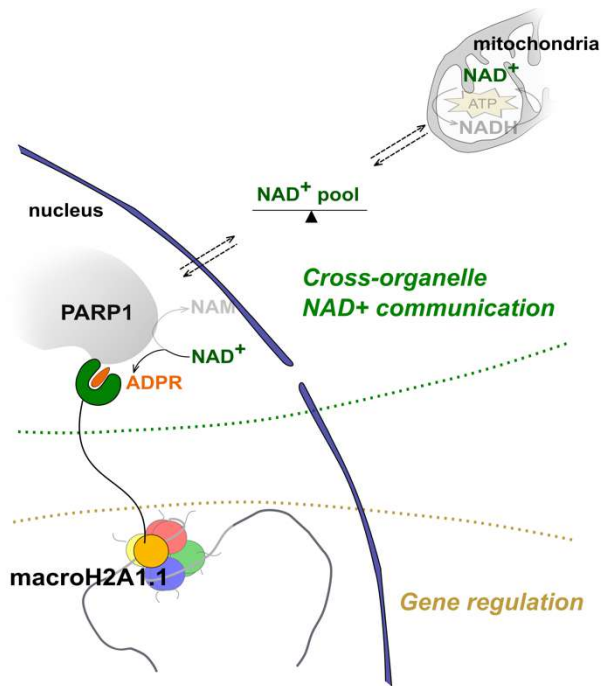


Figure 8. The implication of macroH2A1.1 in the regulation of mitochondrial respiration.

The macroH2A1.1 is incorporated into the nucleosome via the H2A-fold whereby it is implicated in chromatin organization and gene expression. In a gene-regulation de-coupled manner and in the binding pocket related manner, it binds ADP-ribosylated PARP1 and inhibits its NAD⁺ consumption in the nucleus. In this way, it stabilizes cell NAD⁺ pool and enables optimal mitochondrial regulation. This schematic representation is based on Posavec-Marjanović et al., 2017.

1.4.3 The insights into the physiological implications of macroH2A1.2 isoform

Currently available data on the physiological roles of macroH2A1.2 isoform is relatively scarce. MacroH2A1.2 differs from the macroH2A1.1 isoform only in the amino acids encoded by the exon 5 which affect the binding pocket (Karras et al., 2005), however it cannot be excluded that the surface of the macrodomain might be affected. There are several studies addressing the isoform-specific role of macroH2A1.2, however majority of them does not put the observed phenotypes in the context with macroH2A1.1, making it very challenging to deduce if the macrodomain or its binding pocket might be implicated in the function.

MacroH2A1.2 is implicated in the DNA damage repair independently of PARP1. More specifically, it has been implicated in the homologous repair of DNA double strand breaks whereby it was shown that macroH2A1.2 acts in concert with PRDM2 methyltransferase to modulate BRCA1-mediated DNA damage repair (Khurana et al., 2014). Furthermore, macroH2A1.2 seems to interact with EZH2 methyltransferase in a macrodomain-dependent manner which results in increased H3K27me3 whereby it affects the gene expression and consequently inhibits breast cancer-mediated osteoclastogenesis (Jinman Kim et al., 2018). It

is additionally been shown to be enriched in telomeres where it seems to have a protective role during replication stress whereby it prevents double strand DNA breaks (Jeongkyu Kim et al., 2019).

As mentioned, the splicing switch which occurs during the differentiation affects the proteins levels in the cells, with macroH2A1.2 being high in proliferating cells and macroH2A1.1 in the differentiated cells. In their study, Dell'Orso et al. found that the recruitment of Pbx1 to enhancers is required for muscle differentiation, and due to the low expression of macroH2A1.1 in myoblasts, its regulation was attributed to macroH2A1.2 (Dell'Orso et al., 2016). However, upon performing the isoform specific depletion of macroH2A1.2 in myoblasts, our group observed an increase in the fusion of myoblasts to myotubes resulting in a rather opposite phenotype (Hurtado-Bagès, Posavec Marjanovic, et al., 2020).

In a very recent study, Ma et al. reported autistic behavior in the *MACROH2A1*^{-/-} mice which they reported to be a macroH2A1.2-specific phenotype since the wild-type phenotype was rescued upon the overexpression of macroH2A1.2 by *in utero* electroporation (Ma et al., 2021).

2 Aims and Objectives

2.1 Main aims and objectives

Both macrodomains and histones are ancient protein-folds, whereby macrodomains are present in all domains of life including viruses (Rack et al., 2016) and histone homologs have recently been detected in both bacteria and archaea (Alva & Lupas, 2019). MacroH2A proteins likely evolved by the serendipitous fusion of genes encoding for a macrodomain and for a histone. This might have provided the basis for an ancient function related with NAD⁺ metabolism or ADP-ribose signaling. Furthermore, we speculate that younger forms of macroH2A which are unable to bind ADP-ribose, might have acquired independent metabolite-binding activities. The overall goal of my study was to gain a better understanding of the metabolic functions of macroH2A histone variants.

My specific objectives were:

1. To investigate the evolution of the macrodomain-containing histone variant macroH2A1.1. Knowing that it is an integral chromatin component that limits nuclear NAD⁺ consumption by inhibiting PARP1 in vertebrates, we wished to address the question if macroH2A1.1 had a metabolic role from the moment it appeared in the evolution, or rather if its implication in the metabolic regulation was acquired during evolution coinciding with the increasing metabolic complexity of higher organisms.
2. To identify, or at least to determine the key structural properties of the ligand that can be accommodated in the binding pocket of macroH2A1.2 macrodomain. Our technical aim was to develop an untargeted approach to detect yet unknown ligand of macroH2A1.2. Furthermore, we aimed to identify an adequate and relevant cell culture model to study the physiological role of macroH2A1.2 and its capacity to bind ligands.

3 Results

3.1 Preamble

In this section I will present the results obtained during my doctoral thesis. I organized the results in two sections, referring to the two projects I took part in during the last four and a half years in the Chromatin, Metabolism and Cell fate group led by Marcus Buschbeck.

The first section titled “Results I: The search for the evolutionary origin of macroH2A1.1 implication in NAD⁺ metabolism” refers to the project in which we studied the origin of macroH2A histone variants with the focus on the characterization of the role of macroH2A from unicellular organisms to vertebrates. The main results of this section are summarized in a first authored manuscript that is currently under peer-review in Nature Structural and Molecular Biology.

The second section is titled “Results II: Interdisciplinary approach in the search for the macroH2A1.2 ligand” and it refers to the project in which we sought the ligand of the macroH2A1.2 isoform and its related physiological role. The results from this study will be further expanded in the future efforts by the group and will be included in a manuscript with shared first authorship that we estimate to submit in the next two years.

Both studies were performed in an interdisciplinary manner and in collaboration with other groups. In order to be able to coherently discuss the results of the projects as a whole, I included the results obtained by collaborators and other members of our team. In an attempt to be fully transparent, at the end of each section I included a paragraph in which I detail my own contributions and those of others.

3.2 Results I: The search for the evolutionary origin of macroH2A1.1 implication in NAD⁺ metabolism

From the cell biology perspective, there were two major steps in evolution. The first was the fusion of the chromatin-based information system from archaea with the efficient energy-producing system from bacteria. It is postulated that this fusion provided the complexity which allowed for the diversification of eukaryotic life forms (Brunk & Martin, 2019; Lane & Martin, 2010). Indeed, the partitioning of metabolism and other cellular processes into different cell compartments, provides eukaryotes with an increased capacity to adapt to a changing environment (Gabaldón & Pittis, 2015; Zecchin et al., 2015). This increased level of plasticity allowed for the second key step, which was the emergence of cell differentiation and multicellular forms of life.

NAD metabolism is the major model of metabolic compartmentalization. The emergence of its compartmentalized regulation is a likely prerequisite for cellular plasticity, which might have favored the consolidation of multicellular lifestyles. However, we completely lack an understanding of how and when did the compartmental regulation of NAD metabolism evolve. The fusion of a macrodomain to a histone fold provided cells with a system known to limit nuclear NAD⁺ consumption in mouse and human cells.

Here, we addressed the question on when did this regulatory element emerge on an evolutionary scale and how was its function fine-tuned via natural selection.

3.2.1 MacroH2A appeared in protists and remained conserved through evolution

The capacity of chromatin to regulate nuclear NAD⁺ consumption in vertebrates is restricted to the histone variant macroH2A1.1 and depends fully on its ability to bind ADP-ribose and inhibit PARP1. The ADP-ribose binding pocket of macroH2A1.1 is largely defined by the amino acids encoded by exon 5 (**Figure 4A**). To understand when did the role of macroH2A in the compartmental regulation of NAD metabolism evolve, we traced back the origin of macroH2A with a goal to detect when did the fusion between the metabolite-binding macrodomain and the histone-fold occur. For this, we analyzed genomic and transcriptomic sequencing data representing a wide diversity of eukaryotes, and identified a *MACROH2A* gene in 330 of them. We focused our subsequent analysis on the sequences encoding the macrodomain.

The presence of macroH2A was previously reported in the unicellular filasterean, *Capsaspora* (Rivera-Casas et al., 2016). Together with animals and fungi, filastereans belong to the group of opisthokonts (Torruella et al., 2015). Here, we detected the presence of macroH2A sequences in protists that diverged earlier than opisthokonts, such as the haptist *Choanocystis* sp. and the breviate *Pygusua biforma*. This suggests an even older origin of the *MACROH2A* gene than previously reported (**Figure 9A, 9B**). Haptista is a phylum more closely related to plants than to animals, while breviate constitute a group closely related to opisthokonts (Burki et al., 2020). Interestingly, we did not find macroH2A in Fungi, nor in other unicellular opisthokonts.

The presence of macroH2A became more common with the emergence of animal multicellularity. Indeed, we found macroH2A-encoding genes in diverse species of most animal phyla and all vertebrates. On the other hand, and as previously reported (Rack et al., 2016; Rivera-Casas et al., 2016), we observed the absence of macroH2A in several non-vertebrate species, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, as well as in tunicates (**Figure 9A**), which is indicative of lineage-specific losses. The second macroH2A gene corresponding to human *MACROH2A2* appeared after the whole-genome duplication in the last common ancestor of vertebrates, followed by the

appearance of the alternative splicing variant of the ancestral *MACROH2A1* in a common ancestor of jawed vertebrates (**Figure 9A**).

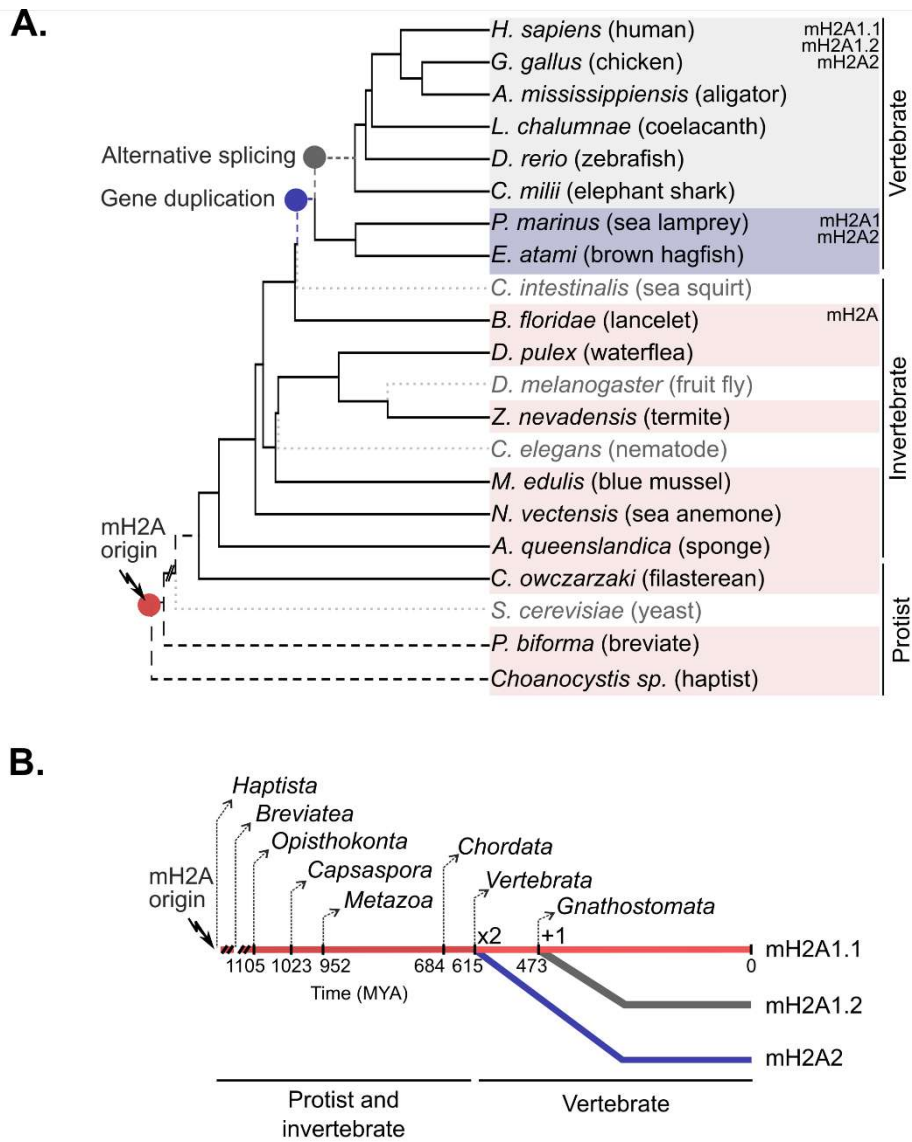


Figure 9. The phylogenetic analysis of the evolution of macroH2A histone variants.

(A) Schematic phylogeny based on the estimated divergence times of selected species. The events of macroH2A origin (arrow), gene duplication and splicing isoform emergence are represented in the tree. Dotted gray lines depict events of macroH2A loss. **(B)** Schematic overview of the evolutionary history of macroH2A. MYA, million years ago.

To assess the evolutionary rate of the three macroH2A macrodomains present in vertebrates, we calculated the number of amino acid and nucleotide substitutions over time. As shown in **Figure 10A**, all three vertebrate macrodomains and sequences corresponding to exon 5 show a significant level of conservation that placed them closer to the highly conserved, replication-coupled histone H2B than the fast-evolving histone variant H2A.Bbd.

The main difference between the three macroH2A isoforms in vertebrates, and the only difference between the two mutually-exclusive macroH2A1 splice variants, is the 30 to 33 amino acid-stretch in the macrodomain encoded by the alternative exon 5 (**Figure 4A**). Comparison of the corresponding amino acid sequences in the non-vertebrate macroH2A indicated that the ancestral macroH2A is most similar to macroH2A1.1 (**Figure 10B**). Importantly, this included a high level of conservation of amino acids required for ADP-ribose binding in human macroH2A1.1, such as aspartate 203 (D203) and glycine 224 (G224) (Kustatscher et al., 2005).

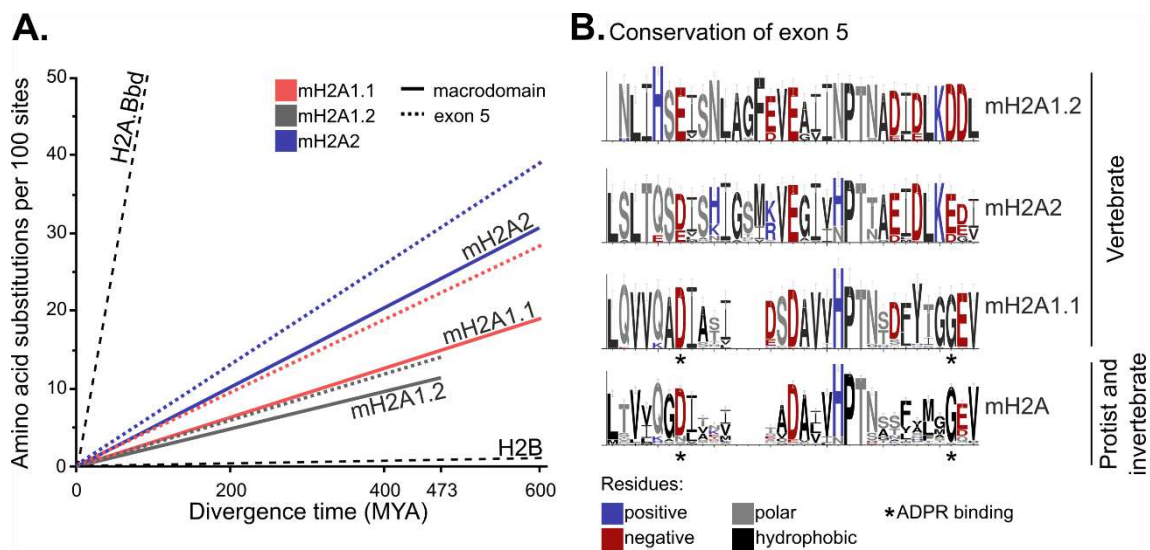


Figure 10. The evolution rate of macroH2A macrodomains and exon 5 corresponding amino acid sequence.

(A) Estimated rates of evolution for macroH2A proteins in vertebrates. For each macroH2A variant, the evolutionary rate of the macrodomain is depicted. Pairwise amino acid identities are represented based on the species divergence times as defined in the TimeTree database. Evolutionary rates for fast-evolving histone H2A.Bbd, slow-evolving histones H2B were included as references. MYA, million years ago. **(B)** Logo plots comparing the amino acid sequence encoded by exon 5 of macroH2A from vertebrate macroH2A1.1, macroH2A1.2 and macroH2A2 and invertebrate macroH2A (including protists). Residue colors are based on their biochemical properties: positively charged in blue, negatively charged in red, hydrophobic in black and polar in gray. Residues important for ADP-ribose (ADPR) binding in vertebrate macroH2A1.1 are marked by asterisk.

Our results allowed us to describe for the first time the evolutionary order of events that resulted in the three different macroH2A histone variants present in vertebrates (**Figure 9A, 9B**). Importantly, we found that macroH2A is much older than previously reported and originated in pre-metazoan protists. The expansion and conservation of macroH2A co-occurred with, and thus might be important for, the emergence of multicellularity in animals. Excitingly, the first macroH2A gene is similar to vertebrate macroH2A1.1. This suggests that the function of macroH2A histone variant in nuclear NAD⁺ metabolism might be ancient.

3.2.2 The structural conservation and ADP-ribose binding of the protist macroH2A macrodomain

Given the origin of macroH2A before metazoans (animals), we sought to determine its potential ancestral metabolic implication by determining the biochemical properties of macroH2A in one of the protist organisms. As a model system we used *Capsaspora*, one of the closest unicellular relatives of animals which displays a multicellular life stage (Sebé-Pedrós et al., 2013). We produced and purified the macroH2A macrodomain from *Capsaspora*. As a reference for a vertebrate macrodomain, we used the macrodomain of mouse macroH2A1.1. As an initial, qualitative method for the detection of interaction with ADP-ribose, we used saturation transfer difference NMR (**Figure 11A**). We found that the *Capsaspora* macrodomain bound ADP-ribose, showing a similar binding spectrum to that of the wild-type murine macroH2A1.1 macrodomain (**Figure 11B**). Mutations introducing negatively charged residues into the ADP-ribose binding pocket, such as G224E, are known to abolish ADP-ribose binding in the human protein (Kustatscher et al., 2005). Similarly, the corresponding mutation in the mouse macroH2A1.1 abolished its interaction with ADP-ribose (**Figure 11B**).

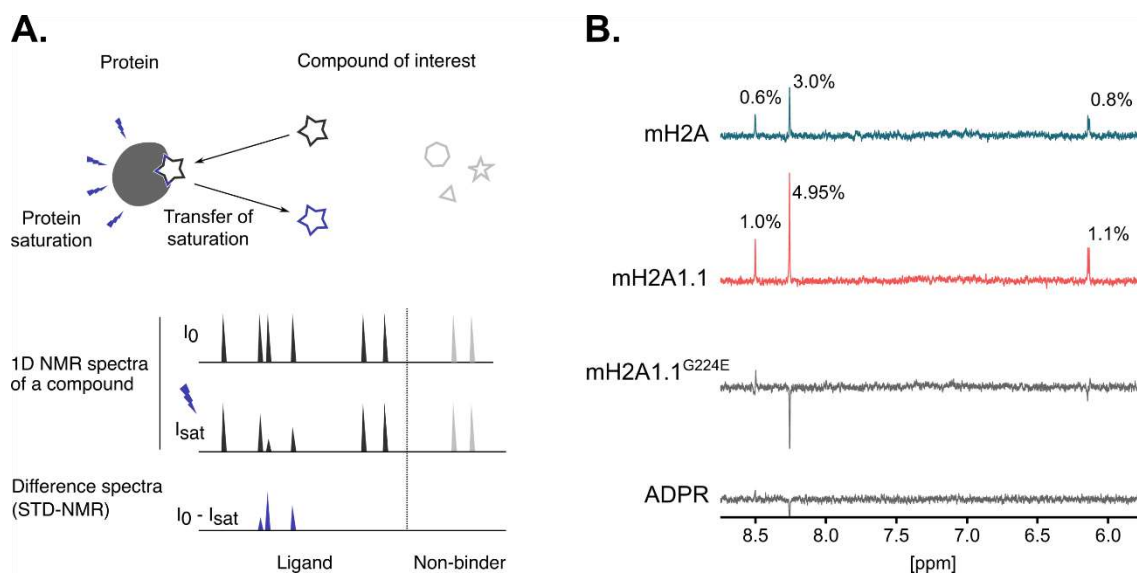


Figure 11. Saturation transfer difference NMR (STD-NMR) for detection of ADP-ribose binding.

(A) Scheme of the saturation transfer difference nuclear magnetic resonance (STD-NMR) experiment, which involves recording two analogous 1D NMR spectra: the reference spectrum of a ligand without saturation of protein signals (irradiating at a position far enough from any signal in the ^1H spectrum to avoid saturation), and a spectrum recorded after selective saturation of the protein. The saturation of the protein can be transferred to the small molecule if transiently bound to the protein, causing a decrease in the signal intensity of binders. The difference spectrum shows only the signals of the binders as plotted in Figure 2A and Supplementary Figure S1B. For non-binders there is no saturation transfer to the ligand and the intensities in the off- and on-resonance spectra are identical and cancel out in the difference spectrum. **(B)** STD-NMR experiments indicate the presence of interaction between *Capsaspora* macroH2A macrodomain (mH2A) and ADP-ribose (ADPR) comparable to the interaction with mouse macroH2A1.1 macrodomain (mH2A1.1) and ADPR. No interaction is observed between ADPR and the loss-of-function mutant mH2A1.1^{G224E} that we included as a control. STD spectrum for ADPR in the absence of protein is shown as a reference. The STD effect (%) is quantified as $I_{STD} = 100 * (I_0 - I_{sat}) / I_0$.

We further used isothermal titration calorimetry to characterize the interaction. *Capsaspora* macrodomain bound ADP-ribose eight times stronger than the macrodomain of mouse macroH2A1.1, with equilibrium dissociation constants of 1,3 μM and 11,3 μM , respectively (**Figure 12A, 12B**). The control experiment shows no significant contribution of heat of dilution (**Figure 12C**). The thermodynamic profiles indicated that the two macrodomains bind ADP-ribose using a different binding mode. *Capsaspora* macrodomain bound ADP-ribose through favourable enthalpic and entropic contributions, whereas the mouse macroH2A1.1 macrodomain bound it mainly in an enthalpy-driven manner, which was partially compensated by an unfavourable entropic contribution (**Figure 12D**). Furthermore, *Capsaspora* macroH2A macrodomain showed high

selectivity for ADP-ribose binding, since its affinity towards ADP was ~50-fold lower, and no interaction was observed with related nucleotides (ATP, AMP, GDP), nor ribose (**Figure 12E**).

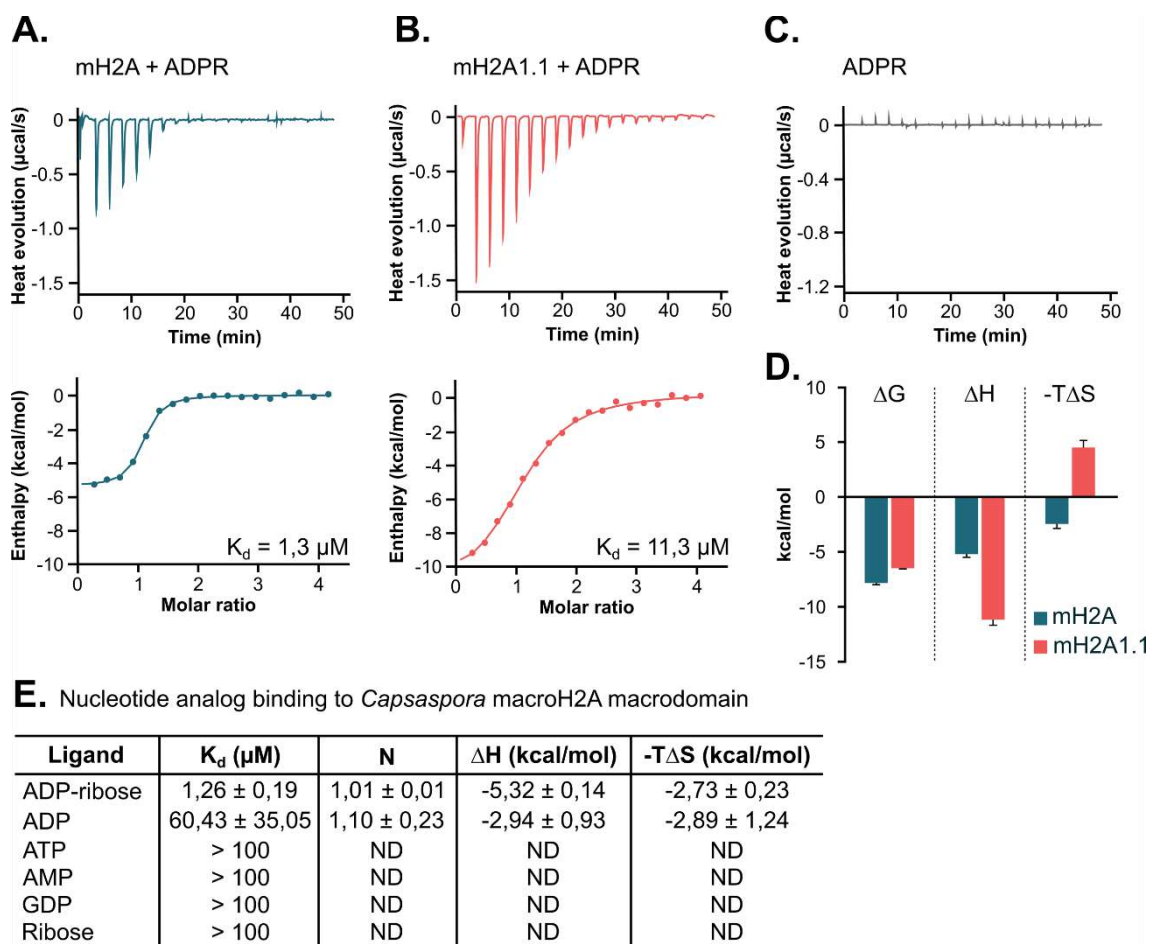


Figure 12. Isothermal titration calorimetry (ITC) for detection of ADP-ribose binding.

(A) Representative plot of raw heat evolution in ITC experiment upon ADP-ribose titration into a solution with purified *Capsaspora* mH2A macrodomain. The integration of the raw heat evolution results in a curve representing the equilibrium binding isotherm, represented below the raw heat evolution plot with indicated K_d value. **(B)** Same as in (A), but for mouse mH2A1.1 macrodomain. **(C)** Representative plot of raw heat evolution in isothermal titration calorimetry (ITC) control experiment upon ADP-ribose titration into the buffer alone as a control. **(D)** Thermodynamic parameters of experiment in (A) and (B) shown as signature plots indicate different contributions of enthalpy (ΔH) and entropy ($-T\Delta S$) to the Gibb's free energy (ΔG). Errors are calculated from four independent experiments. **(E)** Table showing the results of ITC binding affinity measurements (K_d , N, ΔH and $-T\Delta S$) of *Capsaspora* mH2A macrodomain and a range of nucleotides (ATP, ADP, AMP, GDP) and ribose. K_d values for ATP, AMP, GDP and ribose are estimated to be higher than $100\mu\text{M}$, as their ITC binding curves show weak or absence of binding. If no binding occurred or binding was below ITC detection limit, reaction stoichiometries and thermodynamics could not be determined (ND).

To characterize the *Capsaspora* macroH2A macrodomain and its interactions with ADP-ribose at the atomic level, we solved the structure of the protein in the presence and the absence of ADP-ribose by protein X-ray crystallography (**Figure 13A, 13B**). The unliganded macroH2A macrodomain crystallized in the space group P12₁1 and could be refined to 1.4 Å resolution, while the ADP-ribose bound protein crystallized in the space group P3₂21 and was refined to 2.0 Å resolution. The obtained globular structures, which are composed of seven central β-sheets in the characteristic 1276354 order surrounded by α-helices, show high structural similarity to the previously described macrodomains (Allen et al., 2003; Karras et al., 2005; Kustatscher et al., 2005), with a root mean square deviation (RMSD) of only 0.5 Å between the ADP-ribose bound structures of *Capsaspora* and human macroH2A macrodomains (PDB ID: 3IID). Surprisingly, the C-terminal α-helix of the *Capsaspora* macroH2A macrodomain is not visible in the electron density of the ADP-ribose bound structure. This prevented the modelling of the C-terminal 13 amino acids, while the apo structure could be fully refined all the way to the C-terminus of the macrodomain (**Figure 13C, 13D**). Whether this is the result of a conformational change upon ligand binding or an artefact from the crystal packing remains to be determined. However, we previously observed that ADP-ribose binding induces a conformational change in the human macroH2A1.1 macrodomain, specifically by a 30° rotation of the C-terminal α-helix away from the globular macrodomain fold (Timinszky et al., 2009a). We surmise that a similar induced fit is likely triggered by the binding of ADP-ribose to the *Capsaspora* macroH2A.

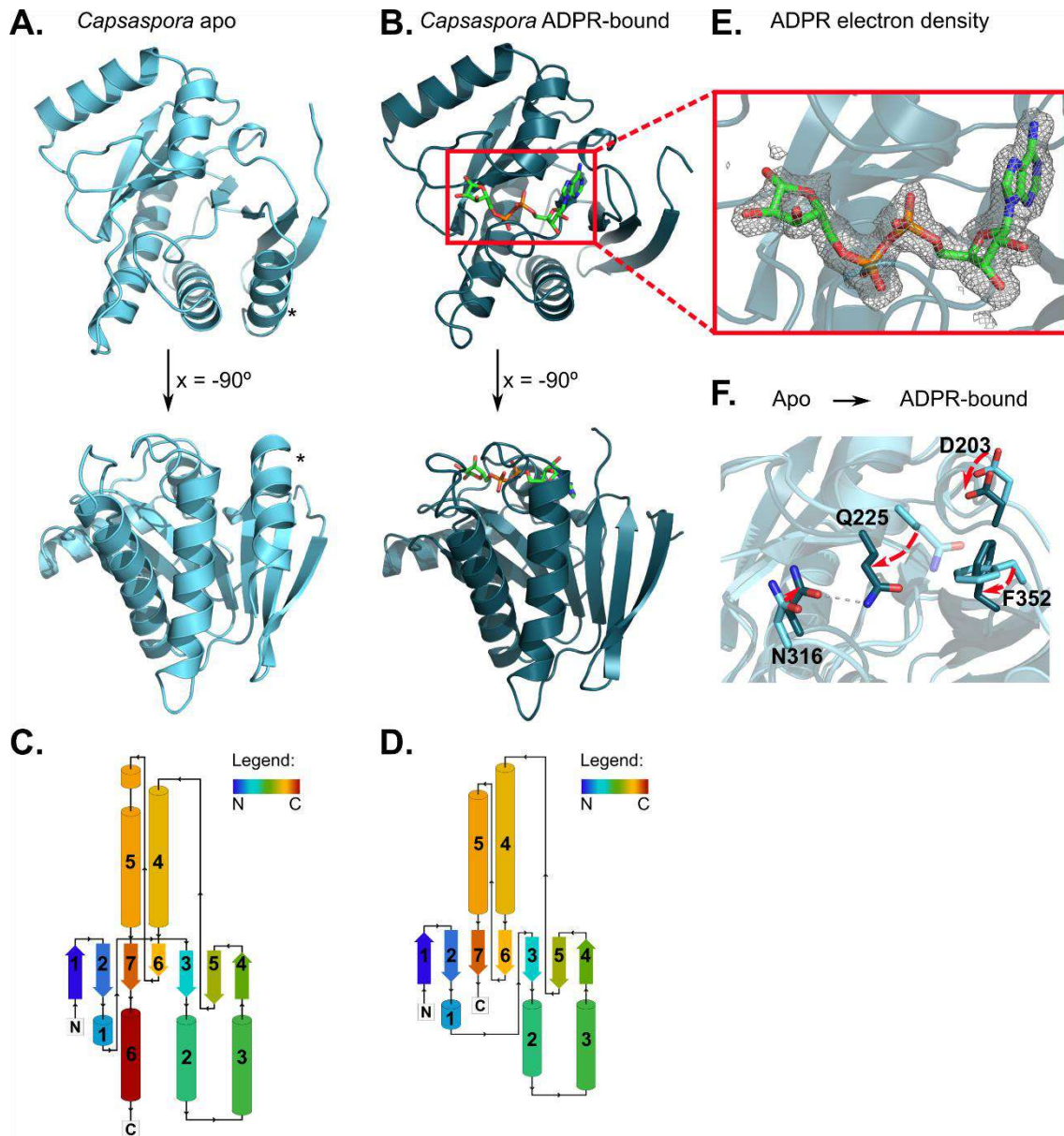


Figure 13. Crystal structures of *Capsaspora macroH2A* macrodomains with and without ADP-ribose.

(A) Crystal structure of apo-form of *Capsaspora* mH2A macrodomain (PDB 7NY6) showing entrance to ADP-ribose binding pocket, and rotated by -90° around x-axis. C-terminal α -helix is marked by asterisk. **(B)** Crystal structure of *Capsaspora* mH2A macrodomain in complex with ADPR (PDB 7NY7) in same orientation as in **(A)**. **(C)** Topology diagram of apo-form of *Capsaspora* mH2A macrodomain. The topology diagram shows β -sheets in a conserved 1276354 order, surrounded by 6 α -helices. **(D)** Topology diagram of *Capsaspora* mH2A macrodomain in complex with ADP-ribose (ADPR). The topology diagram shows β -sheets in a conserved 1276354 order, surrounded by 5 α -helices. The C-terminal α -helix observed in the apo-structure **(A)** was not detected in the ADPR-bound structure. It remains to be determined if this reflects a conformational change upon ligand binding or an artefact from the crystal packing. **(E)** Composite omit 2Fo-Fc electron density map (gray mesh, contoured at 1σ) of ADPR bound to the *Capsaspora* macroH2A macrodomain. **(F)** Zoom into the binding pocket of superimposed crystal structures of apo- (light blue) and ADPR-bound (deep-teal) forms of *Capsaspora* macroH2A macrodomains shows significant reorganization upon ligand binding. Arrows indicate the change in position from apo to bound for residues D203, Q225, N316 and F352. The side chain of Q225 is not modeled due to ambiguous electron density. One possible conformation is shown as a transparent stick.

The electron density of ADP-ribose clearly revealed that the ligand situates within the canonical binding pocket of macroH2A macrodomain (**Figure 13E**). The residues important for ligand binding included aspartate 203 (D203), which established a hydrogen bond (H-bond) with the adenine amino-group, and phenylalanine 352 (F352), which stabilized the adenine ring via π -electron stacking of the aromatic rings. Further, the amino group of the side chain of asparagine 316 (N316) establishes an H-bond with the distal ribose of ADP-ribose (**Annex 8.1A**), which can explain the selectivity for ADP-ribose over ADP (**Figure 12E**). Upon ADP-ribose binding, the side chains of glutamine 225 (Q225) and N316 move towards each other and establish an H-bond, resulting in a conformation that encloses the central diphosphate moiety of ADP-ribose (**Figure 13F**).

In summary, the capacity of the macroH2A macrodomain to bind ADP-ribose is conserved in the protist *Capsaspora*. Indeed, the *Capsaspora* macroH2A macrodomain bound ADP-ribose stronger than its vertebrate counterpart, although the overall globular structure of the macrodomains was similar. This raises the possibility that evolution has selected for decreased ADP-ribose affinity in higher organisms along the vertebrate stem lineage.

3.2.3 Two residues critical for the induced-fit closing of the binding pocket are the key evolutionarily divergent feature

Next, we determined the importance of the conservation of the amino acid sequence in the context of ADP-ribose binding and the conserved structure of the macrodomain. We first analyzed the homology of macroH2A macrodomain sequences from *Capsaspora*, *Pygmaia* and the representative species of 7 animal phyla. For vertebrates, we used sequences corresponding to the ADP-ribose binding isoform, macroH2A1.1. The pairwise sequence identity of these representative macroH2A macrodomains varied between 31 and 98%, with the macrodomains of *Capsaspora* and human sharing only 50% identity at the level of the amino acid sequence (**Figure 14**).

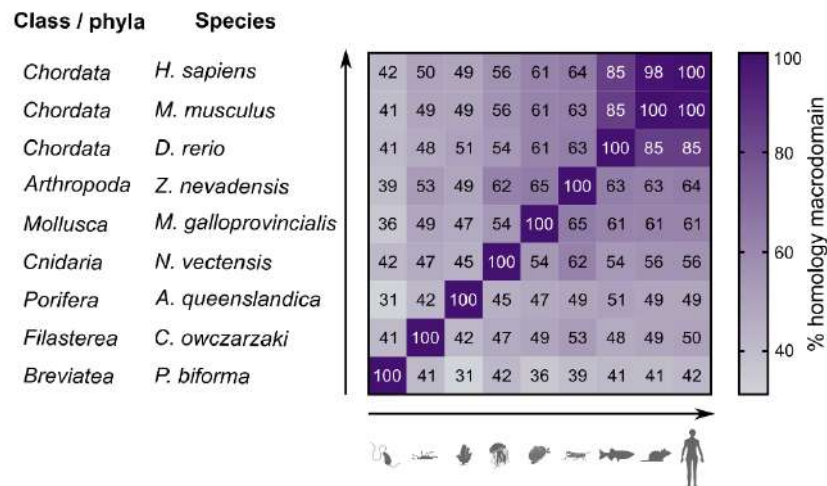


Figure 14. Pairwise sequence homology analysis of representative macroH2A macrodomains.

Pairwise sequence homology of macrodomains from macroH2A representatives of 9 different phyla. For the vertebrates, amino acid sequences corresponding to macroH2A1.1 have been used.

However, the multiple sequence alignment of more than 300 macroH2A macrodomain sequences delineated two well conserved regions of the protein, one located toward the N-terminus and one toward the C-terminus of the macrodomain fold (**Figure 15A**). The N-terminal region highly overlaps with vertebrate exon 5, whose alternative splicing is the critical determinant for ADP-ribose binding of macroH2A1.1 (Kustatscher et al., 2005). Comparison of these two regions between *Capsaspora* macroH2A and human macroH2A1.1 revealed

that they are comprised of a large number of invariant or structurally related amino acids, indicating a high prevalence of conservative substitutions. To zoom in on the conservation of the functionally-relevant ADP-ribose binding surface, we marked all the residues that directly interacted with ADP-ribose in the human (Kustatscher et al., 2005; Timinszky et al., 2009a) and *Capsaspora* proteins (**Figure 15A, Annex 8.1A**). Interestingly, with the exception of serine 275, all ADP-ribose-interacting residues were located in the conserved N- and C-terminal regions (**Figure 15A**). Additionally, the *Capsaspora* macroH2A macrodomain established a lower number of bonds with ADP-ribose compared to the human ADP-ribose-bound macroH2A1.1 macrodomain (**Annex 8.1B**). This is consistent with the lower enthalpic contribution of the *Capsaspora* macrodomain for ADP-ribose binding (**Figure 12D**). However, the majority of the interactions shared between the human and *Capsaspora* macrodomains involved highly conserved residues (**Figure 15A**). Further mapping of the conservation rate on the crystal structure of the apo-form of *Capsaspora* macroH2A macrodomain, indicated a high conservation of the inner part of the ADP-ribose binding pocket (**Figure 15B**), while the surface regions were more variable (**Figure 15C**).

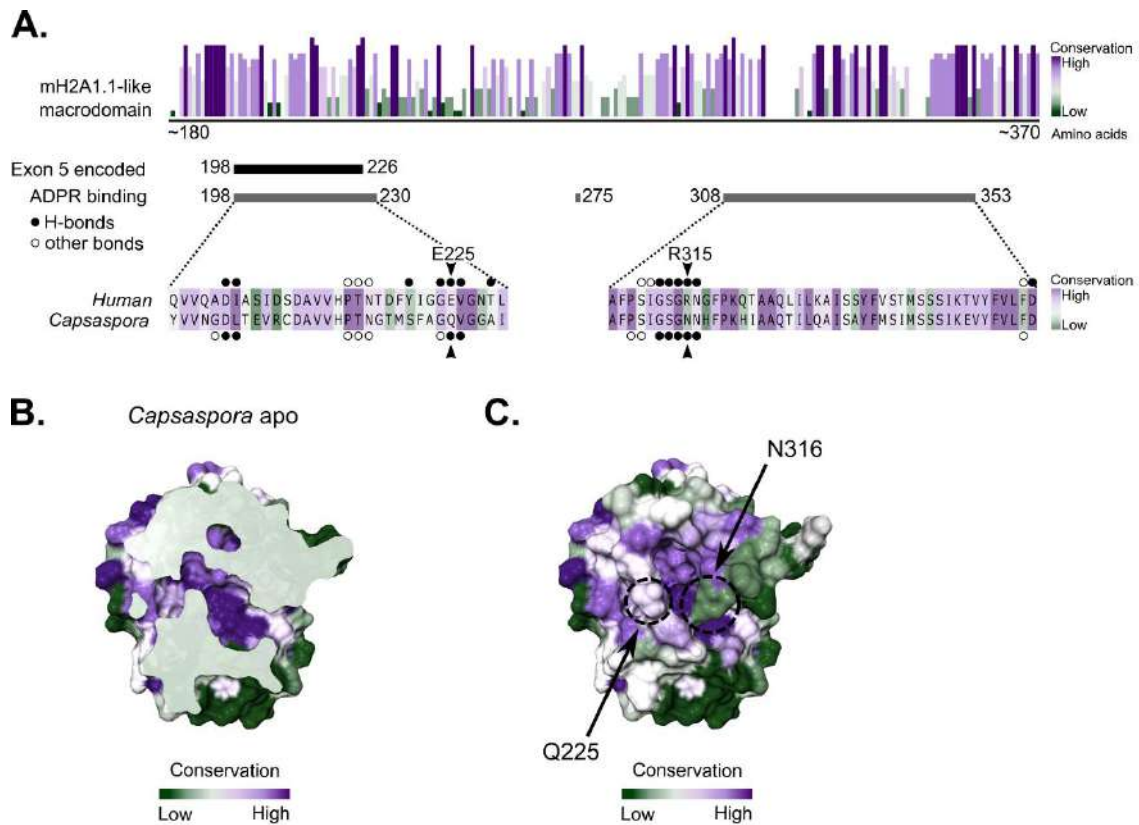


Figure 15. Sequence and structure conservation of macroH2A1.1-like macrodomains through evolution.

(A) Multiple sequence alignment of >300 macroH2A1.1 (mH2A1.1)-like macrodomains indicates two highly conserved regions which highly overlap with ADPR-binding residues. Amino acids encoded by exon 5 and residues establishing bonds with ADPR in *Capsaspora* macroH2A (mH2A) macrodomain or human mH2A1.1 are indicated. Arrows point to the two key less-conserved residues involved in ADPR binding, namely Q225 and N316 in *Capsaspora* macroH2A, which correspond to residues E225 and R315 in human mH2A1.1. (B) Consurf analysis of structural conservation of the binding pocket of mH2A1.1-like macrodomains mapped on the apo-form of *Capsaspora* mH2A macrodomain (PDB 7NY6). The macrodomain is horizontally clipped to reveal the interior of the binding pocket. (C) Consurf analysis of structural conservation of mH2A1.1-like macrodomains as in (B), representing the whole macrodomain and with outlined the less conserved amino acids Q225 and N316.

The key residues with a low degree of conservation that are involved in ADP-ribose binding were Q225 and N316 in the *Capsaspora* macroH2A, which correspond to glutamate 225 (E225) and arginine 315 (R315) in human macroH2A1.1. Both pairs of residues close the binding pocket upon ADP-ribose binding. However, the polar uncharged side chains of Q225 and N316 establish an H-bond (Figure 13F), while the corresponding residues E225 and R315 in human macroH2A1.1 establish a salt bridge to close the binding pocket upon ADP-ribose binding (Kustatscher et al., 2005). Detailed comparisons revealed very few conformational differences between the macrodomain of *Capsaspora*

macroH2A and its human ortholog. The electron density of the Q225 and N316 side chain were weaker compared to adjacent side chains, indicating conformational heterogeneity of these residues, which is in good agreement with the respective side chains in the human macroH2A1.1 macrodomain structures (**Figure 16A**). The side chains of F352 had a slightly different orientation, while D203 shows a similar orientation. Most interestingly, upon ADP-ribose binding, both macrodomains adopt an almost identical conformation (**Figure 16B**).

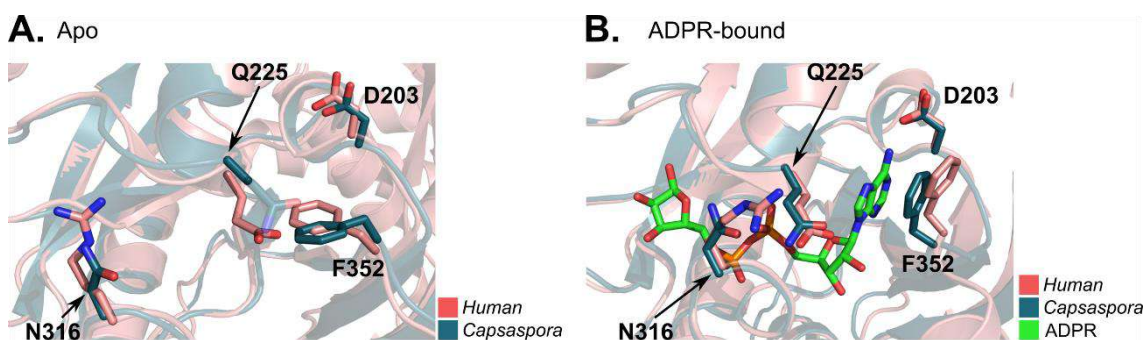


Figure 16. Structural comparison of *Capsaspora* macroH2A and human macroH2A1.1 macrodomains.

(A) Overlay of apo-forms of the macrodomains of *Capsaspora* macroH2A (dark-teal, PDB: 7NY6) and human macroH2A1.1 (light red, PDB: 2FXK) highlighting residues participating in the binding of ADPR. Indicated residue numbers refer to the amino acids in *Capsaspora* protein. **(B)** Same as in **(A)** for the ADPR-bound forms (*Capsaspora* macroH2A PDB: 7NY7; human macroH2A1.1 PDB: 3IID). ADPR is shown in green.

Taken together, the low *a priori* conservation at the level of primary sequence contrasts with the highly conserved three-dimensional structure of the *Capsaspora* and human macrodomains. The residues forming the inner ADP-ribose binding pocket are highly conserved between *Capsaspora* and vertebrates. The identification of two key residues that participate in ADP-ribose binding, but are less conserved, provides a potential explanation for the observed difference in affinity between human and *Capsaspora* macroH2A macrodomains. Intriguingly, these two non-conserved residues correspond to *Capsaspora* Q225 and N316, which are involved in an induced-fit closing of the binding pocket of the macroH2A macrodomain upon interacting with ADP-ribose.

3.2.4 The selection of two residues fine-tuned ADP-ribose affinity through evolution

We hypothesized that the specific substitution of amino acids Q225 and N316 drove the evolution towards the decreased ADP-ribose binding affinity from *Capsaspora* to vertebrates. To test this, we focused on 305 distinct species with a single macroH2A1.1-like isoform and analyzed the identity of the residues corresponding to the positions Q225 and N316 in the *Capsaspora* ortholog. We calculated the average presence of the respective residues in species from nine of the most represented animal phyla, and filastereans (**Figure 17**). Specifically, we classified the residues as protist-characteristic, when identical to *Capsaspora* macroH2A (Q225 and N316), and vertebrate-characteristic when identical to human and mouse macroH2A1.1 (E225 and R315), or other residues (**Annex 8.2**).

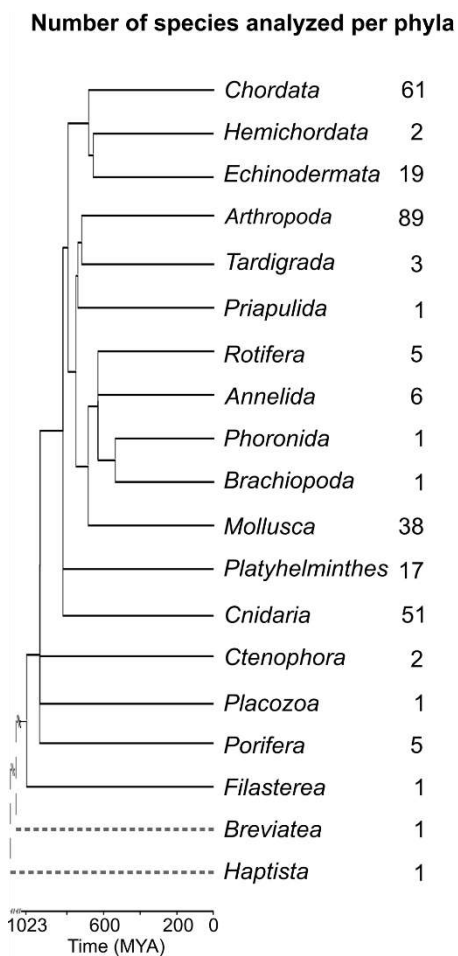


Figure 17. The identity and biophysical consequences of evolution of residues 225 and 316 closing the binding pocket.

Schematic phylogeny representing all phyla analyzed and the number of analyzed species per phyla with a single macroH2A1.1-like gene. Nine most represented phyla (more than 3 species per phyla) were used for the analysis shown in Figure 18A and 18B.

We found that the substitution of the protist-characteristic residue with the vertebrate-characteristic residue at the position 225 (Q225E) occurred first, while the transition at the position 316 (N316R) happened later during evolution and over an extended period of time (**Figure 18A, 18B**). The analysis of residue 225 showed that the majority of the sponges (*Porifera*), the earliest-branching metazoans, had the *Capsaspora*-characteristic Q225, suggesting a likely configuration of a shared ancestor. However, already the majority of the closely diverging *Cnidaria* species contained the vertebrate-characteristic E225 residue, while the molluscs (*Mollusca*) were the oldest animal phylum with a complete vertebrate-characteristic E225 configuration (**Figure 18A**). In contrast, the evolution of the residue 316 was much more variable and lengthy process (**Figure 18B**). Accordingly, the protist-characteristic N316 was still present in a substantial proportion of later diverging species, such as segmented worms (*Annelida*) and rotifers (*Rotifera*).

Additionally, we analysed the pairing of the residues involved in the induced-fit closing of the binding pocket through evolution. The established pairs between residues 225 and 316 show that the vertebrate E225 was accompanied by a non-vertebrate residue 316 in ~50% of analysed species (**Figure 18C**). On the other hand, the vertebrate-characteristic R316 paired with vertebrate-characteristic E225 or a similar residue in more than 90% of the species, thus corroborating the sequential evolution of amino acids at the two positions (**Figure 18D**).

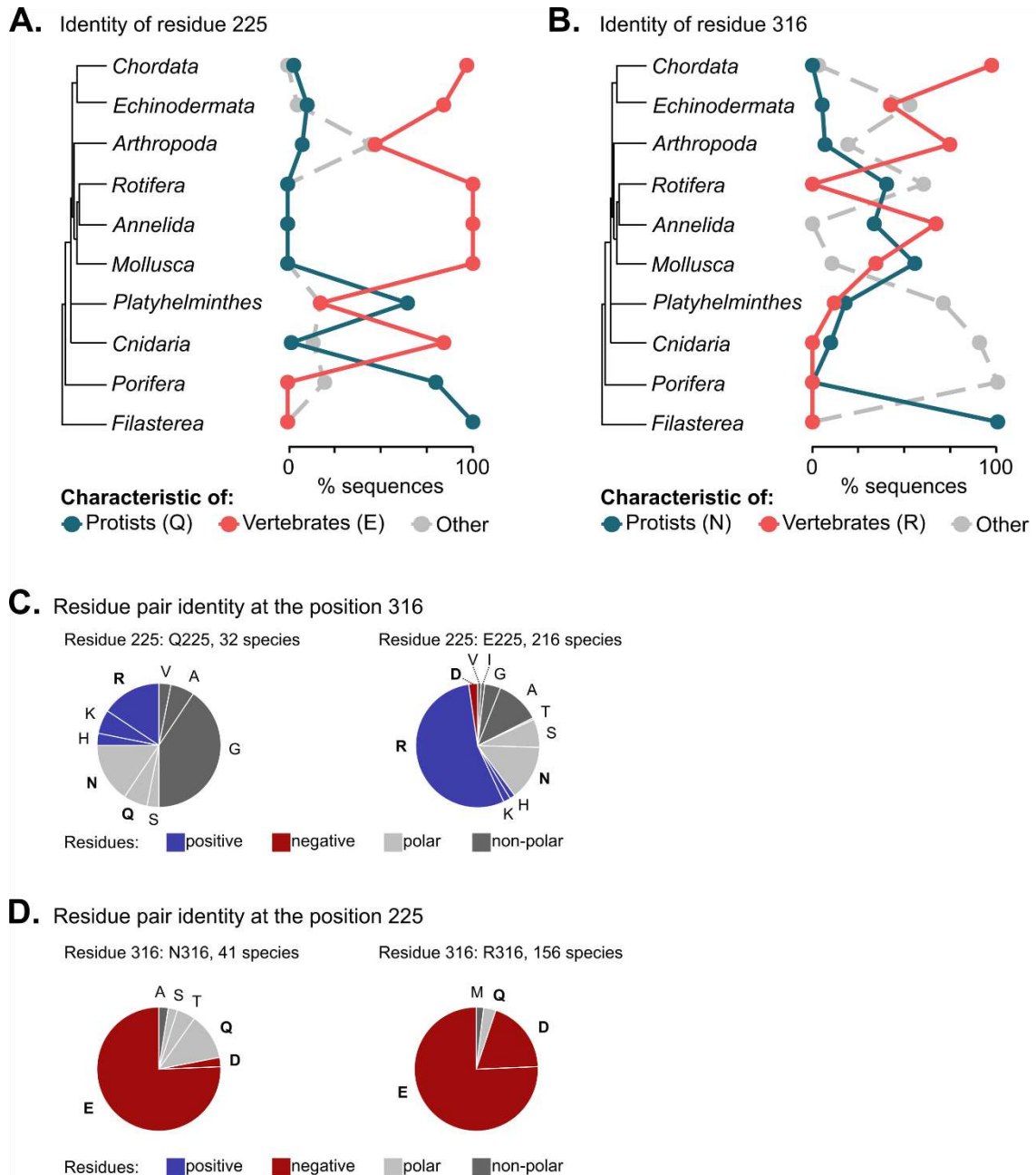


Figure 18. Comparative analysis of evolutionary rate of residues 225 and 316.

(A) The identity of the residue 225 in macroH2A1.1-like macrodomains is classified as protist-characteristic if identical to *Capsaspora*, vertebrate-characteristic or other. The percentage of species per phyla is plotted. (B) Same as (A) for residue 316. (C) The pie charts showing the identity of the residues at the position 316 which pair with either protist-characteristic Q225, or vertebrate-characteristic E225. The numbering of the residues is based on the *Capsaspora* mH2A. In particular, the protist-characteristic residue Q225 was paired with varying residues at 316. Residues are colored same as in (B). (D) The pie charts showing the identity of the residues at the position 225 which pair with either protist-characteristic N316, or vertebrate-characteristic R316. The numbering of the residues is based on the *Capsaspora* mH2A. Strikingly, R316 was paired with E225 or the biochemically similar D225 in more than 95% of the species, while its pairing with Q225 was rare. Residues are colored same as in (B) and (C).

To understand the physiological consequence of this evolutionary course, we individually introduced Q225E and N316R mutations in the *Capsaspora* macrodomain and confirmed that the mutant proteins were folded (results not shown). Interestingly, the Q225E substitution, which simulates the earlier evolutionary change, resulted in a 7-fold decrease of the affinity towards ADP-ribose compared to wild-type *Capsaspora* protein (**Figure 19A**). Thus, the affinity and the thermodynamic properties of ADP-ribose binding by the Q225E *Capsaspora* mutant were more similar to those of the mouse macroH2A1.1 macrodomain (**Figure 19B**). Strikingly, the N316R substitution lost its capacity to bind ADP-ribose, while in the *Capsaspora*-characteristic Q225 context (**Figure 19A**) and might explain why this combination was very rare in our dataset (**Figure 18D**).

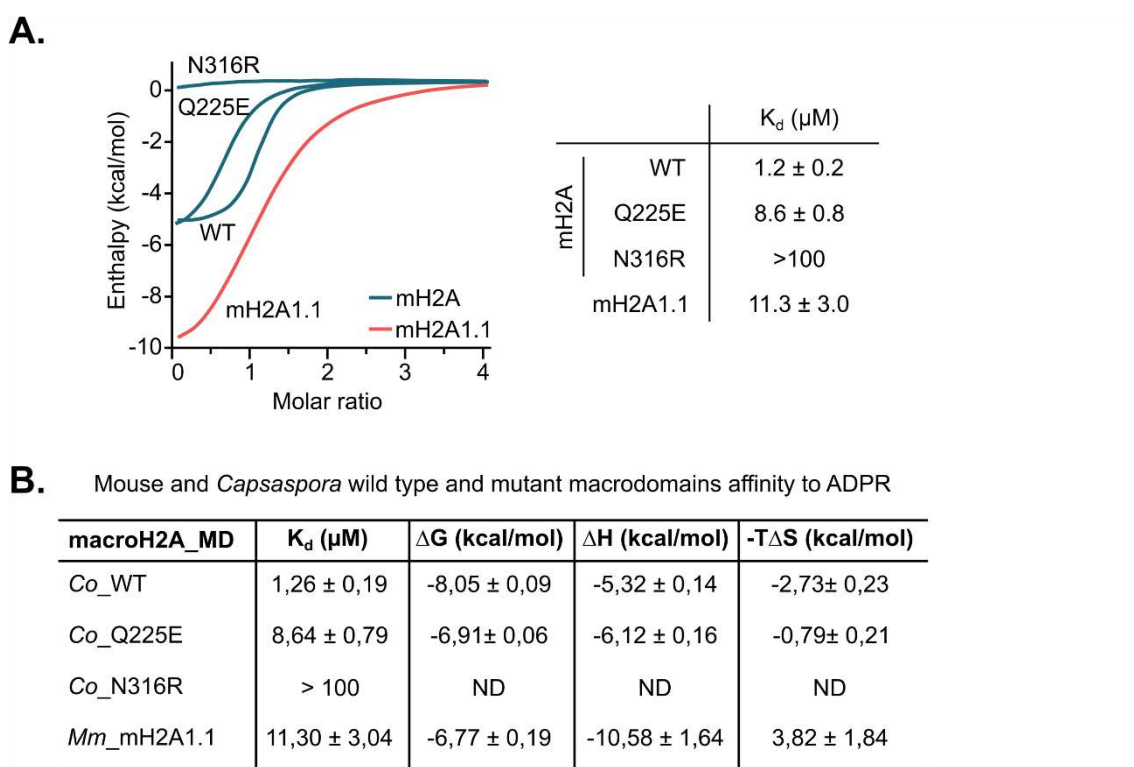


Figure 19. The binding measurements of *Capsaspora* mutant proteins with ADP-ribose.

(A) ITC of *Capsaspora* macroH2A macrodomain mutants (Q225E and N316R) with ADPR. Data of wild-type *Capsaspora* macroH2A macrodomain (WT) and the macrodomain of mouse macroH2A1.1 (mH2A1.1) from Figure 1B were included as reference. The calculated K_d values are indicated. A representative graph of four independent experiments is shown. **(B)** Table showing the thermodynamics of the binding interaction calculated on the basis of ITC data shown in (A).

Taken together, these results suggest that the evolution of the residues at these sites occurred sequentially and in a defined order: first Q225E and then N316R. Whereas the exchange of Q225E proved to be a determining factor for a decrease in ADP-ribose affinity of macroH2A1.1 over the course of evolution, the role and importance of the subsequent change from N316 to R remained unclear. We thus decided to take a closer look at *Mollusca*, the oldest non-deuterostome phylum with a more significant representation of the vertebrate-characteristic R316 residue (**Figure 18B**). Closer inspection of transcriptomics data indicated that a number of species contained two macroH2A isoforms as a result of an alternative splicing event (**Figure 20A**). Interestingly, the usage of the detected alternative exon 7 affected the C-terminal portion of the macrodomain and encoded a different amino acid at the position 316 (**Figure 20B**). In total, we were able to analyze data from 50 species from three different classes of molluscs. The alternatively spliced exon 7 was present in more than 25% of species in two classes, but absent from the third one (**Figure 20B, 20C**). Strikingly, in all of the mollusc species with two macroH2A isoforms the amino acid residue at the position 316 was almost exclusively either protist- or vertebrate-characteristic (**Figure 20D**).

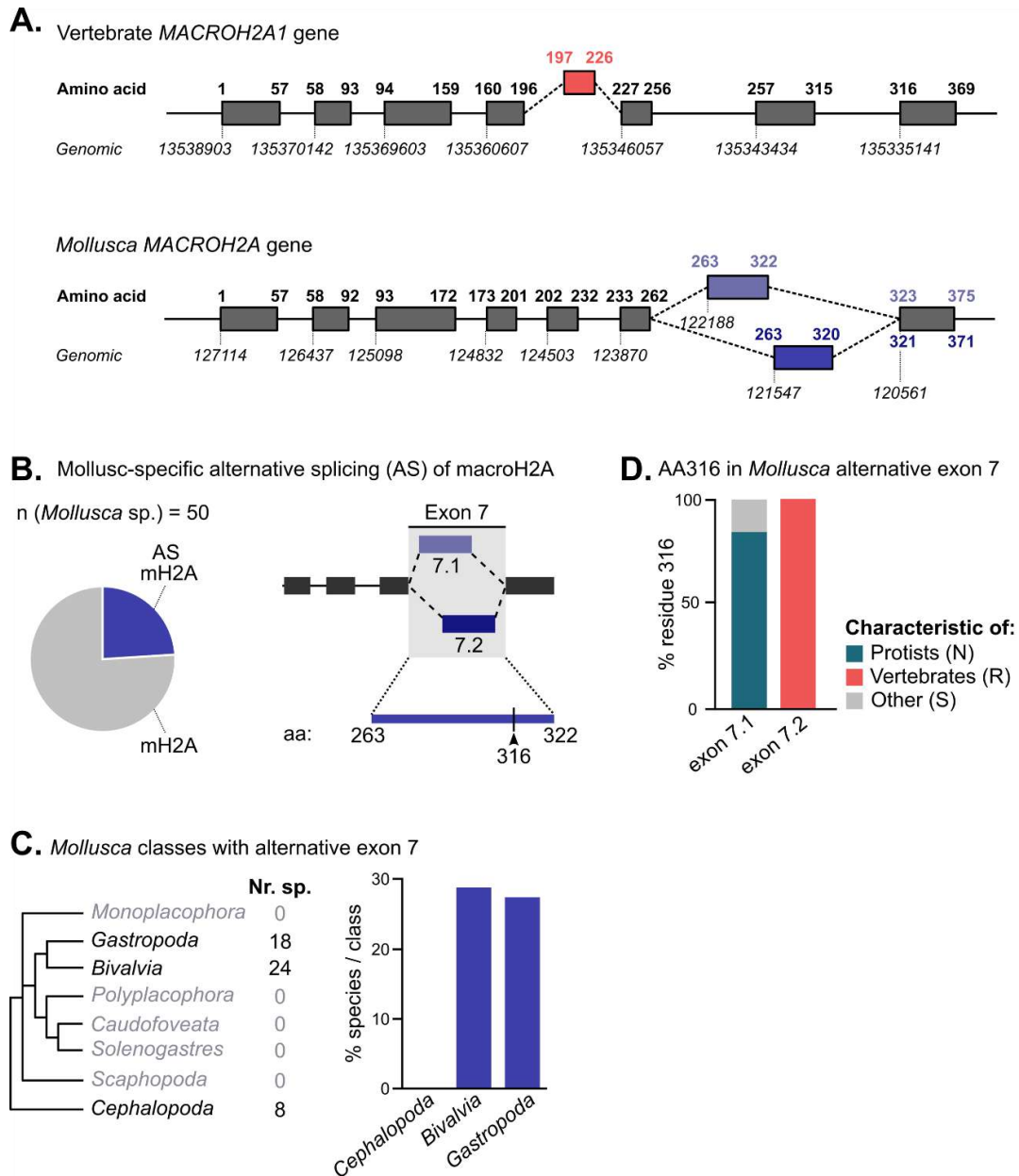


Figure 20. The analysis of the mollusk-specific alternatively spliced exon of macroH2A.

(A) The organization of the human *MACROH2A1* gene and the corresponding gene from the mollusc *Crassostrea gigas* (Gene ID: 105347164) are represented on an arbitrary scale for comparison purposes. For simplification only the alternative exon 5 defining macroH2A1.1 is represented for human *MACROH2A1*. Both alternative exons 7 of the *Mollusca* macroH2A gene are represented. (B) Pie chart showing that almost 25% of available Mollusca species undergo mollusc-specific alternative splicing of macroH2A (AS mH2A). Scheme in the right panel shows the alternatively spliced exon 7, specifying the amino acid residues this splicing affects. Residue 316 is outlined by an arrowhead. (C) Transcriptomic data was available for 50 Mollusca species from three of eight different classes. The number of species per class is indicated (left panel). The percentage of species per Mollusca class containing alternatively spliced macroH2A isoforms are shown (right panel). (D) For the subset of Mollusca species with alternative splicing of exon 7, the identity of the amino acid corresponding to 316 is plotted as percentage of species for both alternative exons. In the majority of cases, isoform 7.1 contains the protist-characteristic residue N316 and the isoform 7.2 has exclusively the vertebrate-characteristic R316.

Taken together, these results allowed us to delineate evolutionary order of events which affected the affinity of macroH2A macrodomains towards ADP-ribose. This occurred by changing the chemical identity of the two amino acids that close the binding pocket upon ADP-ribose binding. The Q225E substitution was determining and occurred early in the evolution of the protein, preceding the change of N316 to R. Further, the presence of alternatively spliced exons in molluscs hints at exon duplication as a potential molecular mechanism for decreasing affinity towards NAD-derived ADP-ribose in metabolically divergent higher organisms.

3.2.5 Dynamic regulation of *Capsaspora* metabolism across the life stages includes the macroH2A – PARP1 – NAD⁺ axis

To understand whether already the ancient macroH2A affected the macroH2A – PARP1 – NAD⁺ axis, as in the case of mammals (Hurtado-Bagès et al., 2018), we used *Capsaspora* as a model organism. *Capsaspora* undergoes a process of temporally-regulated cell differentiation. This results in a dynamic life cycle composed of three stages, namely filopodial, aggregative and cystic stage (**Figure 21A**). As a response to environmental cues, single cells from the filopodial stage transition to the aggregative stage, one of the simple forms of multicellularity. Cells from both these stages are proliferative. In less advantageous environments, they can transition to a spore-like cystic stage, which represents a resistance form with much smaller, non-proliferative cells (Sebé-Pedrós et al., 2013). The metabolic dynamics associated with these life-stage transitions are largely unknown. To describe the *Capsaspora* metabolism, we re-analyzed our previously generated stage-specific RNA sequencing data (Sebé-Pedrós et al., 2013) to which we added the information for mitochondrially encoded genes, which had not been included in the previous analysis. Then, we extracted the information on a curated list of more than 400 metabolic genes corresponding to KEGG pathways. The large majority of all analyzed metabolic genes were differentially expressed in at least one of the three life stages. Differentially expressed genes could be grouped in two larger and two smaller clusters by unsupervised clustering (**Figure 21B**). The larger groups contained

103 and 192 genes that were specifically up- or down-regulated, respectively, in the cystic stage in comparison to filopodial and aggregative stages. We refer to these groups as Cys high and Cys low. 25 and 29 metabolic genes formed the smaller clusters that were differentially expressed between filopodial and aggregative stages, placing the cystic stage in an intermediate position. These results suggested that the three life-stages differ in their metabolism, with the cystic stage having the most divergent metabolic state.

When classifying metabolic genes according to their function, we noted that almost one fifth of the genes in the Cys high cluster were not related to anabolism or catabolism, rather to other metabolic processes (**Figure 21C**). The majority of these genes encode autophagy-related V-ATPase proton pumps related to cell acidification. The relative proportion of anabolism-related genes was the same in the Cys high and the Cys low gene clusters (**Figure 21C**), but they had a different metabolic implication (**Figure 21D**). Most strikingly, NAD and NADP biosynthesis pathways were strongly overrepresented in the Cys high cluster (**Figure 21D**). Additionally, the Cys high-overrepresented genes involved in amino acid biosynthesis were related to very specific amino acids, namely tyrosine and phenylalanine, both of which can serve as substrates for the *de novo* biosynthesis of NAD (Canto et al., 2015; Dosselaere & Vanderleyden, 2001). On the other hand, amino acid synthesis genes present in the Cys low were related to a wide variety of amino acids. Furthermore, although the genes involved in catabolism had a relatively lower representation in the Cys high cluster than in the Cys low, they seemed to be more directed towards efficient energy production. Specifically, the nuclear genes encoding for the components of the respiratory chain were represented in both clusters, while those encoded by the mitochondria were exclusively found in the Cys high (**Figure 21D**). Taken together, the cells in the cystic stage might be channelling the NAD biosynthesis to retain the life-sustaining catabolic processes.

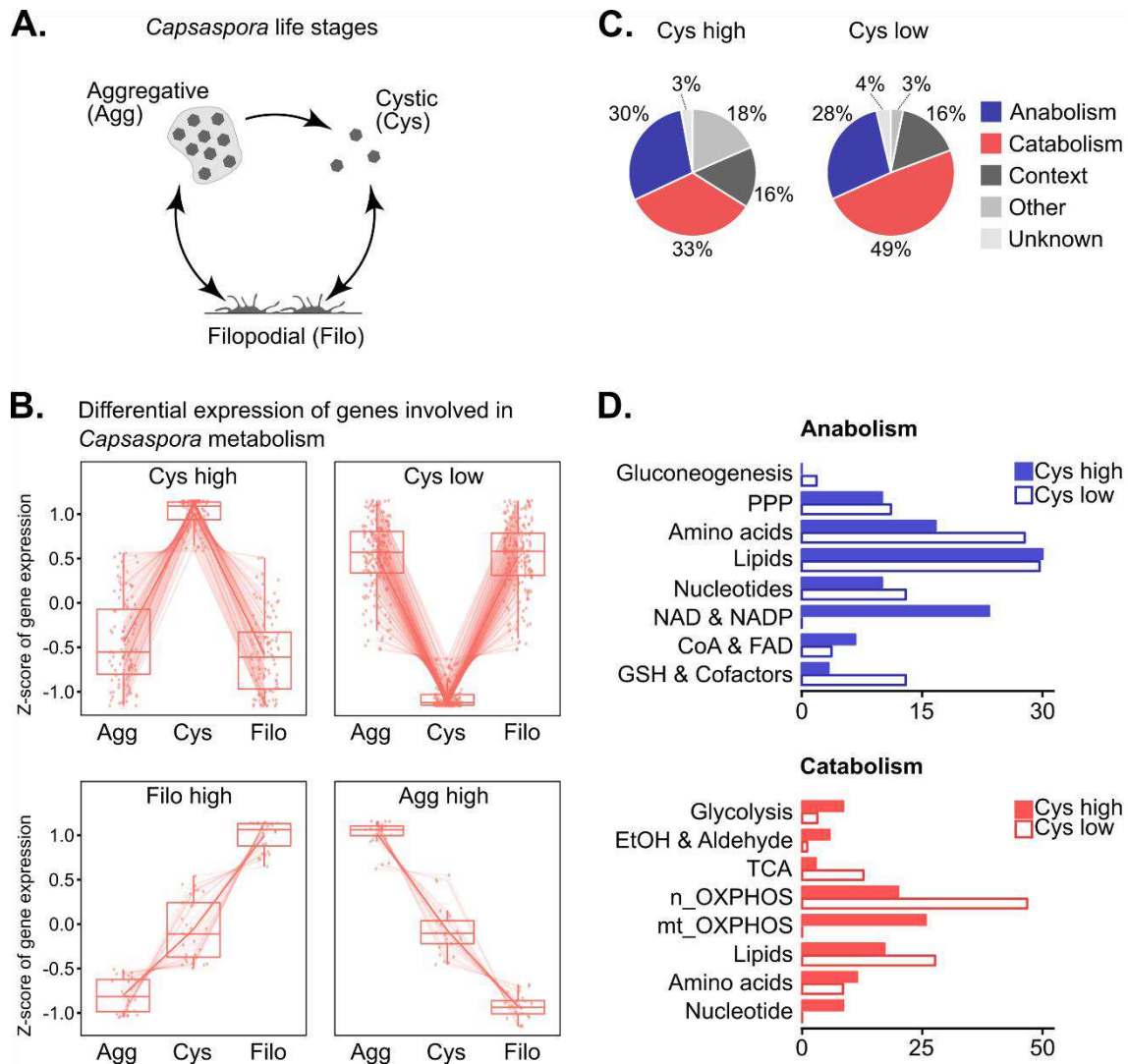


Figure 21. Metabolically dynamic life cycle of *Capsaspora*.

(A) *Capsaspora* life cycle has three life stages: aggregative (Agg), Filopodial (Filo) and cystic (Cys). Cys stage is spore-like and non-proliferative. **(B)** The majority of metabolic genes are differentially expressed in the three life stages of *Capsaspora* and fall into 4 major groups: Cys high, Cys low, Filo high and Agg high. **(C)** Metabolic genes were classified as anabolic, catabolic, context-dependent, other and unknown. Pie charts indicate their proportion in Cys high and Cys low clusters identified in **B**. **(D)** Column charts show the relative contribution of different pathways to the anabolic (top panel) and catabolic (bottom panel) component of the group of genes in Cys high and Cys low. The total number of anabolic and catabolic genes shown in **C** was set to 100%.

Intrigued by the differential regulation of NAD and NADP biosynthesis in the three life stages of *Capsaspora*, we examined the dynamic expression of genes encoding for proteins involved in NAD⁺ dependent processes, including PARP1 and its nuclear inhibitor macroH2A. To be more systematic in this process, we grouped the genes in eight categories: macroH2A, other macrodomain-containing proteins, PARPs, PAR glycohydrolases (PARG), PAR-related DNA

repair proteins, sirtuins, NUDIX proteins and proteins involved NAD biosynthesis. Unsupervised clustering distinguished 4 clusters related to their differential expression in the three life stages (**Figure 22**). Cluster 1 and 2 contained genes with increased expression in the two proliferative stages, and decreased expression in the non-proliferative cystic stage (**Figure 22**). In line with the proliferative characteristics, they included most of the genes encoding for PARP enzymes and other components of the DNA repair machinery, which might allow cells to cope with replication-associated DNA damage. Clusters 3 and 4 contained genes more prominently expressed in aggregative and cystic stages. This included genes encoding components of biosynthetic NAD pathways such as the salvage enzyme NAMPT and the *de novo* synthesis enzyme QPRT (**Figure 22**).

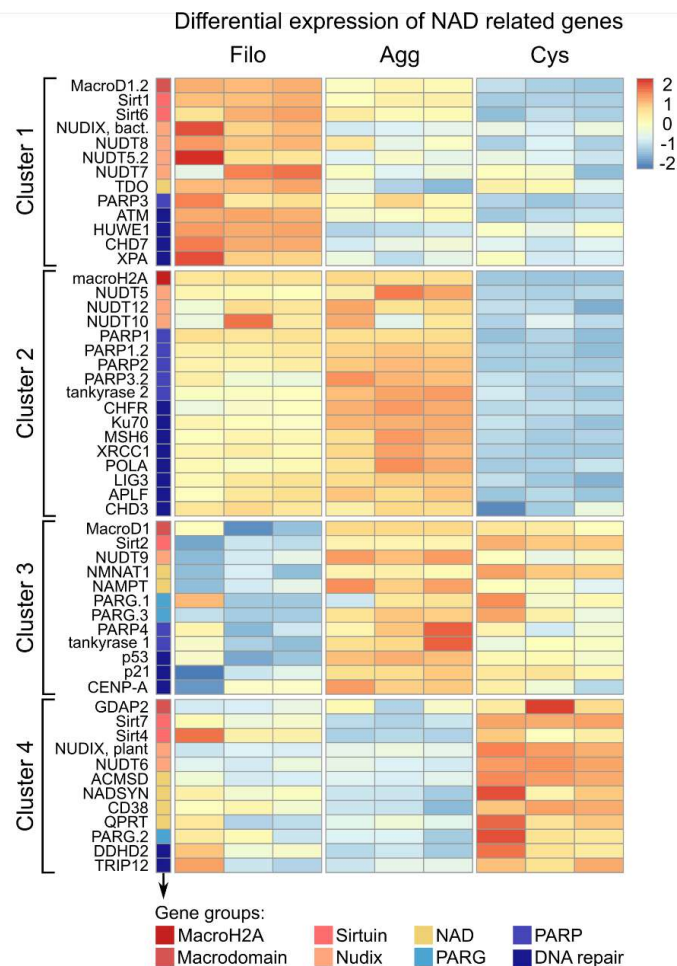


Figure 22. Differential expression of NAD metabolism-related genes in *Capsaspora*. Differentially expressed genes related to NAD⁺ metabolism group into 4 clusters. Genes were classified in 8 groups and color coded as shown in the legend: macroH2A, other macrodomain proteins, sirtuins, Nudix proteins, PARPs, PARG-type hydrolases, DNA repair and NAD biosynthesis.

Noteworthy, the expression patterns of PARP1 and macroH2A in cluster 2 were highly similar, suggesting co-regulation at the transcriptional level (**Figure 22**). However, we know from vertebrate cells that the relative protein levels of PARP1 and macroH2A1.1 are decisive, and that macroH2A1.1 needs to be equally or more abundant than PARP1 to efficiently inhibit its activity and dampen the nuclear NAD⁺ consumption. To assess the protein levels, we generated antibodies specific for *Capsaspora* PARP1 and *Capsaspora* macroH2A (**Figure 23A, 23B**). PARP1 was readily detected in both proliferative stages, but its levels dropped in the cystic stage (**Figure 23C**). For macroH2A, we detected a doublet at the expected size, which collapsed into a single band in the cystic stage (**Figure 23C**). To confirm this change in ratio, we extracted the information on PARP1 and macroH2A from available shotgun mass proteomic data (Sebé-Pedrós, Peña, et al., 2016). Although this approach is not quantitative on the absolute level, it demonstrated that the relative ratio of macroH2A and PARP1 was the highest in the cystic stage (**Figure 23D**), consistent with our western blot results.

Taken together, gene expression data indicated that metabolism is dynamically regulated between the three life stages of the unicellular filasterean *Capsaspora*. The relative ratio of macroH2A and PARP1 was the highest in the cystic stage. This suggests that the macroH2A-dependent inhibition of nuclear NAD⁺ consumption is most likely to occur in the non-proliferative stage, which coincides with the urgent need to use salvaged and *de novo* synthesized NAD⁺ for life-sustaining catabolic processes.

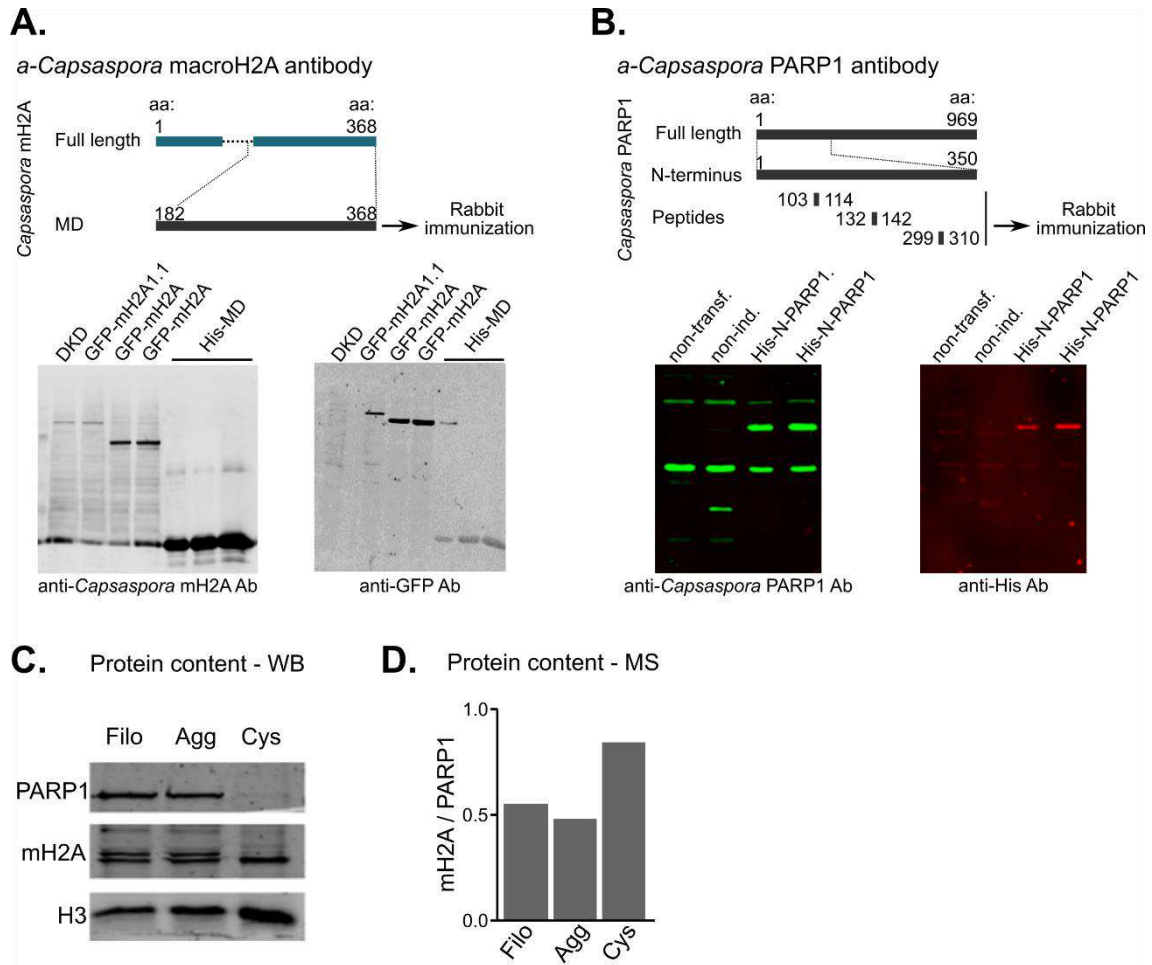


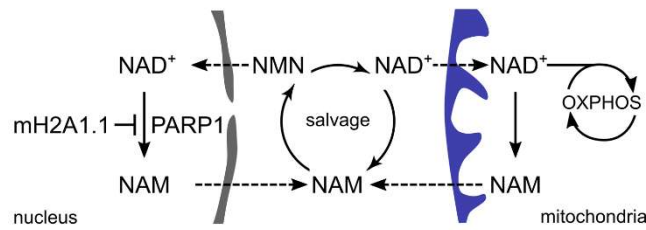
Figure 23. The macroH2A and PARP1 protein expression levels in *Capsaspora*.

(A) Antibody specific for *Capsaspora macroH2A*. Rabbits were immunized with purified His-tagged macroH2A macrodomain (AA: 182-368). The specificity was corroborated by immunoblotting using stable cell lines expressing GFP-tagged constructs of mouse macroH2A1.1 (mH2A1.1) and *Capsaspora macroH2A* (mH2A) and purified His-tagged macrodomain of *Capsaspora mH2A*. **(B)** Antibody specific for *Capsaspora PARP1*. Rabbits were immunized with three different peptides corresponding to the N-terminal part of PARP1. The antibody specificity was corroborated by immunoblotting bacterial lysates: non-transformed (non-transf.), transformed with a plasmid encoding for N-terminal tagged fragment of PARP1 corresponding to amino acids 1-350 but not induced (non-ind.), and bacteria producing His-tagged PARP1 (His-N-PARP1). **(C)** Immunoblot (WB) of total *Capsaspora* cell extracts from the 3 life stages using *Capsaspora* specific antibodies for macroH2A and PARP1 and histone H3 as a loading control. **(D)** Changes in the relative ratio of macroH2A (mH2A) and PARP1 as determined by mass spectrometry (MS). Value in filopodial has been set to 1.

3.2.6 *Capsaspora* macroH2A is a stronger PARP1 inhibitor than mouse macroH2A1.1 and has a more pronounced impact on mitochondrial respiration

The inhibition of PARP1 by macroH2A1.1 in vertebrate cells reduces nuclear consumption of NAD⁺, thereby increasing its availability for the reactions occurring in other organelles, such as the respiration in mitochondria (**Figure 24A**). We wondered how did the evolutionary changes in the macrodomain modulate the metabolic communication between distant organelles. For this, we decided to compare the impact of *Capsaspora* macroH2A and mouse macroH2A1.1 macrodomains by analyzing the two key aspects of this metabolic axis: the capacity to inhibit PARP1 and the ability to increase mitochondrial respiration. The results show that the *Capsaspora* macrodomain was more potent in inhibiting PARP1 activity *in vitro* than the murine macroH2A1.1 macrodomain (**Figure 24B**). To check whether the *Capsaspora* macroH2A macrodomain inhibition of PARP1 depends on its ability to bind ADP-ribose, we used G224E and G314E macrodomain mutants. Indeed, these mutations in the ADP-ribose-binding pocket abolished its ability to inhibit PARP1 (**Figure 24B**). This indicates that the *Capsaspora* macrodomain strongly affects PARP1 activity in an ADP-ribose-binding dependent manner.

A.



B.

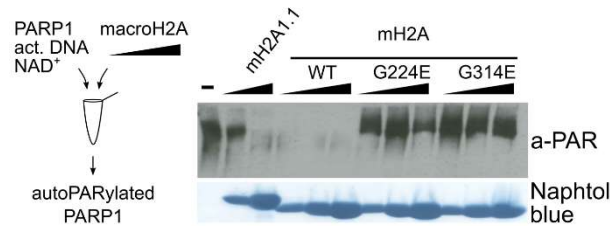


Figure 24. Evaluation of inhibitory activity of macroH2A macrodomains towards PARP1.

(A) Scheme showing how the inhibition of PARP1 by macroH2A1.1 (mH2A1.1) in the nucleus is connected with mitochondrial respiration through NAD⁺ metabolism in vertebrates. **(B)** *In vitro* PARP1 auto-PARylation activity induced by nicked DNA is measured by anti-PAR immunoblotting. Naphtol blue staining shows the increasing amounts of macrodomains that were titrated into the reaction. A representative experiment of three independent replicates is shown.

Next, we sought to determine the impact of ADP-ribose binding by the *Capsaspora* macrodomain on metabolism *in vivo*. Currently, there are no available tools that would allow for the genetic manipulation of *Capsaspora*. Therefore, we used an orthogonal approach and introduced the *Capsaspora* protein into human cells. As cellular model, we chose the well-described HepG2 cells, stably depleted of all macroH2A isoforms (Douet et al., 2017). To focus on the effects of the ADP-ribose binding macrodomain and avoid confounding influences caused by any differences in histone-fold or linker sequences, we fused either the wild-type (WT) or the G224E-mutant macrodomains of *Capsaspora* macroH2A to the histone-fold and linker region of mouse macroH2A1.1 (**Figure 25A**). Based on the detection of an N-terminal GFP-tag, we found that the expression levels of the WT and mutant chimeras was similar in polyclonal stable cell lines and reached approximately half of the level of full-length mouse macroH2A1.1 (**Figure 25B, 25C**). The achieved expression levels were in a similar range or slightly higher than the endogenous levels of macroH2A proteins in parental HepG2 cells (**Figure 25D**). We confirmed that the GFP-

tagged chimeric and murine proteins were fully incorporated into chromatin (Figure 25E) and in contact with PARP1 (Figure 25F).

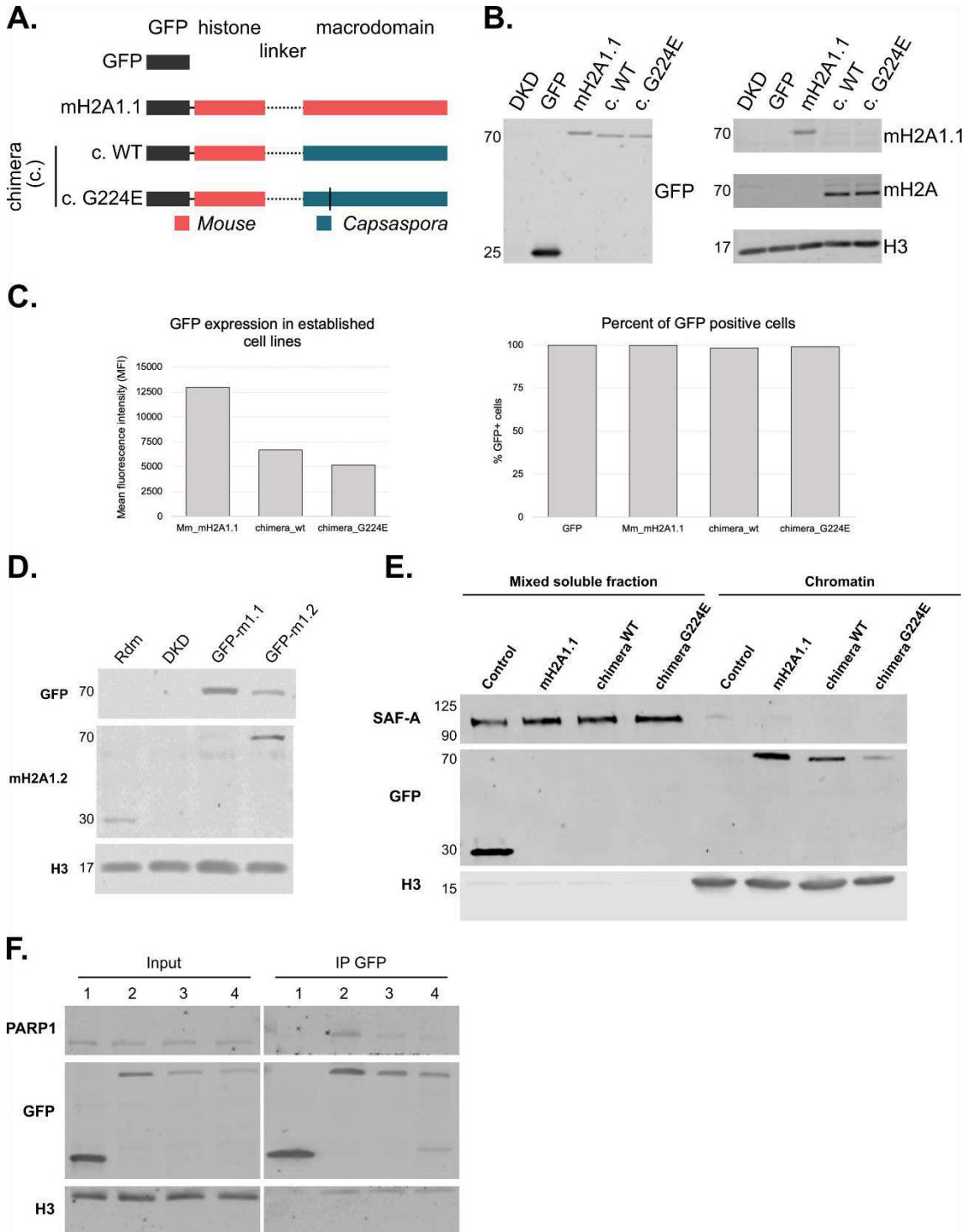


Figure 25. Characterization of orthogonal cell system for the metabolic study of *Capsaspora* macroH2A.

(A) Schematic overview of the constructs that have been introduced into macroH2A-deficient HepG2 cells (referred to as DKD cells). **(B)** The exogenous expression of constructs illustrated in **C** is shown by immunoblotting using anti-macroH2A1.1, anti-*Capsaspora* macroH2A, and anti-GFP. Anti-Histone H3 is included as a loading control. **(C)** Flow cytometric detection of GFP in cell lines from **A**. Bar plot to the left shows GFP fluorescence in cell lines stably expressing macroH2A constructs. The bar plot to the right shows high percentage ($\geq 99\%$) of GFP positive cells. **(D)** Stable cell lines expressing the GFP-tagged mouse mH2A1.1 were compared to a similar cell line expressing GFP-tagged mouse macroH2A1.2 (mH2A1.2), control DKD cells and the parental HepG2 cell line. Immunoblotting using anti-GFP and anti-mH2A1.2 specific antibodies allows evaluating the expression level of the exogenous mH2A1.1 relative to endogenous mH2A1.2. **(E)** Immunoblot analysis after cell fractionation shows that all macroH2A constructs but not GFP alone are incorporated in chromatin. **(F)** The chimeric *Capsaspora*-mouse macroH2A construct interacts with PARP1 as detected by immunoblotting after co-immunoprecipitation.

Having established this orthogonal cell system, we tested for the impact of the distinct macroH2A proteins on cellular metabolism. We found that the expression of key genes representing major metabolic pathways such as glycolysis, fatty acid metabolism and oxidative phosphorylation were largely unaffected by the expression of the different macroH2A constructs (**Figure 26A**). The same was true for *PARP1* and the genes encoding for the components of NAD⁺ salvage pathway. Furthermore, the mitochondrial content was similar between the four cell lines (**Figure 26B**). However, when analysing the oxygen consumption, we found that both the basal and the maximal mitochondrial respiration increased in the presence of mouse macroH2A1.1. Strikingly, despite its lower expression level, this was even more pronounced in the case of the chimeric protein containing the *Capsaspora* wild-type macrodomain (**Figure 26C**). This was also true when we calculated the resulting ATP production (**Figure 26D**). The G224E mutant macrodomain was inert in its impact on metabolism, further substantiating the requirement for a functional, intact ADP-ribose pocket. In sum, our results show that the increased ADP-ribose affinity of *Capsaspora* macroH2A translated into stronger PARP1 inhibitory capacity and a more pronounced impact on the mitochondrial respiration, when compared to mouse macroH2A1.1.

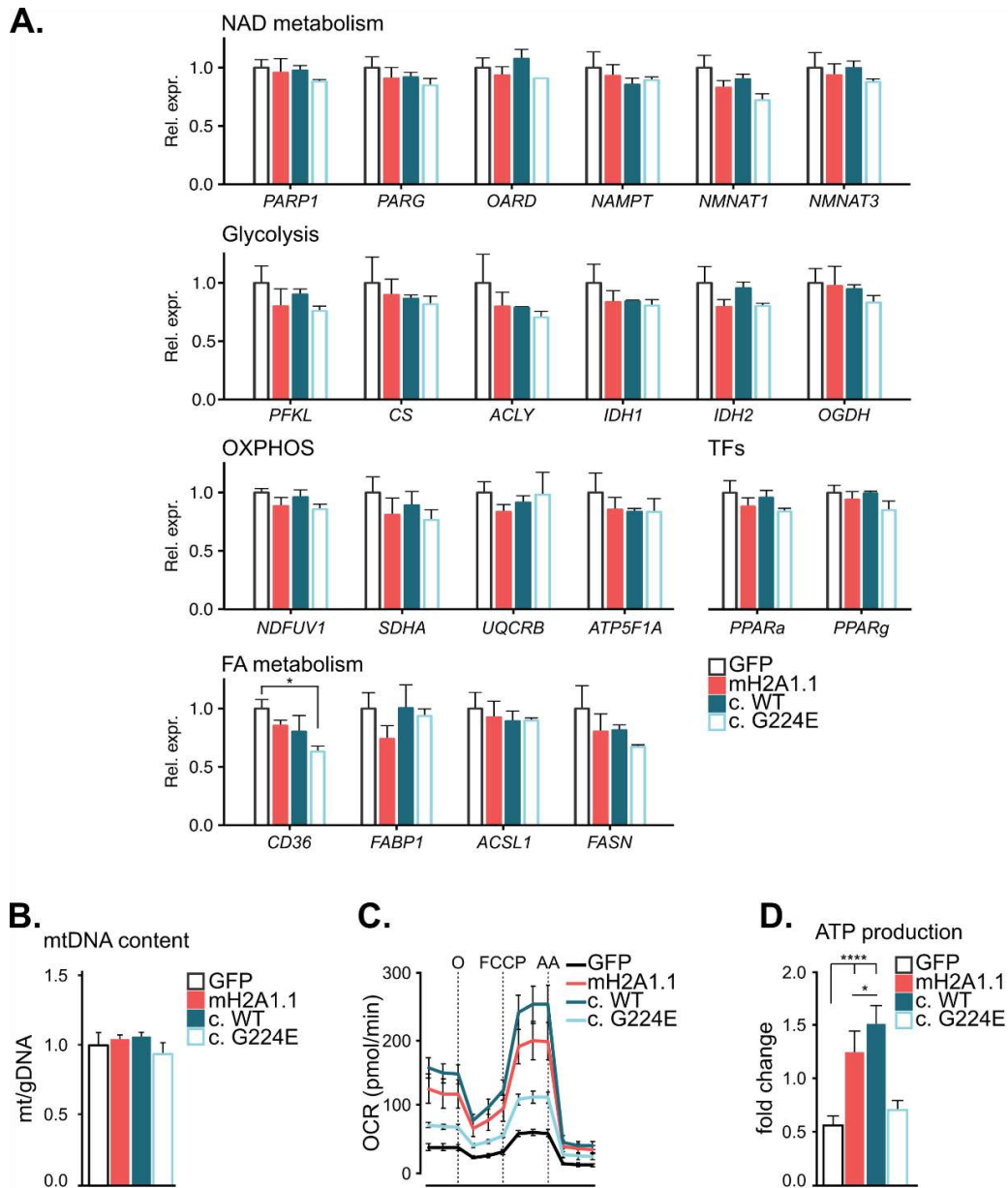


Figure 26. Characterization of the gene expression and metabolic effects of macroH2A macrodomains.

(A) The expression of genes encoding key components and regulators of indicated metabolic pathways was analyzed by RT-qPCR in the cell lines described in Figure 25. The data are calculated based on three independent experiments performed in triplicates. A two-tailed Student's t-test was used to make the indicated pairwise comparisons, * $p < 0.05$. **(B)** Mitochondrial content was assessed by measuring the relative amount of unique sequences in mitochondrial and genomic DNA. Mitochondrial DNA (mtDNA) content normalized to nuclear DNA (gDNA) with MT-ND2 and NDUFV1. A two-tailed Student's t-test to analyze the presence of statistical differences. * $p < 0.05$. **(C)** The oxygen consumption of stable cell lines described in Figure 25 was measured in routine culture and after subsequent addition of the ATPase inhibitor oligomycin (O), the uncoupling compound FCCP and the electron transport chain inhibitors rotenone and antimycin A (AA) (left panel). **(D)** The bar plot shows the resulting ATP production. The data is calculated based on the results of two independent experiments performed in at least three replicates. Ordinary one-way ANOVA was used to make the indicated comparisons between different groups of samples. * $p < 0.05$, **** $p < 0.0001$.

In conclusion, the results of our study suggest that the role of macroH2A in the regulation of cross-organelle NAD communication is ancient. The fusion of a macrodomain to a histone-fold in a common ancestor of breviate, fungi and animals provided protists with a mechanistic tool to dampen nuclear NAD⁺ consumption. During evolution, a reduction in ADP-ribose affinity, mediated by changes in the residues that close the binding pocket upon ligand binding, fine-tuned the stringency of this mechanism. Our results provide important insight on how a compartmental regulatory mechanism of NAD⁺ metabolism evolved.

3.2.7 Contributions

This project was conceived and supervised by Marcus Buschbeck, Sarah Hurtado-Bagès and myself. I took part in designing, performing and/or analyzing of all biophysical and biochemical assays, interpretation of phylogenetics results, performed multiple sequence alignment, performed the analysis of the residue conservation through the evolution, took part in design and validation of *Capsaspora* antibodies, manually annotated and chose the genes involved in *Capsaspora* metabolism and interpreted the obtained results. I also took part in cloning the plasmids for the panel of HepG2 cell lines and their characterization, whereby majority of the characterization was done by Vanesa Valero with the help of Ainhoa Perez. Sarah Hurtado-Bagès initiated the project and started the collaboration with Eirin-Lopez group, took part in interpretation of the phylogenetics results, took part in the design of *Capsaspora* antibodies, performed the western blot in *Capsaspora*, took part in the design and plasmid preparation for the *in cellulo* experiments and performed preliminary *in cellulo* experiments in Hek293T cells. Ciro Rivera-Casas performed all of the phylogenetic analysis together with Jose M. Eirin-Lopez. Gunnar Knobloch taught me and helped with design, analysis and interpretation of the biophysical and biochemical assays, sample preparation for protein crystalization and interpretation of the protein crystalization results in the lab of Andreas Ladurner, whereas Jerome Basuquin solved the crystal structures. Michelle Leger provided us with *Capsaspora* samples and took part in data analysis and interpretation for the *Capsaspora* part, together with Inaki Ruiz-Trillo. Roberto Malinverni the RNAseq analysis of the

data from *Capsaspora*. Jesus Garcia and Marta Frigole taught me how to prepare samples preparation for STD-NMR and JG performed and analyzed STD-NMR experiments in the lab of Xavier Salvatella. Marta Gomez de Cedron performed and analyzed the Seahorse experiment in the lab of Ana Ramirez de Molina. Manjinder S. Cheema and Juan Ausio performed codon optimization and provided us with *Capsaspora* macroH2A cDNA and designed and produced the *Capsaspora* anti-PARP1 antibody.

3.3 Results II: Interdisciplinary approach in the search for the macroH2A1.2 ligand

The phylogenetic analysis of the origin of macroH2A histone variants indicates that macroH2A1.2 was the last of the three macroH2A isoforms to appear. It originated in jawed fish (*Gnathostoma*) and underwent very few substitutions within its macrodomain and exon 5, thus indicating it might have a binding pocket-related physiological role. However, it does not bind ADP-ribose and its ligand remains unknown. The research on macroH2A1.1 histone variant indicates that the choice of the optimal cell line with a relatively high expression of macroH2A isoform of interest is crucial for detecting their binding pocket-related physiological role. However, by now there were no clear indications on the physiological role of macroH2A1.2.

The goal of this project was to detect the ligand that can be accommodated in the binding pocket of macroH2A1.2 macrodomain and to find the physiologically relevant cell line or tissue of interest which would be optimal for studying the physiological role of macroH2A1.2. For this, we chose to use a wide interdisciplinary approach with simultaneous *in silico*, *in vitro*, *in cellulo* and *in vivo* experimental approaches.

3.3.1 Structural characterization of macroH2A1.2 macrodomain

As a result of the alternative splicing of mutually exclusive exons 5, the two macroH2A1 isoforms differ in 30 and 33 amino acid residues within the macrodomain of macroH2A1.1 and macroH2A1.2, respectively (**Figure 27A**). Consequently, macroH2A1.1 binds ADP-ribose, while macroH2A1.2 does not interact with ADP-ribose and its ligand remains unknown. To underline the important characteristics of macroH2A1.2, we performed its detailed characterization on the level of amino acid sequence and macrodomain structure. As the benchmark for comparison, we used the known important features of macroH2A1.1 and several other representative ADP-ribose-binding macrodomains.

3.3.1.1 Comparison with ADP-ribose binding macrodomains

The superposition of the macrodomain structures of macroH2A1.1 and macroH2A1.2 shows a high overall structural homology, in accordance with the 86.9% identity between the two macrodomains. Furthermore, the exon 5 encoded amino acid sequence specifically affects the structure of the macrodomain binding pockets (**Figure 27B**), and as a result affects their ligand-interaction properties. Some of the key macroH2A1.1 ADP-ribose-interacting residues are encoded by the exon 5, such as aspartate 203 (D203) which stabilizes the adenine ring (Karras et al., 2005), glycine 224 (G224) whose mutation to glutamate (G224E) abrogates the binding (Kustatscher et al., 2005), and glutamate 225 (E225) which we here show to be responsible for the decrease in ADP-ribose affinity through evolution. The alignment of the exon 5 encoded sequences of the two isoforms outlines key substitutions which significantly affect the binding pocket (**Figure 27C**). More specifically, macroH2A1.2 has an insertion of three residues (EIS) which correspond to the D203 in the macroH2A1.1 sequence. This is followed by a sequence with a relatively high conservation, while the C-terminal part of the exon 5 sequence of macroH2A1.2 has a highly charged amino acid stretch *in lieu* of ADP-ribose-relevant G224 of macroH2A1.1 (**Figure 27C**). To understand the structural consequences of the outlined substitutions, we compared the structure of exon 5 encoded amino acid stretch of macroH2A1.1 and macroH2A1.2 (**Figure 27D**). The side chain of

isoleucine 204 (I204), a part of the EIS loop, protrudes into the binding pocket and blocks the adenine portion of the binding pocket. Thus, the EIS loop insertion disables the interaction between macroH2A1.2 and ADP-ribose. The ensuing conserved amino acid stretch corresponds to the highly structurally conserved β -sheet 3 (β 3) that is important for the structural integrity of the globular macrodomain structure. In addition to the EIS loop insertion, the second hallmark of macroH2A1.2 which significantly affects the structure and properties of the binding pocket is the highly charged amino acid stretch (DLKDD), referred to as D-rich loop (**Figure 27D**).

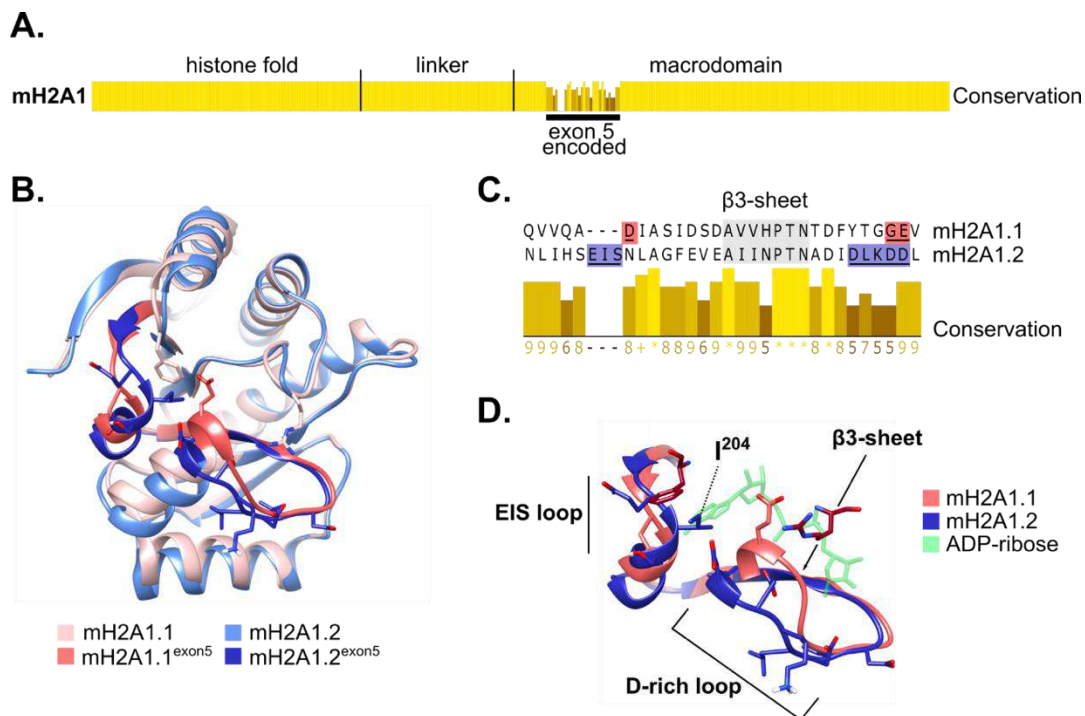


Figure 27. The alternative splicing affects the binding pockets of macroH2A1.1 and macroH2A1.2.

(A) Protein sequence alignment of human macroH2A1.1 (NP_613075.1) and macroH2A1.2 (NP_001035248), represented as a conservation score indicating difference in the macrodomain part of the sequence affected by the alternative splicing of exon 5. The conservation score of amino acid biochemical properties is represented by the height and color of the bars. **(B)** Overlay of the apo-forms of macroH2A1.1 (light coral, PDB 2FXK) and macroH2A1.2 (sky blue, PDB 1ZR5) macrodomains. Exon 5 encoded sequences highlighted in bright coral for macroH2A1.1 and bright blue for macroH2A1.2. **(C)** Protein sequence alignment of amino acids encoded by exon 5 of macroH2A1.1 and macroH2A1.2. The residues implicated in macroH2A1.1 ADP-ribose binding highlighted in coral; the corresponding residues in macroH2A1.2 highlighted in blue. β 3-sheet of proteins highlighted in light grey. The lower part of the panel shows the conservation score of amino acid biochemical properties and is represented by the height and color of the bars on 0-11 scale. **(D)** The overlay of the exon 5-corresponding crystal structure of ADP-ribose (green) bound to macroH2A1.1 (coral, PDB 3IID) and apo macroH2A1.2 (blue, PDB 1ZR5). The β 3-sheet important for structural integrity of both proteins and segments presumed to be relevant for macroH2A1.2 binding pocket function are highlighted: EIS loop with I204, and D-rich loop.

To validate the three observed hallmarks of macroH2A1.1 and macroH2A1.2, increased the number of sequences to analyze. We first compared the exon 5-corresponding amino acid sequences (**Figure 28A**) and structures (**Figure 28B**) of representative ADP-ribose binding macrodomains for which we could find the crystal structures available in the Protein Data Bank (PDB) database, namely macroH2A1.1, Af1521, MacroD1 and MacroD2. As detected upon the comparison of macroH2A1.1 and macroH2A1.2, all of the analyzed macrodomains have a highly conserved sequence coinciding with the structurally conserved β 3-sheet (**Figure 28A, 28B**). Furthermore, the alignment shows that all ADP-ribose interacting macrodomains have conserved residues corresponding to D203 and G224 in macroH2A1.1, as expected. Additionally, Af1521, MacroD1 and MacroD2 macrodomains have a glycine (G225) instead of glutamate (E225) at position 225, which enables the accommodation of the water molecule important for their hydrolytic activity (Jankevicius et al., 2013). Interestingly, one of the apparent hallmarks of all ADP-ribose binding macrodomains seems to be the presence of at least two glycine residues surrounding the position 225, referred to as G-rich loop (**Figure 28A, 28B**). We postulate this loop enables the flexibility and the closure of the binding pocket upon ADP-ribose binding. Thus, by comparing the macroH2A1.1-corresponding exon 5 sequences and structures of ADP-ribose binding macrodomains, we validated the three main hallmarks important for ADP-ribose binding, namely: (I) D203 in adenine accommodating portion of the binding pocket; (II) β 3-sheet important for structural integrity of the macrodomain; and (III) G-rich loop important for binding pocket flexibility and closure.

3.3.1.2 Comparison with non-ADP-ribose binding macrodomains

Majority of the macrodomains discovered and described up to date take part in ADP-ribose binding and metabolism. The only macrodomains which do not bind ADP-ribose, and thus remain orphan proteins up to date, are the macrodomains of macroH2A1.2 (Kustatscher et al., 2005), macroH2A2 (Kozłowski et al., 2018b) and ganglioside-induced differentiation-associated protein 2 (GDAP2) (Rack et al., 2016). To underline the important properties of the non-ADP-ribose binding macrodomains with a goal to take a more informed approach in ligand prediction,

we compared the sequences (**Figure 28C**) and structures (**Figure 28D**) of the macroH2A1.2-corresponding exon 5 between macroH2A1.2, macroH2A2 and GDAP2.

The analysis further consolidated the importance of the conservation of the β 3-sheet sequence which defines the structural axis of the protein and is well conserved across macrodomains independently of the ligand binding (**Figure 28C, 28D**). Similarly to macroH2A1.2, macroH2A2 has a three amino acid insertion at N-terminus of the exon 5-corresponding sequence. The DIS-loop of macroH2A2 is biochemically highly similar to the EIS-loop of macroH2A1.2, thus interfering with ADP-ribose binding in the similar manner. Interestingly, GDAP2 has an aspartate (D) corresponding to D203 of macroH2A1.1 and other ADP-ribose binding macrodomains, although currently available data indicate it does not interact with ADP-ribose (Neuvonen & Ahola, 2009). However, none of the three non-ADP-ribose binding macrodomains have a G-rich loop characteristic of ADP-ribose binding macrodomains. Instead, all three non-ADP-ribose-binding macrodomains have a corresponding charged amino acid loop, termed D-rich loop in macroH2A1.2. The D-rich loop of macroH2A2 is highly biochemically similar to that of macroH2A1.2 (**Figure 28C**). Based on the protein dynamics observed in macroH2A1.1 upon ADP-ribose binding, the macrodomains of macroH2A1 have a flexible loop which can close the upper part of the binding pocket (**Figure 16**). However, the corresponding sequence in macroH2A2 is proline rich (PPFP) and is thus expected to be more rigid than macroH2A1.2 (PSIG). On the other hand, the D-rich corresponding loop of GDAP2 has proline (P) in addition to charged amino acid residues. For this reason, the D-rich loop of GDAP2 is expected to be very rigid and probably in a fixed position, unlike the rest of the described macrodomains (**Figure 28C, 28D**).

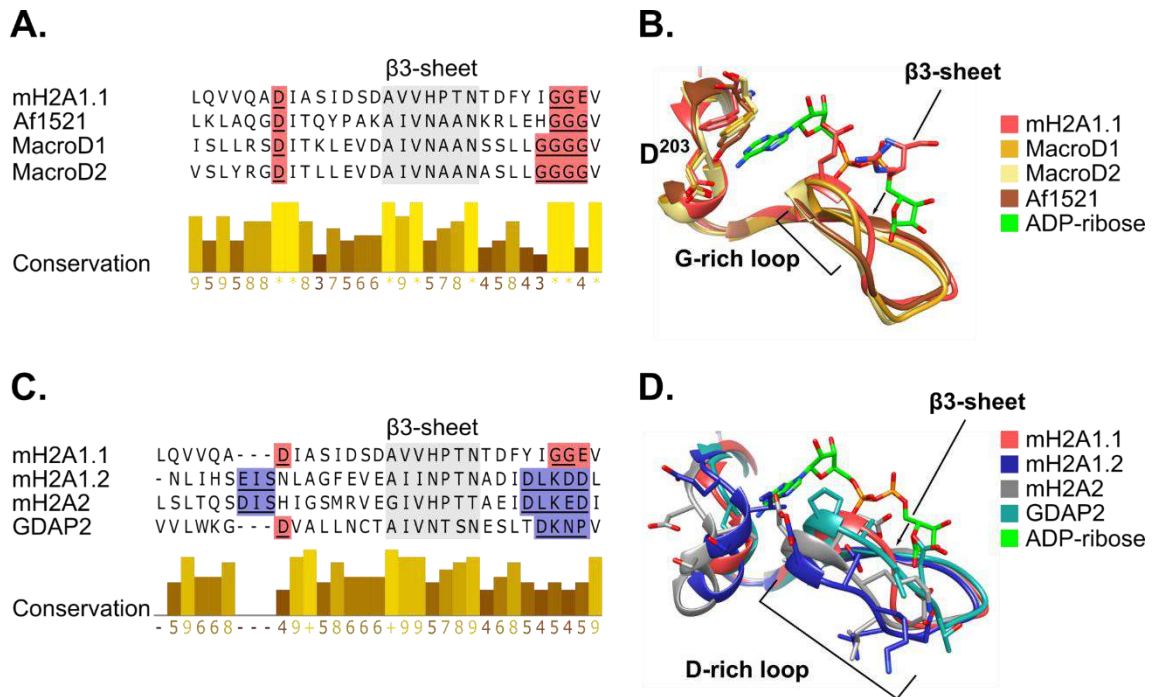


Figure 28. The comparison of macroH2A1.1-analogous exon 5 encoded sequence in different ADP-ribose-binding and non-ADP-ribose binding macrodomains outlines sequence hallmarks.

(A) The protein sequence alignment of exon 5 encoded sequence of human macroH2A1.1 (PDB 3IID) and corresponding sequences of Af1521 (PDB 2BFQ), MacroD1 (PDB 6LH4) and MacroD2 (PDB 4IQY) at the top of the panel. The residues important for ADP-ribose binding are highlighted in coral (D203, G-rich loop). Sequence corresponding to β3-sheet of proteins is highlighted in light grey. The lower part of the panel shows the conservation score of amino acid biochemical properties and is represented by the height and color of the bars on 0-11 scale. **(B)** The overlay of the exon 5-corresponding crystal structure segments of ADP-ribose (green) bound macrodomains outlined in **(A)**. The β3-sheet important for structural integrity of proteins and segments relevant for ADP-ribose binding are highlighted: D203, and G-rich loop. **(C)** The protein sequence alignment of exon 5 encoded sequence of human macroH2A1.1 (PDB 3IID) and corresponding sequences of macroH2A1.2 (PDB 1ZR5), macroH2A2 (PDB 6FY5) and GDAP2 (PDB 4UML) at the top of the panel. The residues important for ADP-ribose binding are highlighted in coral, and the corresponding residues in non-ADP-ribose binding macrodomains are highlighted in blue (D/EIS-loop, D-loop). Sequence corresponding to β3-sheet of proteins is highlighted in light grey. The lower part of the panel shows the conservation score of amino acid biochemical properties and is represented by the height and color of the bars on 0-11 scale. **(D)** The overlay of the exon 5-corresponding crystal structure segments of ADP-ribose (green) bound macrodomain of macroH2A1.1 and non-ADP-ribose binding macrodomains outlined in **(C)**. The β3-sheet important for structural integrity of proteins and macroH2A1.2 D-rich loop are highlighted.

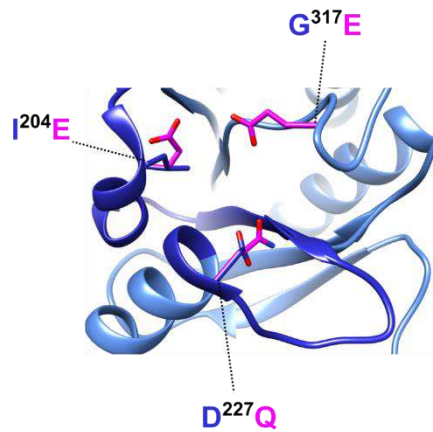
3.3.1.3 The design of the macroH2A1.2 binding pocket mutants

Based on the characterization and comparison of the ADP-ribose-binding and the non-ADP-ribose-binding macrodomains, we postulate that the EIS loop insertion and the D-rich loop are the key portions of the macroH2A1.2 specific binding pocket important for ligand binding. Furthermore, we speculate that the glycine 317 (G317) might be important in macroH2A1.2 ligand interaction, since the corresponding G315 has an important implication in the binding pocket flexibility and stabilization of ADP-ribose in macroH2A1.1. Based on these findings, we designed the expected loss-of-binding mutants of macroH2A1.2 binding pocket which we could later use for studying the physiological role of macroH2A1.2 binding pocket (**Figure 29A**):

- I204E: hydrophobic to negatively charged residue; mutant affecting the EIS loop; expected to abrogate the binding if the ligand protrudes within the binding pocket to interact with the nonpolar I204 side chain;
- D227Q: negatively charged to polar residue resulting in a loss of charge in the D-rich loop and thus abrogating any salt bridges stabilizing the ligand;
- G317E: small nonpolar to large negatively charged residue; expected to abrogate the interaction due to difference in charge, size and flexibility of the loop.

We checked the stability of single, double and triple mutants of macroH2A1.2 binding-pocket by thermal shift assay, which indicates the protein melting temperature and can be used as a proxy for protein folding. Results indicated the melting temperature of single and double mutants are very similar to that of wild type (less than 1°C difference), with the exception of D227Q_G317E (~1.4°C lower). The triple binding pocket mutant had a melting temperature ~2°C lower than the wild type, thus indicating lower stability of this mutant (**Figure 29B**).

A. Binding pocket mutant mH2A1.2



B. T_m WT vs mutant mH2A1.2

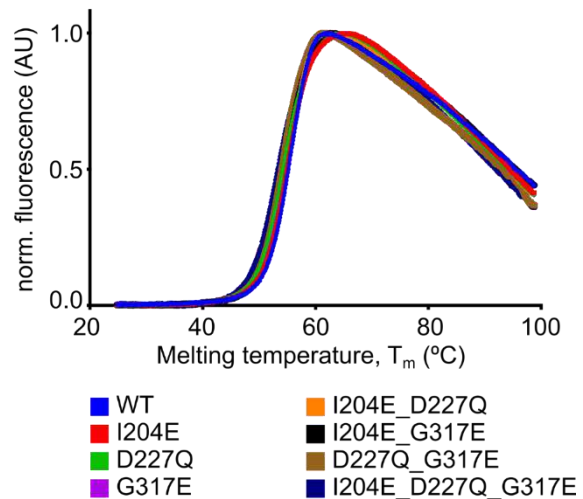


Figure 29. The design and analysis of the macroH2A1.2-binding pocket mutants.

(A) The design of the binding pocket mutants of macroH2A1.2. *In silico* mutagenesis of macroH2A1.2 (PDB 1ZR5) was performed using UCSF Chimera. The residues in the wild type binding pocket are represented in blue (I204, D227, G317), and the side chains of mutated residues are represented in magenta (E204, Q227, E317). **(B)** The thermal shift assay was performed using GST-tagged macrodomains of the wild type (WT) and a panel of single (I204E, D227Q, G317E), double (I204E_D227Q, I204E_G317E, D227Q_G317E) and triple (I204E_D227Q_G317E) mutant macroH2A1.2 macrodomains. The results of one assay performed in triplicates were analyzed and represented using GraphPad Prism.

Taken together, after the extensive sequence and structure characterization of both the ADP-ribose binding and the non-binding macrodomains, we found the possible hallmarks of the non-ADP-ribose binding macrodomains. Thus, we postulate that the EIS-loop and the D-rich loop are important features of macroH2A1.2 for ligand binding.

3.3.2 The targeted approach in the search for the ligand of macroH2A1.2

3.3.2.1 Molecular docking predicts a list of putative ligands

Following the analysis of macroH2A1.2 sequence and structure in the search for the hallmarks of the binding pocket, we used the molecular docking for *in silico* prediction of the molecules with the highest probability to interact with macroH2A1.2 binding pocket.

In the search for the physiologically relevant ligand we used the Human Metabolome Database (HMDB) and Kegg Database as a source of candidate compounds, namely small molecule metabolites found in human body and ecosystem, respectively. As a bait, we used macroH2A1.2 macrodomain crystal structure (PDB 1ZR5). The output of the molecular docking is a top hit list of the small molecules which are ranked by their probability to interact with the macroH2A1.2 binding pocket (**Figure 30A**). From the extracted list of 500 best scored molecules based on the docking calculations, 9 molecules were ranked to have a very good probability for the interaction with macroH2A1.2 binding pocket (**Figure 30B**). The candidate molecules had diverse chemical structures and are expected to be present in human body as drug derivatives or metabolites (compounds 1-4), as naturally occurring metabolites in mammals (compounds 5-7), or as derivatives of plant metabolites (compounds 8 and 9) (**Figure 30B**). Due to their intricate and complex stereochemistry, natural compounds still present a great challenge for synthetic chemistry (Paterson & Lam, 2018), and consequently only a very few natural compounds are available on the market. Thus, from our top hit list of compounds, we were able to obtain only three out of nine small molecules to test them experimentally, namely compounds 2, 4 and 6.

3.3.2.2 The binding validation of the top predicted ligands from molecular docking

To test if any of the predicted top hit molecules interacts with macroH2A1.2 macrodomain, we used STD-NMR (**Figure 11**). As expected based on the chemical structure of the molecules, the solubility of the compounds was relatively poor in the polar solvent. However, we did not observe any precipitation of the compounds 2 and 4 in the presence of 1% DMSO, while the compound 6 could not be solubilized. Additionally, 1% DMSO in the solvent did not affect the protein folding (results not shown). For method validation we used ADP-ribose, and confirmed the absence of interaction between macroH2A1.2 and ADP-ribose, while confirming the interaction between macroH2A1.1 and ADP-ribose (**Figure 30C, left panel**). From the two compounds with adequate solubility, upon the irradiation of the compound 4 in the absence of protein and while specifically targeting the protein signal, the STD profile presented with peaks characteristic of artefacts, which persisted in the sample containing macroH2A1.2 (**Figure 30C, right panel**). Based on this result and absence of change upon the addition of macroH2A1.2, the compound 4 is not expected to interact with macroH2A1.2 macrodomain. The artefacts in the STD spectrum of compound 2 were much smaller and the observed peaks were similar to those obtained in the presence of macroH2A1.1, thus indicating the absence of interaction between macroH2A1.1 and compound 2. On the other hand, the STD signal in the presence of macroH2A1.2 showed significant signal saturation transfer from macroH2A1.2 to compound 2, resulting in the >1% peaks, the threshold we set to define the presence of interaction (**Figure 30C, middle panel**). Thus, the compound 2, namely 3-hydroxypivacaine (3-OHR), was shown by STD-NMR to specifically interact with the macrodomain of macroH2A1.2 and not with macroH2A1.1.

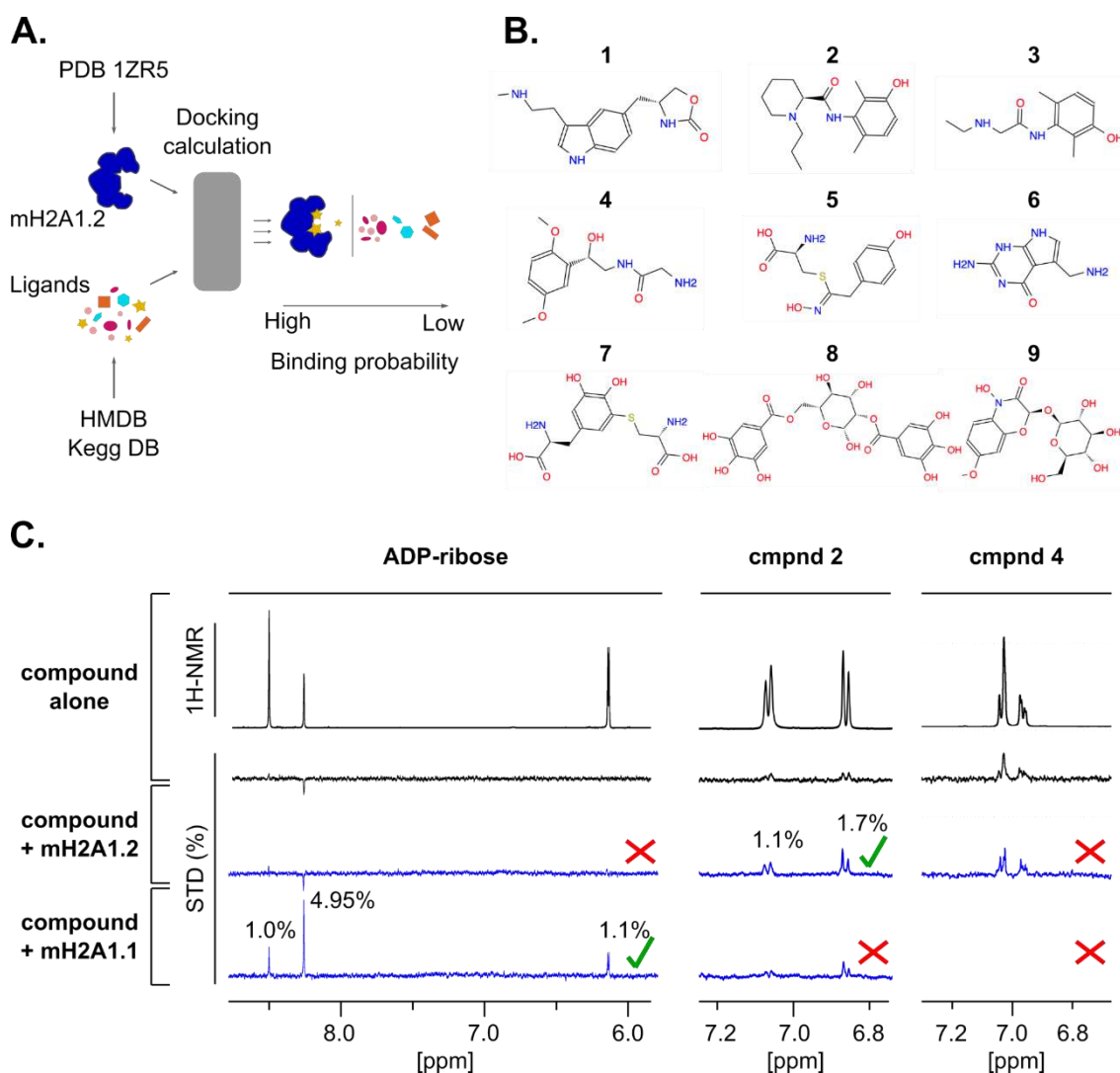


Figure 30. Targeted in silico and in vitro approach for the detection of macroH2A1.2 ligand. (A) Molecular docking schematic indicating the source of macroH2A1.2 structure (PDB 1ZR5) and ligands (HMDB and Kegg) for the analysis. (B) The structures of 9 top hit compounds from the molecular docking which are rated to have the high probability to interact with the macrodomain of macroH2A1.2 in a binding pocket-related manner. The exact identities of the compounds are listed in **Annex 8.3**. (C) The top spectra show 1H-NMR and STD-NMR spectra in black of the tested compounds alone, namely ADP-ribose, and compounds 2 and 4 from **B**. The STD-NMR spectra of the compounds in the presence of macroH2A1.2 (top blue spectra) or macroH2A1.1 (bottom blue spectra) macrodomains are shown for ADP-ribose, compound 2 and compound 4 (in the presence of macroH2A1.2 only). The saturation transfer difference percentage is written for all peaks equal or above 1%. The red cross mark indicates the absence of binding and the green tick mark indicate the presence of binding.

As a parallel binding assay for validation of the interaction between 3-OHR and macroH2A1.2, we used isothermal titration calorimetry (ITC). One of the prerequisites for successful ITC measurements is to have matched buffer between ligand and protein. Thus, due to the presence of 1% of DMSO necessary for the solubilization of 3-OHR, we added 1% DMSO to the protein containing buffer. We checked for any potential confounding effects of DMSO addition, we first validated the interaction between ADP-ribose and macroH2A1.1 in the presence of DMSO. The obtained ITC results confirmed the binding between macroH2A1.1 macrodomain and ADP-ribose in the presence of DMSO, however with an increased heat of dilution (**Figure 31A, left panel**). However, we were not able to detect the interaction between macroH2A1.2 and 3-OHR using this method. This was a consequence of very low enthalpic change (ΔH), resulting in ΔH of -1.49 kcal/mol measured for macroH2A1.2 with 3-OHR, compared to -12.0 kcal/mol for macroH2A1.1 and ADP-ribose. These small heat changes were further masked by high levels of background, i.e. heat of dilution (**Figure 31A, right panel**). Taken together, the results indicate ITC is not a suitable method for characterizing macroH2A1.2 – 3-OHR interaction, suggesting other biophysical methods, which are less sensitive to buffer mismatches, should be considered.

Furthermore, to try obtain more detailed information on the 3-OHR binding mode we set about to try and crystallize macroH2A1.2 macrodomain in the presence of 3-OHR. Interestingly, we obtained what seemed to be protein crystals in several conditions (**Figure 31B, 31C**). However, although we obtained some big crystals (**Figure 31B**) indicating that the binding between macroH2A1.2 and 3-OHR is stable enough to form ordered structures prone to crystallization, the crystals dissolved before we were able to pick them. In various other conditions, we obtained small crystals which were most probably salt crystals (**Figure 31C**).

Due to the scarcity of the natural compounds on the market, we were able to test only three top candidates from molecular docking using STD-NMR. The compound 3-OHR specifically bound to macroH2A1.2, and not to macroH2A1.1, as detected by STD-NMR, indicating that macroH2A1.2 macrodomain might indeed accommodate a ligand in a macrodomain specific manner. ITC as a parallel method for detection of binding was not optimal and thus no binding was

observed, but the crystal appearance adds to the robustness of the results obtained by STD-NMR. However, the binding affinity seems to be very low, implying that 3-OHR has some relevant chemical groups in its structure, but is not optimal and thus not the true physiological ligand.

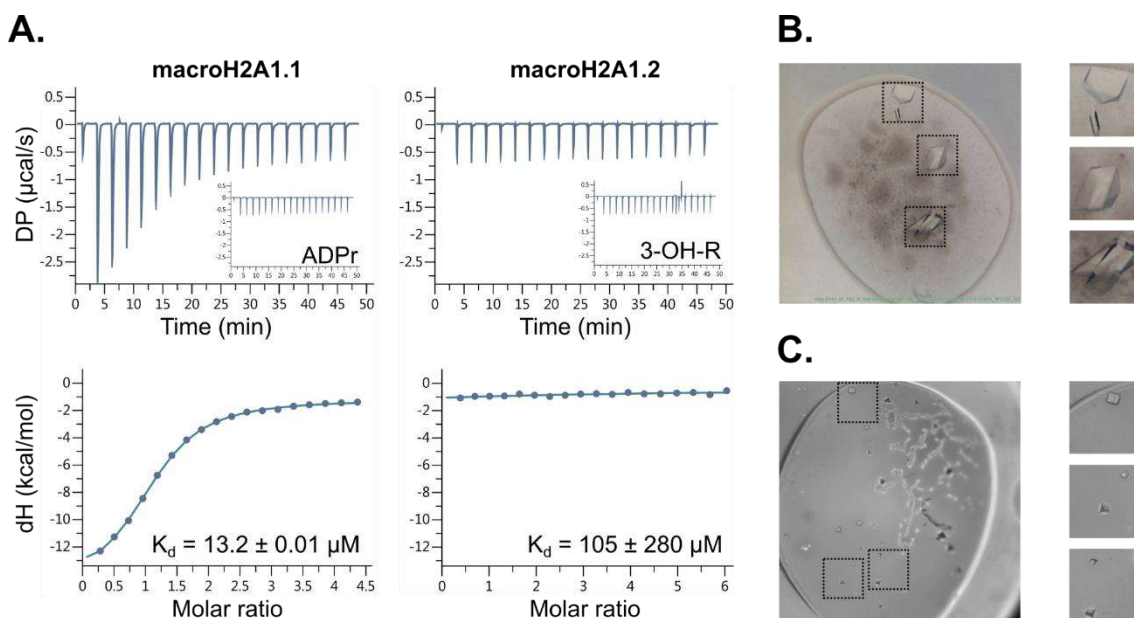


Figure 31. Isothermal titration calorimetry (ITC) and crystallization assay as parallel binding assays.

(A) The results of the ITC binding measurements between macroH2A1.1 and ADP-ribose (left panel) and between macroH2A1.2 and 3-OH-ropivacaine (3-OHR, right panel) in the presence of 1% DMSO. The top panels show heat development and the bottom panel shows the integrated heat plot with the calculated K_d values as indicated. The inset in the top panel shows the heat development of the compound titration to the buffer, without the presence of the proteins. **(B)** The light microscope image showing the protein-corresponding crystal formation after three weeks of macroH2A1.2 macrodomain and 3-OHR incubation. **(C)** The light microscope image showing the salt-corresponding crystal formation after six weeks of macroH2A1.2 macrodomain and 3-OHR incubation.

3.3.2.3 The similarity search to refine the list of potential ligands

We used the obtained experimental data on the low affinity interaction between macroH2A1.2 and 3-OHR, to perform a refined similarity search using ZINC15 database as a source of 120 million synthetic molecules. The resulting top 11 compounds had one of two specific structural motifs: (I) a nitrogen (N)-containing, non-aromatic heterocycle linked via an amide linker to an aromatic ring, or (II) a heterocycle linked via an aliphatic linker to a terminal amide (**Figure 32**).

The N-containing heterocyclic ring was predicted to be protonated and positively charged. Consequently, the docking model predicted it to be positioned in the negative region spanned by the aspartic acids, termed D-rich loop (**Figure 27D**). The amide linkers are docked to form bridging interactions with D227 and the backbone of R318. Other chemical functions attached to the aromatic part found to be forming interactions with R318 and N319.

The compounds were checked for interaction with macroH2A1.2 macrodomain using microscale thermophoresis (MST). The experiment was designed to give the information on the presence or absence of the interaction between the compounds and macroH2A1.2 macrodomain, based on the presence or the absence of the signal (results not shown). The measurements confirmed the interaction of macroH2A1.2 with 3-OHR (**compound 2**) and two additional compounds with an amide linker. The compounds were further ranked based on the descending probability of interaction with macroH2A1.2 from probable (compounds 2.1–2.5), possible (compounds 2.6 and 2.7) and no interaction (compounds 2.8–2.11) (**Figure 32**).

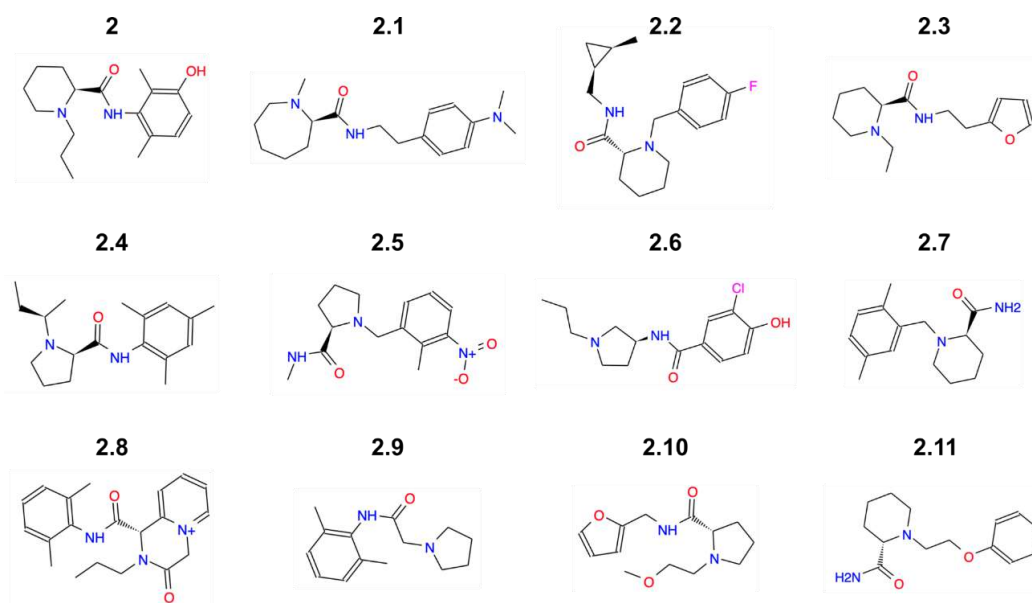


Figure 32. The chemical structures of 11 compounds similar to compound 2 from the *in silico* similarity search.

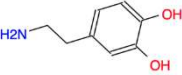
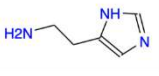
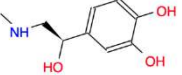
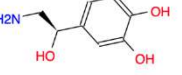
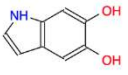
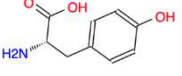
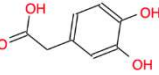
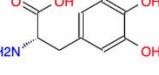
The compounds are shown in the order of the decreasing probability of the interaction with macroH2A1.2 macrodomain as measured by MST. The compound 2 corresponds to the 3-OHR shown to interact with macroH2A1.2 in **Figure 30**. The compound identities for compounds 2.1–2.11 are shown in **Annex 8.3**.

3.3.2.4 The catecholamine neurotransmitters as candidate ligands

The previous observations on the mice with a knock-out of both macroH2A1 isoform (macroH2A1 KO mice) indicated they are resistant to anesthesia. More specifically, it took significantly longer for the macroH2A1 KO than the WT mice to go under anesthesia upon an intraperitoneal administration of a mix of ketamine (10mg/mL) and medetomidine (0.2mg/mL) (results not published). Intriguingly, one of the prospective hits from the molecular docking was the catecholamine metabolite, 5-S-Cysteiny-DOPA (**Figure 30B**, compound 7). Unfortunately, we were unable to experimentally validate 5-S-Cysteiny-DOPA since it is not available on the market and attempts to synthesize it failed due to its complicated stereochemistry. However, the simple catecholamines are widely available. Thus, we decided to test a panel of catecholamine neurotransmitters and their metabolites for the interaction with macroH2A1.2 macrodomain using MST.

Catecholamines are tyrosine-derived monoamine neurotransmitters, characterized by the catechol ring which is coupled to an amine side chain (**Table 1**) (Nagatsu, 2006). Surprisingly, the MST measurements indicated all of the tested catecholamines, namely dopamine, histamine, epinephrine and norepinephrine, interact with macroH2A1.2 macrodomain with increasing affinity. The interaction affinity was significantly affected by the pH of the buffer especially in the case of histamine, thus outlining the protonation state of the molecules as one of the important prerequisites for the interaction with macroH2A1.2. The highest macroH2A1.2 affinity was observed for epinephrine or norepinephrine, with K_d values of 140 and 213 μ M, respectively. None of the catecholamine precursors or intermediates interacted with macroH2A1.2 macrodomain (**Table 1**).

Table 1. The characterization of macroH2A1.2 – catecholamine interaction using MST.

| Compound name | Structure | K_d (μM) | pH |
|---------------------|---|-------------------------------------|------------|
| Dopamine |  | 792.3 ± 201.9 | 6.4 |
| Histamine |  | 46.3 ± 37.9 509.3 ± 48.3 | 6.4 7.4 |
| Epinephrine |  | 213.7 ± 61.3 | 7.4 |
| Norepinephrine |  | 140.3 ± 21.7 | 7.4 |
| 5,6-dihydroxyindole |  | ND | 7.4 |
| Tyrosine |  | ND | 7.4 |
| DOPAC |  | ND | 7.4 |
| Levodopa |  | ND | 7.4 |

3.3.3 The untargeted approach in the search for the ligand of macroH2A1.2

3.3.3.1 Experimental design to overcome the challenges of available methods for protein deorphanization

Deorphanization of domains is a challenging task. Methods described for the deorphanization of enzyme domains rely on measuring enzymatic activity, which is not possible with the proteins with an unknown or no enzymatic activity (Orsak et al., 2012). The majority of approaches directly assessing binding are usually limited to a predefined subset of ligand candidates. In order to make binding assays compatible with high-throughput screening, they usually require the labeling of the protein or a competitor ligand and thus carry the risk of interfering with the interaction of interest. Screens in living cells require a physiological readout such as a reporter gene expression and are not feasible less characterized proteins without tangible cell biological effects (Vuignier et al., 2010). Finally, a prerequisite for the direct identification of a ligand by mass spectrometry measurements is a strong interaction persisting through washing steps that allow removing unspecifically bound molecules (M. V. Chakravarthy et al., 2009).

Previous attempts of our group failed to detect the binding of ADP-ribose to macroH2A1.1 being a known ligand. More specifically, the GFP-macroH2A1.1 was overexpressed in cells and pulled-down, and followed by the elution of the unspecifically bound molecules prior to obtaining the metabolite extract. However, no ADP-ribose could be detected. These results imply the interaction between macroH2A1.1 and ADP-ribose is not stable enough for it to remain bound during washing steps (results not shown). Thus, since we expected the affinity between macroH2A1.2 and its ligand to be in a similar range as macroH2A1.1 and ADP-ribose, we needed to apply a different methodology which would not involve the washing steps which could potentially elute the ligand of interest.

With that in mind, we adapted the approach reported in Karras et al. The authors incubated various macrodomains with ADP-ribose and subsequently measured

the absorbance of the remaining ADP-ribose in the supernatant. This enabled the distinction between the proteins which were capable to bind ADP-ribose from those which were not without any washing steps.

For a proof-of-concept experiment, we immobilized the GST, GST-tagged macroH2A1.1 or macroH2A1.2 macrodomains to beads and incubated them with ADP-ribose. We then measured ADP-ribose absorbance in the remaining supernatant, or the ADP-ribose released from the proteins upon heat-denaturation. We refer to the first approach as the depletion experiment since we measured the depletion of ADP-ribose from the supernatant upon incubation with the proteins. The second approach is referred to as the enrichment experiment, since we measured the absorbance of ADP-ribose that was bound to the proteins during the incubation and without washing released to the supernatant after protein denaturation (**Figure 33A**). The results showed a significant depletion of ADP-ribose upon the incubation with GST-macroH2A1.1 macrodomain, but not with GST or GST-macroH2A1.2 (**Figure 33B, left panel**). The results of the enrichment experiment showed an even more striking difference, whereby the ADP-ribose signal was almost exclusively coming from the sample obtained after the denaturation of GST-macroH2A1.1 (**Figure 33B, right panel**). Thus, these results indicate that it is possible to detect the interaction between macroH2A1.1 and ADP-ribose by performing a pull-down experiment without washing steps. Furthermore, the approach allowed for the specific detection of the ADP-ribose interaction with GST-macroH2A1.1 and not with GST or GST-macroH2A1.2 despite any unspecific binding or dilution effects.

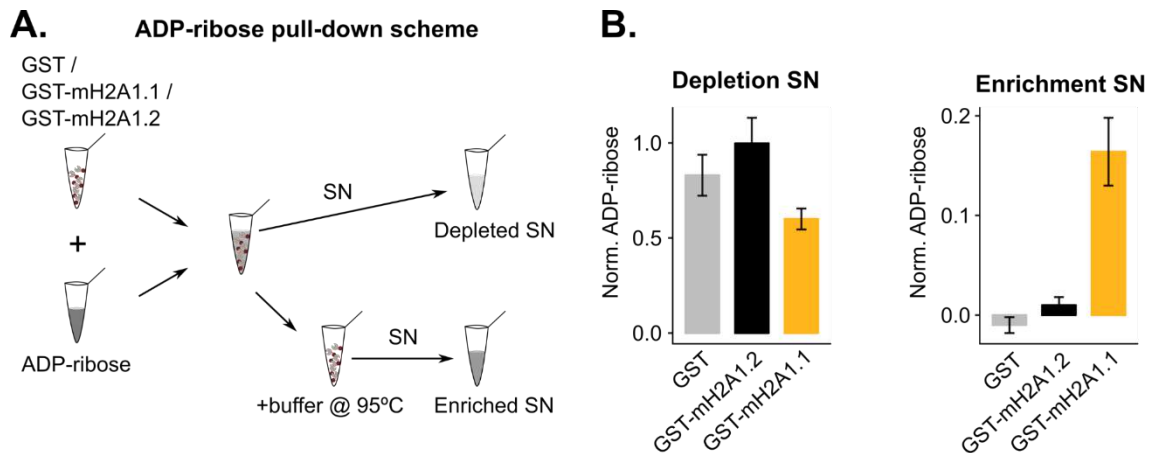


Figure 33. The design and results of proof-of-concept ADP-ribose pull-down assay without elution steps.

(A) The schematic of the proof-of-concept ADP-ribose pull-down assay. GST, GST-macroH2A1.1 and GST-macroH2A1.2 macrodomains were immobilized and incubated with ADP-ribose solution. The supernatant was used for the analysis of ADP-ribose depletion, and the proteins were heat-denatured with the addition of buffer, following which the remaining supernatant was used for the analysis of ADP-ribose enrichment absorbance measurements. **(B)** The left panel shows the ADP-ribose absorbance measurements at 260nm. The mean absorbance measurements from the depletion supernatants of GST and GST-macroH2A1.2 were normalized to 1. The right panel shows the absorbance measurement of the enrichment supernatants in which the mean measurements for the enrichment supernatants of GST and GST-macroH2A1.2 were normalized to 0. The graphs represent measurements for two independent experiments measured in quadruplicates.

3.3.3.2 Untargeted metabolomics for identification of macroH2A1.2 ligand

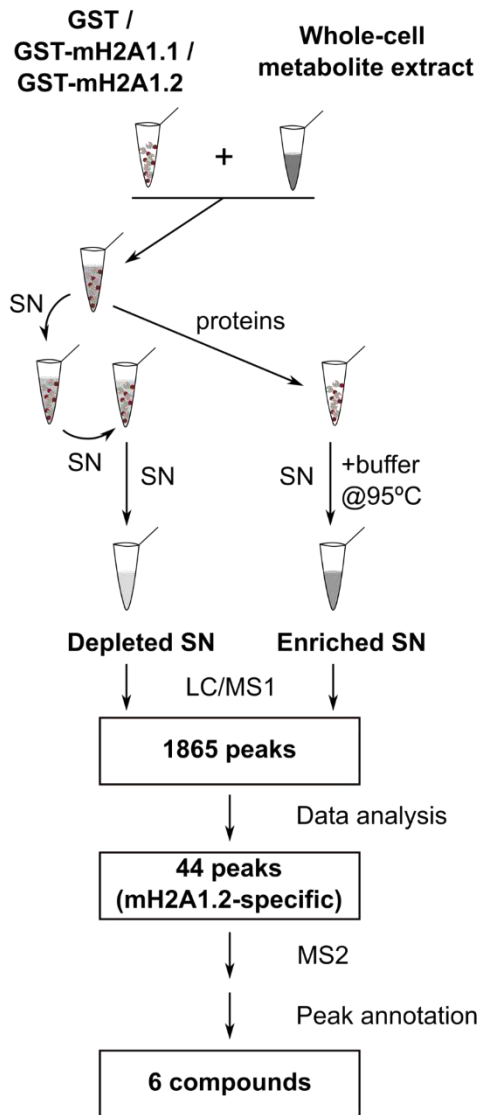
As a source of metabolites in the search for the ligand of macroH2A1.2, we used a whole-cell metabolite extract. In order to avoid any loss of metabolites, we performed metabolite extraction using two different solvents with increasing polarity. Furthermore, to ensure for an efficient depletion of the ligand(s) of interest, the depletion assay was performed in a sequential manner by transferring the depleted metabolite extract to a tube containing fresh proteins, thus enabling efficient depletion of the metabolite of interest. On the other hand, the enrichment assay was performed with the proteins after the first incubation with metabolite extract, which we presumed has the most efficient enrichment of the ligand of interest (**Figure 34A**). The obtained depletion and enrichment supernatants were analyzed by untargeted LC/MS in which the metabolites were separated using two columns with increasing polarity. We detected the total of 1865 peaks representing small molecules interacting with either of the three proteins. To detect the peaks corresponding to molecules specifically interacting with the macrodomain of macroH2A1.2, we filtered for the peaks enriched in macroH2A1.2 in comparison to GST. We proceeded by setting a threshold to exclude any peaks which were similarly enriched in both macroH2A1.1 and macroH2A1.2 samples. As a result of the analysis, we obtained 44 peaks corresponding to either specifically enriched or specifically depleted molecules by macroH2A1.2 macrodomain. Finally, we performed MS/MS fragmentation of 44 macroH2A1.2-specific peaks to detect the exact molecular mass, annotate the peaks and in this way identify the metabolites (**Figure 34A**).

The peaks could be annotated if the corresponding mass and fragmentation spectra coincided with the fragmentation pattern of the chemical standards available in databases. However, databases are not comprehensive and due to a high diversity of metabolites, not all metabolites can be annotated. With some further manual curation, we were able to successfully annotate peaks corresponding to 6 compounds. Four of these were identified as highly saturated polymer chains, presumably the contaminants from the columns and other plastic materials used in the experiment. The two compounds that we identified as natural metabolites were the phospholipids phosphatidylethanolamine [PE(35:1)]

and phosphatidylcholine [PC(34:2)] (**Figure 34B**). Interestingly, additional three peaks from those which could not be annotated had a similar profile as identified phospholipids, indicating there could be more phospholipids interacting with macroH2A1.2 (data not shown). To find which phospholipids correspond to the identified molecules, we searched the Lipid Maps Database and found 30 different phosphatidylcholines with the mass and chemical formula corresponding to PC(34:2) and 16 different phosphatidylethanolamines corresponding to PE(35:1) (**Figure 34C**). On the basis of the obtained results we concluded that macroH2A1.2 macrodomain interact with phospholipids with choline or ethanolamine as polar heads while the exact identity of the fatty acid chains remains unknown (**Figure 34D**).

We hypothesized that the interaction between macroH2A1.2 and phospholipids happens via a phospholipid polar head, otherwise we would expect to see more lipids among the detected metabolites. We tested this hypothesis by performing a molecular docking of the polar heads of phosphatidylcholine (PC) and phosphatidylethanolamine (PE) and macroH2A1.2. Interestingly, the phospholipid polar heads were modelled to be binding within the canonical binding pocket of macroH2A1.2, and predicted to be stabilized either by side chains of D224 and R318 in the case of ethanolamine of the PE (**Figure 34E**), or by D227 side chain and R318 backbone in the case of choline of the PC (**Figure 34F**).

A.



B.

Identification of annotated compounds

| ID | Formula | Origin |
|------------|---|-----------|
| 112-27-6 | C ₆ H ₁₄ O ₄ | Synthetic |
| 4792-15-8 | C ₁₀ H ₂₂ O ₆ | Synthetic |
| 126-71-6 | C ₁₂ H ₂₇ O ₄ P | Synthetic |
| 94159-75-8 | C ₃₈ H ₇₈ O ₁₃ | Synthetic |
| PE(35:1) | C ₄₀ H ₇₈ NO ₈ P | Natural |
| PC(34:2) | C ₄₂ H ₈₀ NO ₈ P | Natural |

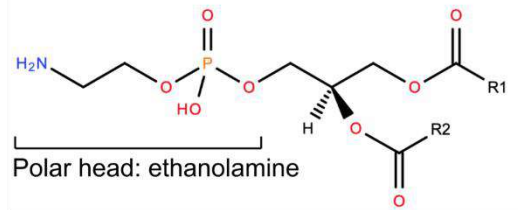
C.

Identification of mH2A1.2-interacting phospholipids using Lipid Maps Database

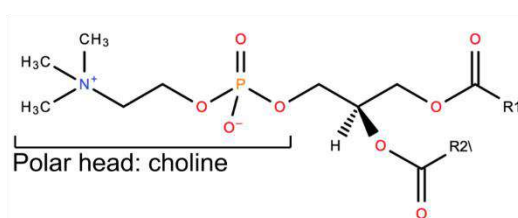
| Compound ID | Formula | Mass | Nr of hits |
|-------------|---|----------|------------|
| PC(34:2) | C ₄₂ H ₈₀ NO ₈ P | 757.5622 | 30 |
| PE(35:1) | C ₄₀ H ₇₈ NO ₈ P | 731.5465 | 16 |

D.

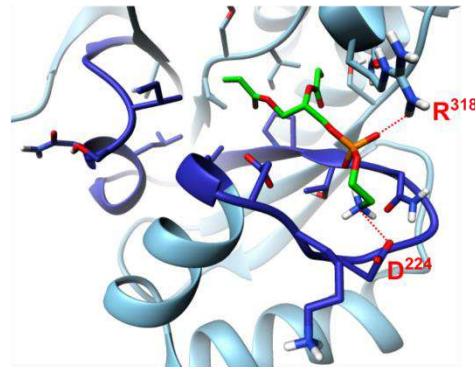
Phosphatidylethanolamine



Phosphatidylcholine



E. Molecular docking: ethanolamine



F. Molecular docking: choline

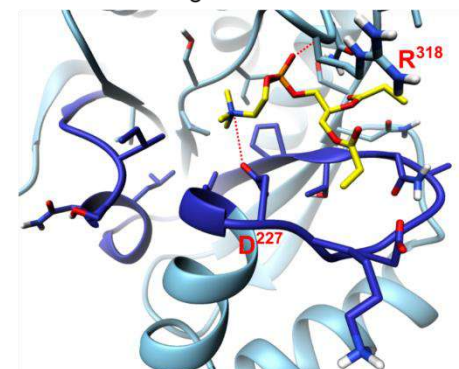


Figure 34. Untargeted metabolomics for the detection of macroH2A1.2 specific ligands from the whole-cell metabolite extract.

(A) The schematic outline of the experiment design and analysis. GST, GST-macroH2A1.1 and GST-macroH2A1.2 macrodomains were immobilized and incubated with the metabolite extract from P19 cells. The beads with the proteins after the first incubation were denatured at 95°C with the addition of buffer and the leftover supernatant was used for the metabolite enrichment analysis. The supernatant from the first incubation step was transferred to the tube with fresh proteins and incubated in the second depletion step. This was repeated one more time to obtain the supernatant sequentially depleted of the metabolites for the depletion analysis. The supernatants were analyzed by LC/MS1 in which 1865 peaks corresponding to small molecules were detected. The analysis outlined 44 peaks specific to macroH2A1.2 interaction which underwent fragmentation (MS2) followed by peak annotation to detect 6 small molecules interacting specifically with macroH2A1.2 macrodomain. **(B)** Table outlining the identities and chemical formulas of the compounds detected as described in **(A)** and whether they are expected to be of synthetic or natural origin. **(C)** Results of the Lipid Maps Database search the compounds corresponding to the only natural compounds detected by **(A)**, as shown in **(B)**, namely PE(35:1) and PC(34:2). The table shows the compound identity, formula, molecular mass and the number of molecules in the database which have the corresponding chemical formula and molecular mass. **(D)** The general structure of phosphatidylethanolamine (PE, top panel) and phosphatidylcholine (PC, bottom panel), outlining their differing polar heads, ethanolamine and choline, respectively. **(E)** Molecular docking results of PE polar head, ethanolamine (green), to the macrodomain of macroH2A1.2 (PDB 1ZR5). The hydrogen bonds (H-bonds) with D224 and R318 expected to stabilize ethanolamine binding are marked in red. **(F)** Molecular docking results of PC polar head, choline (yellow), to the macrodomain of macroH2A1.2 (PDB 1ZR5). The hydrogen bonds (H-bonds) with D227 and R318 expected to stabilize choline binding are marked in red.

3.3.4 The analysis of the available transcriptomics dataset in the search for the physiological role of macroH2A1.2

There are several gene expression datasets which explore the isoform specific roles of macroH2A1.1 and macroH2A1.2 isoforms (Dardenne et al., 2012; Hurtado-Bagès, Posavec Marjanovic, et al., 2020; J. M. Kim et al., 2018; Jinman Kim et al., 2018), of which the analysis of the effects of isoform-specific knock-down on the transcriptomic level is reported only in Hurtado-Bagès, Posavec Marjanovic, et al. We used this dataset obtained to inspect the cell processes specifically regulated by macroH2A1.2. This study shows that macroH2A1 isoforms regulate gene expression in an isoform specific manner in differentiated muscle cells. More specifically, 995 and 372 genes are differentially expressed in macroH2A1.1- and macroH2A1.2-deficient muscle cells, respectively (Hurtado-Bagès, Posavec Marjanovic, et al., 2020).

To better understand which cell processes are specifically regulated by macroH2A1.2 isoform, we used the gene ontology (GO) enrichment analysis of the genes specifically down- and upregulated in muscle cell lacking either macroH2A1.1 or macroH2A1.2. Surprisingly, the GO enrichment analysis of the downregulated genes in macroH2A1.2 depleted cells showed only two of the top ten affected biological processes to be related with muscle development, while eight of them were related with the alcohol, sterol and lipid metabolism (**Figure 35A**). On the other hand, the analysis of upregulated genes showed that the top two enriched biological processes are related to cell adhesion, while the rest of the processes are involved in cell and organism differentiation and development (**Figure 35B**).

To filter for the processes which are specifically affected by macroH2A1.2 and not by macroH2A1.1, we checked the top 10 biological processes affected by the depletion of macroH2A1.1 in the same cells. We did not observe any significant overlap of the biological processes affected by the depletion of macroH2A1.1 and macroH2A1.2. What is more, in the analysis of genes downregulated by the depletion of macroH2A1.1 the top affected processes were involved in vasculature function and development, ossification and extracellular matrix organization (**Figure 35C**). On the other hand, the top enriched processes upon

the analysis of the upregulated genes are implicated in response to virus or generally in immune system and inflammation regulation (**Figure 35D**).

Hence, the analysis of the available RNAseq dataset from the muscle cells depleted of macroH2A1 in an isoform specific manner, indicates that macroH2A1.2 might be specifically implicated in the regulation of alcohol, sterol and lipid metabolism in muscle cells.

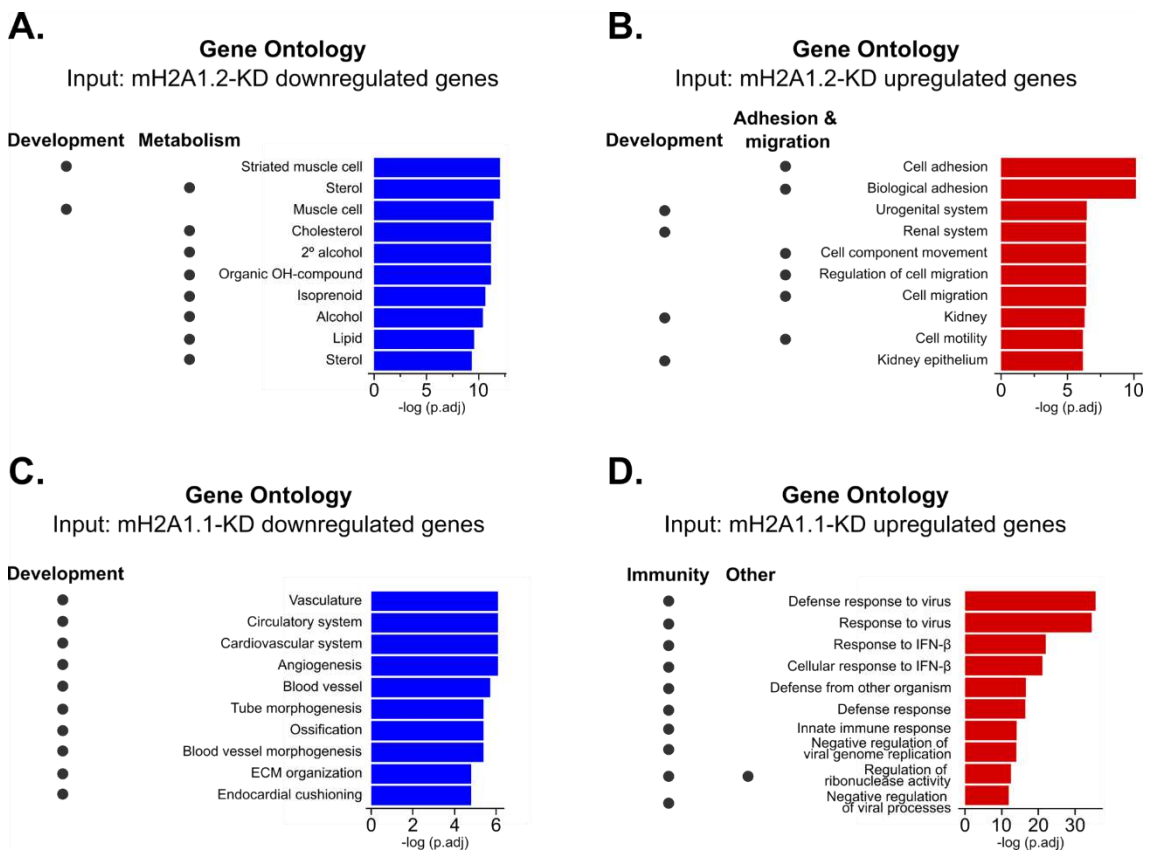


Figure 35. Biological processes related with isoform-specific depletion of macroH2A1.1 or macroH2A1.2.

(A-D) Gene Ontology (GO) enrichment analysis in C2C12 myotubes depleted of macroH2A1.1 or macroH2A1.2. We have used differentially expressed genes with an absolute log₂FC equal or greater than 0.8 (FDR < 0.05) described in Hurtado-Bages et al, 2020. All comparisons refer to the isoform-specific knockdown (KD) condition compared to the control siRNA-treated sample. Top 10 enriched GOs are shown. Exact GO IDs are given in **Annex 8.4**. **(A)** GO analysis using the subset of downregulated genes upon macroH2A1.2-specific knock-down in the muscle cells. To the left is shown the simplified scheme outlining the terms related to development and metabolism. **(B)** GO enrichment analysis using the subset of upregulated genes upon macroH2A1.2-specific knock-down in the muscle cells. To the left is shown the simplified scheme outlining the terms related to development or adhesion and migration. **(C)** GO enrichment analysis using the subset of downregulated genes upon macroH2A1.1-specific knock-down in the muscle cells. To the left is shown the simplified scheme outlining the terms related to development. **(D)** GO enrichment analysis using the subset of upregulated genes upon macroH2A1.1-specific knock-down in the muscle cells. To the left is shown the simplified scheme outlining the terms related to immunity or other processes.

3.3.5 The search for the physiologically relevant cell line: macroH2A1.2 is expressed in mouse brain neurons

The study of our lab which describes the physiological role of the macroH2A1.1–ADP-ribose interaction (Posavec-Marjanović et al., 2017) accentuated the importance of the choice of the cell line for studying the role of macroH2A histone variants. More specifically, the implication of macroH2A1.1 in the regulation of NAD metabolism and mitochondrial respiration was most pronounced in the muscle cells with high expression of macroH2A1.1. Due to possible low affinity ligand–macrodomain interaction, we postulated that the choice of the optimal cell line with a relatively high expression of macroH2A isoform of interest would be crucial for detecting a binding pocket-related physiological function.

MacroH2A1.2 has been reported to be mainly expressed in proliferating cells (Novikov et al., 2011; Judith C. Sporn & Jung, 2012). The analysis of the macroH2A expression in different mouse tissues showed macroH2A1.2 is predominantly expressed in lung, kidney, liver, pancreas, colon and brain of adult mice (Kozłowski et al., 2018b). Unlike the other tissues with significant macroH2A1.2 expression, brain has a limited cell proliferation rate and tissue regeneration capacity. Interestingly, after adipose tissue, brain has the highest lipid content with lipids representing 50% of its dry mass, mainly phospholipids (Bruce et al., 2017). Taken together with the indication that phospholipids and catecholamine neurotransmitters interact with macroH2A1.2 (**Figure 34, Table 1**), the implication of macroH2A1.2 in the regulation of genes involved in lipid metabolism (**Figure 35**) and the decreased sensitivity of total macroH2A1 KO mice to anesthesia (as mentioned on page 75, section 3.3.3), we decided to check whether brain might be most relevant tissue for studying the physiological role of macroH2A1.2.

To delineate the brain regions of interest, we used the publicly available data from Allen Mouse Brain Atlas (Lein et al., 2007) to check for the regions with a significant expression of macroH2A1 mRNA. The hybridization probe for *in situ* hybridization (ISH) detected the mRNA region of macroH2A1 shared by the both alternatively spliced isoforms. The significant expression of macroH2A1 was observed in three brain regions, namely cerebral cortex, (CTX) hippocampal

region (HIP) and cerebellar cortex (CBX). The further image analysis indicates high level macroH2A1 mRNA are present in the hippocampal region and cerebellar cortex, and a much lower expression in cerebral cortex (**Figure 36A**).

Having this in mind, we checked the macroH2A1.2 protein levels in the mouse brain sections by immunofluorescence, using macroH2A1.2 specific antibody. Since macroH2A1.2 is reported to be predominantly expressed in proliferative cells, we first checked its expression in the dentate gyrus of hippocampus, the brain region which maintains proliferative capacity into adulthood (Nelson et al., 2020). The results showed macroH2A1.2 specific signal in the dentate gyrus, with a differential staining in different regions of dentate gyrus (**Figure 36B**, DG). Surprisingly, we also observed a significant macroH2A1.2 expression in brain regions comprised of terminally differentiated cells. More specifically, a prominent macroH2A1.2 signal was observed in cerebellar cortex, seemingly in the nuclei of Purkinje cells (**Figure 36B**, CBX). The staining in the cerebral cortex was less pronounced (**Figure 36B**, CTX). The observations on macroH2A1.2 expression in the non-proliferative brain regions were quite unexpected. To detect if its expression is cell type-specific, we co-stained the brain sections with neuronal and glial cell markers, NeuN and Aldh1, respectively. The results indicate that macroH2A1.2 is specifically expressed in the neurons and not in astrocytes of the dentate gyrus (**Figure 36C**).

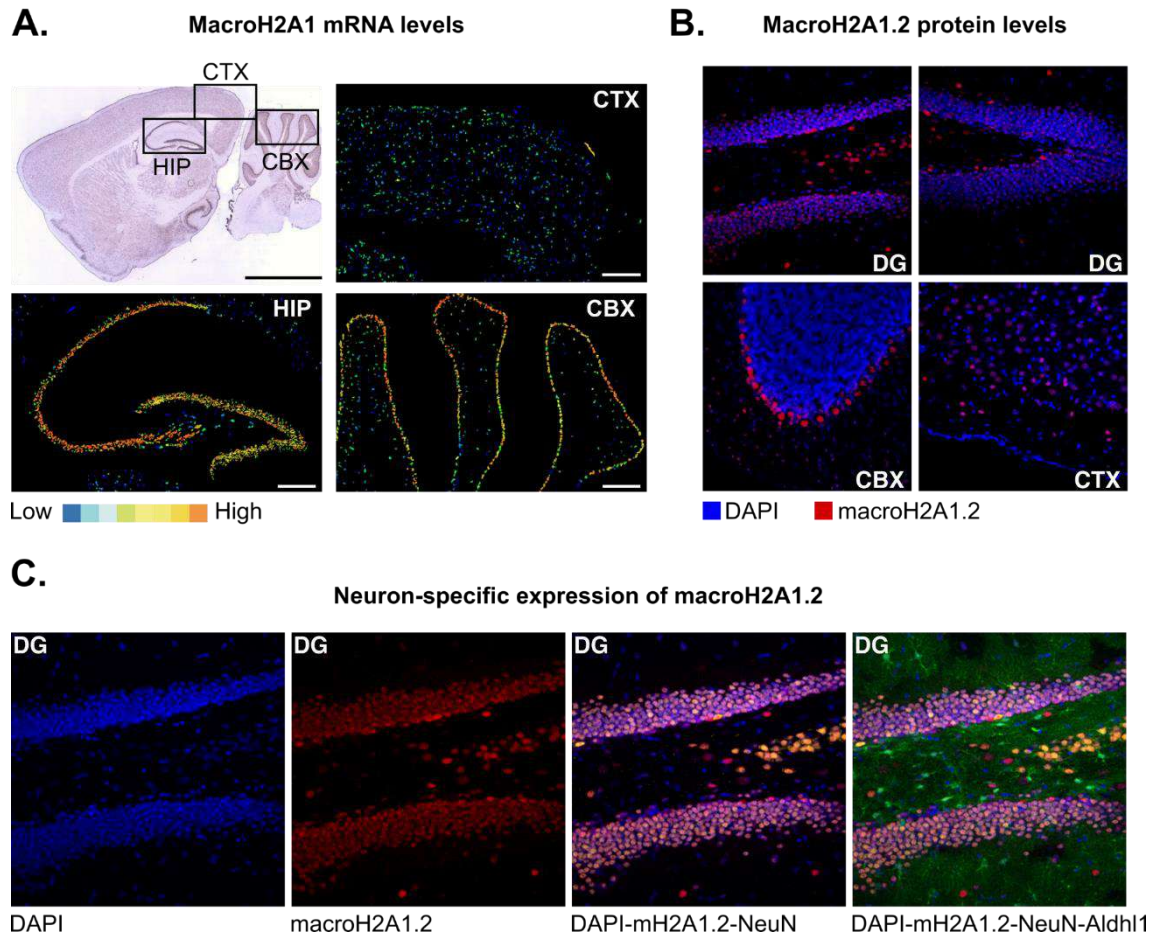


Figure 36. MacroH2A1 and macroH2A1.2 expression in mouse brain sections.

(A) MacroH2A1 mRNA levels from Allen Mouse Brain Atlas (Lein et al., 2007) as analyzed by *in situ* hybridization assay with a probe designed to detect the total macroH2A1 mRNA. The top left panel shows the staining of the sagittal brain section of P56 mouse in which the regions most enriched by macroH2A1 are outlined: hippocampal region (HIP), cerebral cortex (CTX) and cerebellar cortex (CBX). Images with black background show the result of digital image analysis to remove the background staining and indicate macroH2A1 expression levels of each of the outlined regions. (B) Immunofluorescence analysis of the protein levels of macroH2A1.2 using macroH2A1.2 specific antibody in different regions of mouse brain sections: dentate gyrus (DG) as a part of HIP, cerebellar (CBX) and cerebral cortex (CTX). DAPI staining in blue, macroH2A1.2 staining in red. (C) Immunofluorescence results of the DG region as stained by DAPI in blue, macroH2A1.2 in red, neuronal marker (NeuN) in yellow and astrocyte marker (Aldh1) in green.

3.3.6 P19 cell line as a model for studying the role of macroH2A1.2

The significant macroH2A1.2 expression in the non-proliferative regions of the brain, more specifically in terminally differentiated neurons, was unexpected. To check if we can reproduce this result in an *in cellulo* setting, we used the P19 cells, a widely used dynamic cell model which can be differentiated into a mixed neuronal cell population (Jones-Villeneuve et al., 1982).

The expression of macroH2A1.2 in the proliferating P19 cells is relatively low when compared to control hepatoblastoma cells, HepG2 (**Figure 37A**). To inspect for the dynamics of macroH2A expression during cell differentiation, we used retinoic acid to induce the neuronal differentiation of P19 cells (**Figure 37B**). After 12 days, the cells showed a morphology characteristic of a mixed neuronal cell population (Jones-Villeneuve et al., 1982), with clusters of neurons connected via well-defined axons and surrounded by supporting cell types, such as oligodendrocytes, astrocytes and glia cells (**Figure 37B, arrowheads**).

Contrary to the expectation, macroH2A1.1 induction upon neuronal cell differentiation was relatively low on the protein level. On the other hand, the protein level of macroH2A1.2 was significantly upregulated in the differentiated cells, as confirmed by two different macroH2A1.2-specific antibodies. The result becomes even more striking upon comparison with the control HepG2 cells, characterized as cells with a relatively high expression of macroH2A1.2 (H. Chen et al., 2014). As expected (Kozlowski et al., 2018a), we confirmed a significant upregulation of macroH2A2 isoform in the differentiated neuronal cell population (**Figure 37C**).

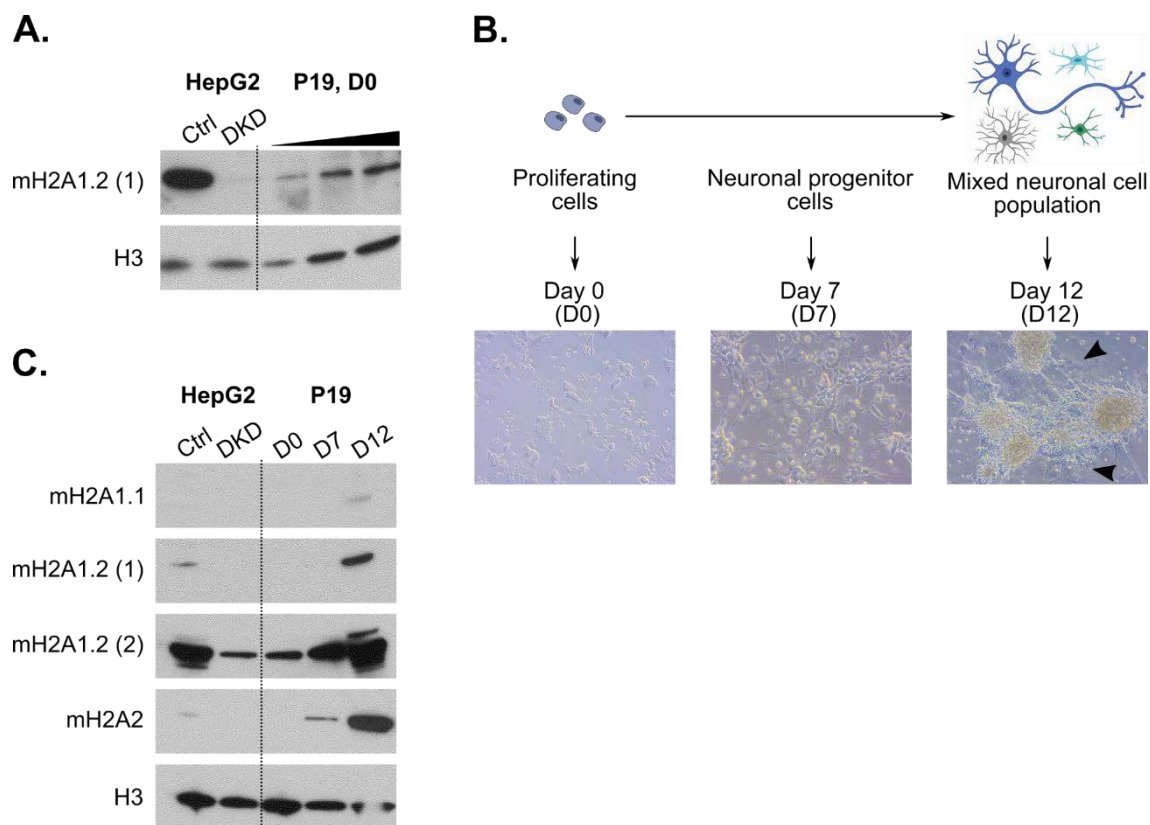


Figure 37. The analysis of the expression of macroH2A isoforms in HepG2 and P19 cells before and after differentiation.

(A) Comparison of the expression of macroH2A1.2 in control (Ctrl) and macroH2A-depleted (double knock-down, DKD) HepG2 cells and proliferating P19 cells. The total proteins extracted from proliferating P19 cells were loaded in 3 different concentrations, in an increasing manner. H3 was used as a loading control. **(B)** A schematic of the differentiation of P19 cells. Proliferating (pluripotent) P19 cells at day 0 undergo differentiation upon retinoic acid (RA) treatment towards neuronal progenitor cells (NPCs) which appear after 7 days of RA treatment, and they differentiate to the mixed neuronal cell population after 12 days (top panel). Representative light microscopy images taken at day 0, day 7 and day 12 are shown in the bottom panel. Arrowheads indicate the supporting cell population. **(C)** The comparison of the expression of macroH2A1.1, macroH2A1.2 and macroH2A2 isoforms in the HepG2 Ctrl and DKD to the P19 cells at all three differentiation stages described in **(B)** (D0, D7 and D12). MacroH2A1.2 expression was evaluated by two different antibodies detailed in the Material and Methods. Histone H3 was used as a loading control.

Due to significant upregulation of macroH2A1.2 expression, we hypothesized that the dynamic model of neuronal cell differentiation might be optimal for studying the physiological role of macroH2A1.2. Thus, we used three different previously described hairpins (Creppe, Janich, et al., 2012) to stably deplete either both macroH2A1 isoforms, or specifically macroH2A1.2 isoform from the P19 cells. On the mRNA level, we could detect a basal macroH2A expression of all three macroH2A isoforms in the proliferative stage of P19 cells. Furthermore, results clearly showed the depletion of both macroH2A1.1 and macroH2A1.2 when total

macroH2A1 is targeted using hairpins 1 and 2, with hairpin 1 showing a better efficiency. We further observed a specific macroH2A1.2 depletion by hairpin 3, while there was no compensatory expression of macroH2A1.1 or macroH2A2. In line with our previous observation on the protein level, we observed the upregulation of all three macroH2A isoforms during differentiation (**Figure 38A**).

The proliferation rate of the macroH2A1-depleted panel of P19 cell lines remained unaffected, as validated by two parallel methods (**Figure 38B, 38C**). On the level of cell morphology, macroH2A1-depletion did not impede the neuronal cell differentiation. More specifically, the differentiated cells successfully formed neuronal clusters that were well-connected via thick axon fibers. However, the cells depleted of macroH2A1 by hairpin 1 and the cells specifically depleted of macroH2A1.2 by hairpin 3 showed an apparent decrease in the supporting cell population (**Figure 38D, arrowheads**). The immunoblotting results confirmed the specificity and efficacy of the depletion of both macroH2A1 by hairpin 1, which blocked the induction of both macroH2A1.1 and macroH2A1.2 isoforms upon the differentiation with a higher efficiency than hairpin 2 (**Figure 38E**). Furthermore, hairpin 3 specifically hindered the induction of macroH2A1.2 upon differentiation, while it did not affect the expression of macroH2A1.1. None of the macroH2A1 directed hairpins affected the protein expression of macroH2A2, thus ruling out compensatory effects (**Figure 38E**).

To further inspect the cell differentiation efficiency upon macroH2A1 depletion, we checked the expression of the neuronal cell markers NeuN, Map2 and b-III-tubulin, and the glial cell marker, Gfap. Supporting the observed morphological changes, the upregulation of all three neuronal markers upon the differentiation of macroH2A1-depleted cells indicated successful neuronal differentiation. However, we observed a marked absence of the upregulation of the glial cell marker Gfap further reinforcing the observations made on the level of cell morphology (**Figure 38F**).

Taken together, our preliminary analysis of macroH2A1.2 in brain, showed a particular pronounced expression of macroH2A1.2 in neurons. This is the first time a high level of macroH2A1.2 expression was observed in terminally differentiated cells. The successful validation of the high expression of

macroH2A1.2 *in cellulo* upon neuronal differentiation of P19 cells, indicated that this model system might be optimal for studying the physiological role of macroH2A1.2. Finally, our preliminary data in macroH2A1 and 1.2 depleted P19 cells suggested that macroH2A1.2 might have a role in cell fate choice. Specifically, we observed that macroH2A1-depleted cells successfully differentiated into neurons, while the differentiation of glial cells was impaired.

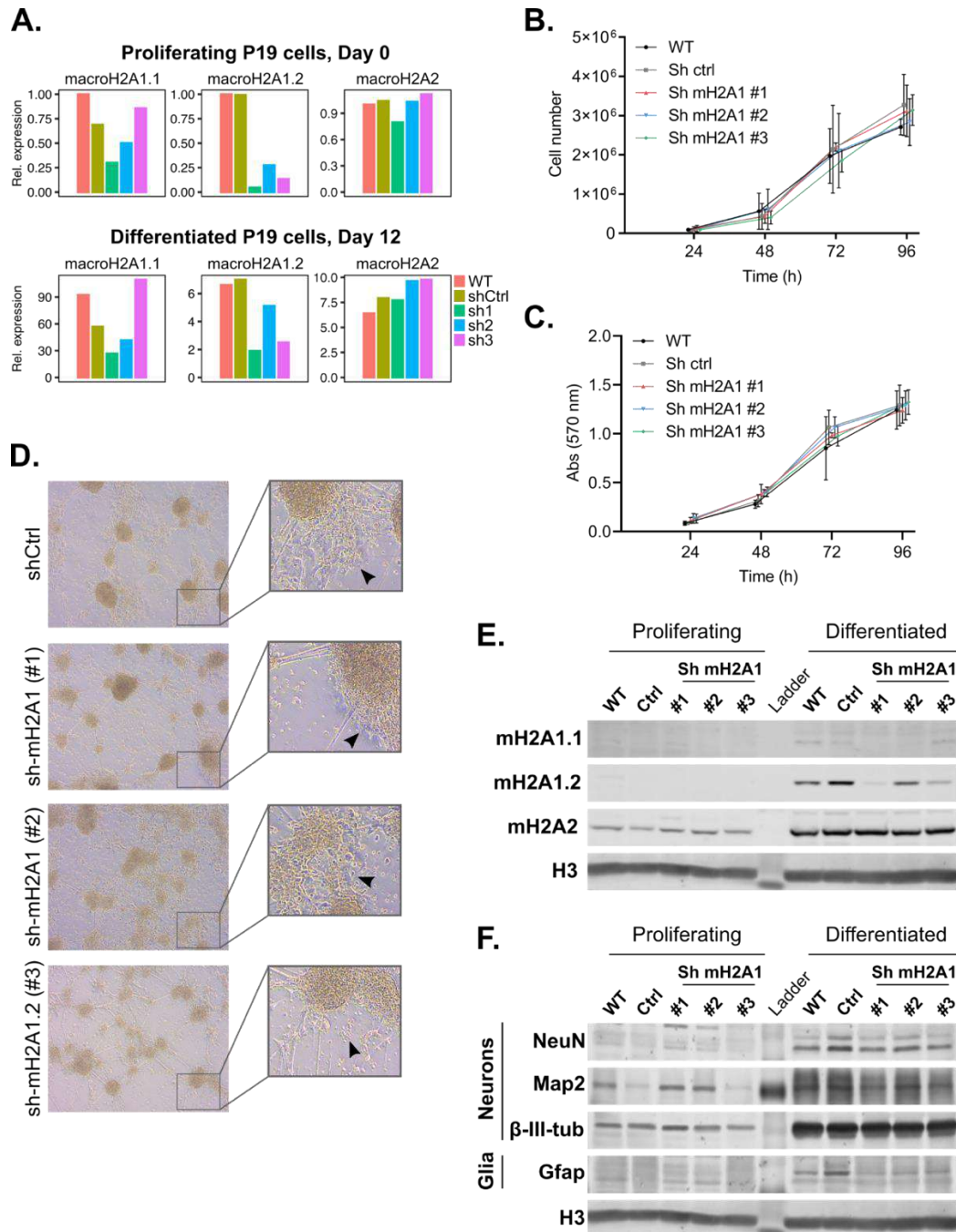


Figure 38. The characterization of macroH2A1 depleted P19 cells.

Figure 38. The characterization of macroH2A1 depleted P19 cells.

(A) The expression of macroH2A1.1, macroH2A1.2 and macroH2A2 in the proliferating (Day 0) and differentiated (D12) P19 cells was measured by RT-qPCR in the wild-type (WT), shRNA control (shCtrl), macroH2A1-depleted cells with two different hairpins (sh1 and sh2) and macroH2A1.2-depleted cells (sh3). The expression of each of the isoforms in the WT proliferating cells was normalized to 1. **(B, C)** The proliferation rate of proliferating (Day 0) cells described in **(A)** was measured by cell counting **(B)** or by MTT assay followed by measuring the absorbance at 570nm **(C)** daily, every 24 hours, for four days. The mean of three independent experiments is shown. Nudging of datasets has been done to allow proper visualization of the error bars (SD). **(D)** Light microscopy images of differentiated cells described in **(A)** (magnification 5x). Arrowheads represents the expected supportive cell population. **(E)** Western blot showing the expression of macroH2A1.1, macroH2A1.2 and macroH2A2 in all cell lines described in **(A)** at the proliferating and differentiated states. H3 was used as loading control. **(F)** Analysis of neuronal markers NeuN, Map2 and β -III tubulin, and glial cell marker Gfap by western blot. Expression levels are shown for all cell lines from **(A)** at the proliferating and differentiated states. Histone H3 was used as loading control.

3.3.7 The characterization of macroH2A1.2 knock-out mice

MacroH2A1.2 specific whole-body knock-out mice were established by using Dre-Rox system, in which the macroH2A1.2 specific exon was surrounded by Rox sites (**Figure 39A**), recognized by Dre recombinase. The Dre-deleter mice with a Rox-flanked macroH2A1.2 specific exon had a mixed 129/Ola (agouti) and C57BL/6 (black) genetic background. To obtain the C57BL/6J genetic background, animals were backcrossed twelve generations into C57BL/6J genetic background. By definition, each successive generation has ~50% less of the donor genome, hence the twelfth generation is expected to have more than 99.9% of the recipient genome, in this case C57BL/6J.

3.3.7.1 The skewed genotype and sex ratio

We used the backcrossed mice to obtain the macroH2A1.2 knock-out mice by crossing the heterozygous Rox-flanked macroH2A1.2 mice (referred to as Het x Het). Surprisingly, even after twelve generations of backcrossing into the C57BL/6J background, we still observed agouti coat occasionally appearing in the offspring (not quantified). Perplexingly, the analysis indicated a slightly skewed sex ratio in favor of male offspring (**Figure 39B**) (in reference to Kruger et al., 2019) accompanied by the skewed genotype towards the heterozygous animals (**Figure 39C**). More specifically, although we did not observe any effect on the litter size upon mating heterozygous mice, we observed an increased number of heterozygous and decreased number of homozygous pups, with the ratio WT : HET : KO ~ 1 : 3.5 : 1 (**Figure 39C, right panel**).

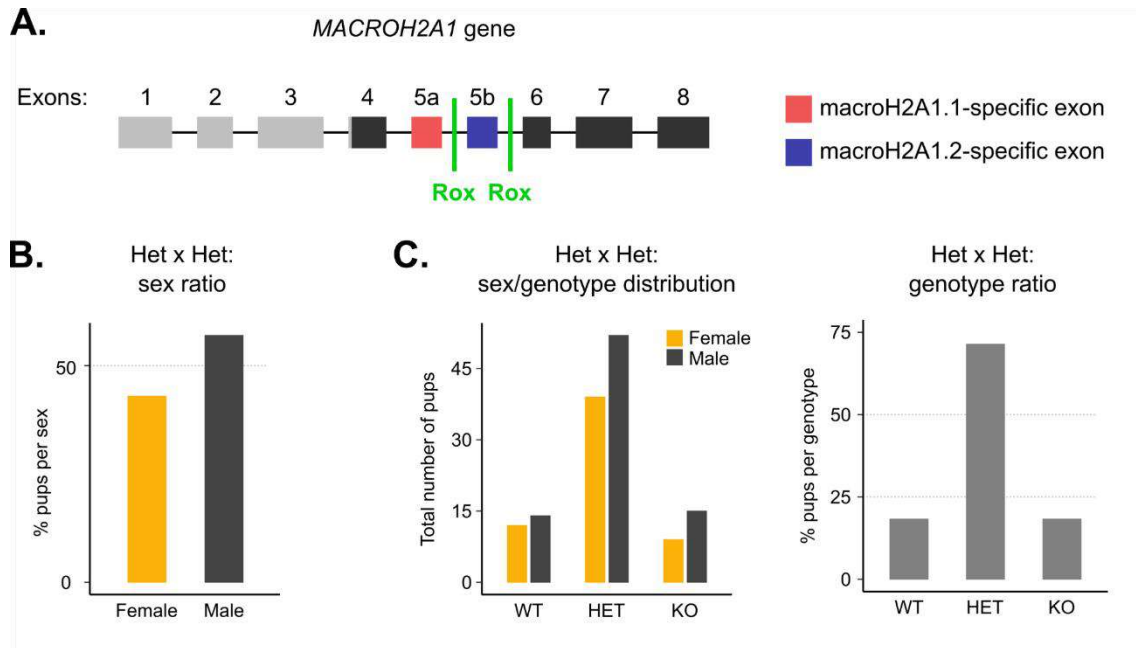


Figure 39. Design of macroH2A1.2 knock-out mice and offspring sex and genotype ratio analysis.

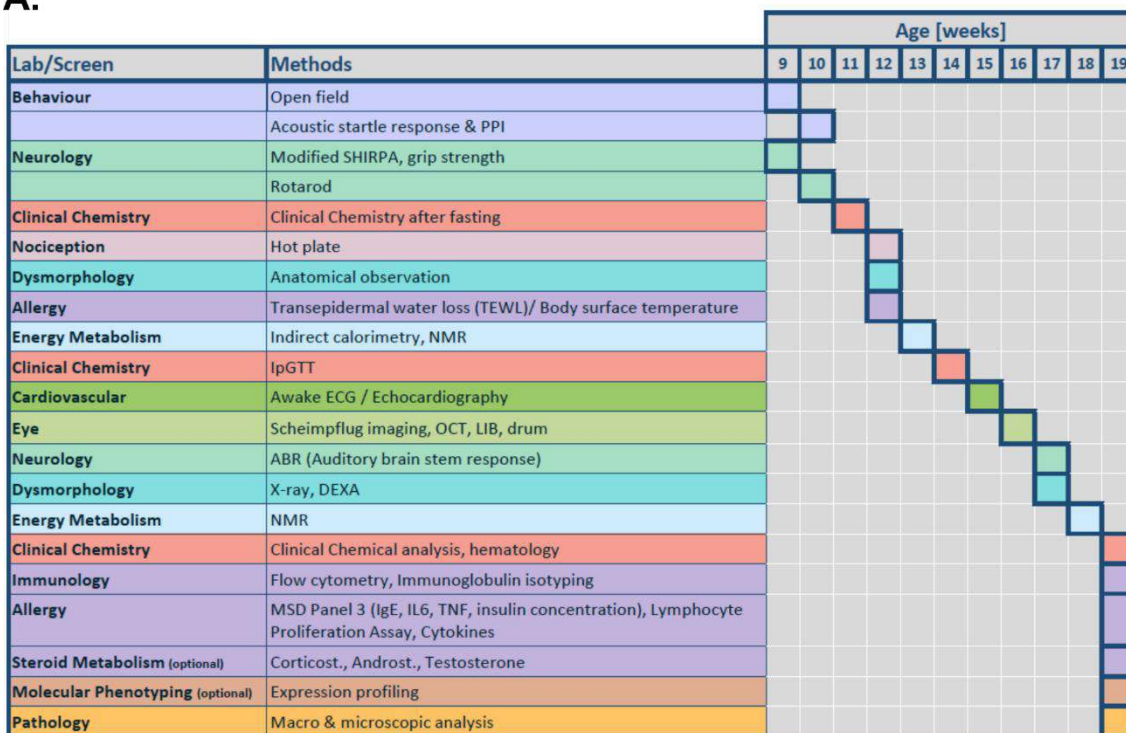
(A) MacroH2A1.2 KO mice were obtained by Dre-Rox system in which the Rox sites were introduced to surround macroH2A1.2 specific exon 5 (5b). (B) The sex ratio upon breeding of macroH2A1.2-floxed heterozygous mice was done to get the KO mice was skewed toward an increased number of male and decreased number of female mice, independently of the genotype. The line at 50% indicates the expected ratio as a result of Het x Het crossing. (C) The sex and genotype of the pups obtained by Het x Het breeding was affected. The left panel shows the total number of pups obtained with respect to both the sex and genotype. The right panel shows the ratio of the genotypes of the animals obtained by Het x Het breeding, independently of sex. The lines at 25% and 50% indicate the expected Mendelian distribution of the phenotype, namely 1:2:1.

3.3.7.2 The phenotyping pipeline for the identification of the physiological role of macroH2A1.2

To evaluate if the macroH2A1.2 knock-out affects the mouse physiology in any way, we performed a comprehensive phenotyping screen using a well-established phenotyping pipeline offered by the German Mouse Clinic. This involved a systematic analysis of different aspects of mouse physiology over ten weeks (Figure 40A). To obtain robust phenotyping results, the standard procedure requires for the screening to be performed on 15 wild type and 15 knock-out animals in one batch. Due to the unusually skewed genotype toward heterozygous animals, the obtained number of homozygous pups upon crossing heterozygous macroH2A1.2 floxed mice was a serious limitation. Thus, we were unable to obtain the number of pups necessary for the phenotyping screen to be

done and analyzed in one batch, rather it was done in three separate batches of mice. Additional problem arose from the skewed sex ratio. Thus, the total number of animals that entered phenotyping was affected by both of these breeding issues, resulting in 5 and 9 wild type and knock-out female mice, and 11 and 10 wild type and knock-out male mice to enter the screening (**Figure 40B**).

A.



B. Genotypes of three batches of macroH2A1.2 KO mice entering phenotyping screen

| Female | WT | KO | Male | WT | KO |
|--------------|----------|----------|--------------|-----------|-----------|
| Batch 1 | 4 | 5 | Batch 1 | 7 | 4 |
| Batch 2 | 0 | 0 | Batch 2 | 1 | 5 |
| Batch 3 | 1 | 4 | Batch 3 | 3 | 1 |
| TOTAL | 5 | 9 | TOTAL | 11 | 10 |

Figure 40. The number of animals entering the outlined GMC phenotyping screen.

(A) The outline of the standardized GMC phenotyping pipeline indicating the broad range of tests performed over 10 weeks, in mice from 9-19 weeks old, with a goal to identify novel phenotypes. (B) Table indicating the numbers and genotypes of female (left) and male (right) mice that entered the phenotyping screen in three separate batches.

3.3.7.2.1 MacroH2A1.2 knock-out causes mild metabolic alterations in mice

First observed differences between KO mice and wild-type littermates were related to metabolism. The body weight was decreased in both the male and

female macroH2A1.2 KO animals, with the more pronounced differences in females (**Annex 8.5, top panel**). Interestingly, the significantly decreased food intake was observed only in male mice (results not shown). Body composition analysis indicated that both sexes had a decreased body weight due to a reduced lean body mass and increased fat content, which remained consistent over time (**Annex 8.5, top panel**). Furthermore, both sexes showed decreased carbohydrate and increased lipid oxidation during the active phase (**Annex 8.6**), while the male animals showed significantly lower RER (data not shown), presumably due to pronounced lipid oxidation rate. When fasted, KO females showed slightly more pronounced weight loss, while fasting non-HDL cholesterol and triglycerides were slightly decreased in KO males (data not shown). Since the animals entered the phenotyping in different batches, we need to consider that high dissipation of the phenotyping results might be the consequence of a batch effect. Taken together, macroH2A1.2 seems to have a complex effect on metabolism which might have something to do with lipid metabolism but more hypothesis driven tests need to be performed to obtain conclusive data.

3.3.7.2.2 MacroH2A1.2 knock-out causes significant behavioral alterations in mice

The most exciting KO-specific phenotypes were related to behavior. Behavioral tests in 9-week-old mice showed a slightly decreased vertical activity in the KO mice of both sexes, with a more pronounced effect in females, but no effect on horizontal activity (**Figure 41A**). Additionally, the KO mice spent more time in center with a slightly increased distance from the center, although the results were not statistically significant (**Figure 41B**). However, the most significant result in the phenotyping screen was observed when 10-week-old mice were exposed to the acoustic startle click at 110 decibels (db) with a broad frequency spectrum (2-25kHz). The results showed that KO mice have a significantly impaired acoustic startle reactivity in a sex-independent manner (**Figure 41C**), without an effect on the prepulse inhibition (**Figure 41D**). These results indicate the KO mice do react to the loud sounds and show normal sensory adaptation, but the overall amplitude of the reaction is lower. The hearing sensitivity was checked in 17-week-old mice by determining the sound pressure threshold for the detection of neural activity in auditory brainstem response test (ABR). There

was a subtle increase in ABR threshold in female KO mice at 24 and 30kHz, while no effect was observed in male mice (**Figure 41E**). These results indicate that 17-week-old female mice have impaired hearing of the high frequency sounds, while male mice remain unaffected.

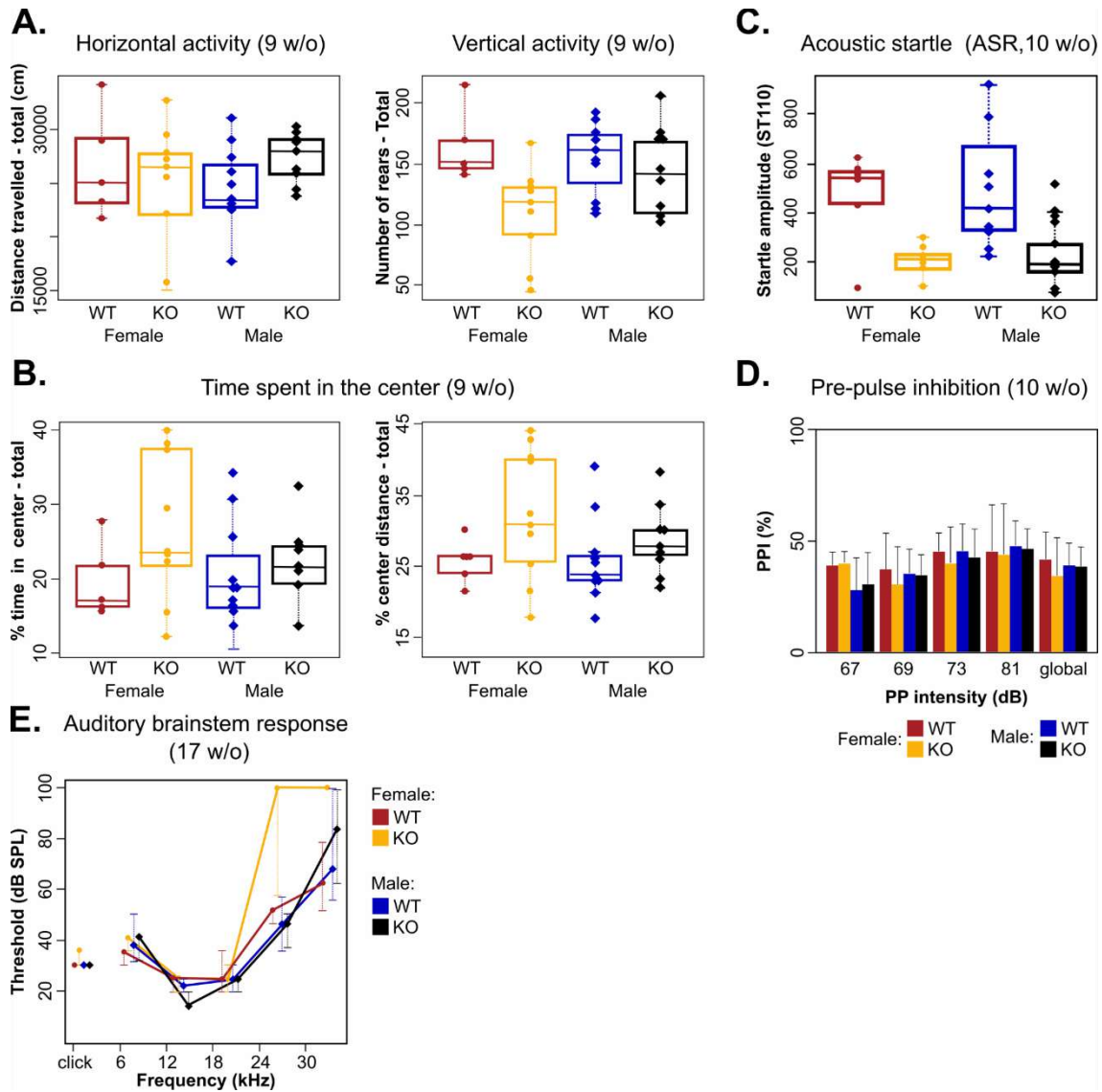


Figure 41. Phenotyping results of neurological and behavioral experiments on macroH2A1.2 wild-type (WT) and knock-out (KO) female and male mice.

(A) Open field test results in 9 weeks old mice (9 w/o) show no effect on horizontal locomotion (left panel), and a slightly decreased vertical locomotion in macroH2A1.2 KO mice (right panel). **(B)** Open field test results in 9 w/o show slightly increased time spent in the center in the KO mice of both sexes. **(C)** Acoustic startle reactivity in 10 w/o mice to the broad frequency spectrum noise (2-25kHz) at 110 dB is significantly decreased in macroH2A1.2 KO mice in a sex-independent manner (2-way ANOVA genotype effect, $p < 0.001$). **(D)** Pre-pulse inhibition experiment in which 10 w/o mice were exposed to an increasing sound pressure pre-pulse (67-81dB) followed by a 110dB startle, show normal sensory adaptation of macroH2A1.2 KO mice. **(E)** Auditory brainstem response (ABR) test in 17 w/o mice shows reduced hearing sensitivity to high frequency tones (24 and 30kHz) in the female macroH2A1.2 KO mice.

Obtained results indicated decreased exploratory activity with an increased novelty-seeking and risk-taking behavior or possibly a marked indifference to risk. This conclusion is based on increased time spent in center, accompanied by the decreased reactivity to acoustic startle. The decreased reactivity to startle was observed when mice were exposed to broad frequency sound at 110db in both sexes, whereas the impaired hearing was observed in older mice and only in females exposed to high frequency sounds.

Taken together, loss of macroH2A1.2 affects the metabolism of mice in a complex and in part sex-dependent manner. The key sex-independent phenotype of macroH2A1.2 null mice are changes in behavior and reactivity to external stimuli.

In conclusion, the results of this study indicate that the putative physiological ligands of macroH2A1.2 are phospholipids. Furthermore, macroH2A1.2 seems to be implicated in the transcriptional regulation of genes involved in lipid metabolism. There are some indications it might be involved in normal membrane function due to previously observed fusion defects in myotubes (Hurtado-Bagès, Posavec Marjanovic, et al., 2020), but also due to the non-Mendelian genotype distribution in mice, without affecting their viability. Finally, due to the significant increase of macroH2A1.2 during neuronal cell differentiation, its high expression in mouse brain and the observed behavioural defects in macroH2A1.2-specific knock-out mice, we postulate that macroH2A1.2 has a physiological role in the brain and propose neuronal differentiation cell system and brain to be optimal models for the study of physiological roles of macroH2A1.2.

3.3.8 Contributions

This project was conceived and supervised by Marcus Buschbeck and myself. I took part in the design, performance and/or analysis of all of the experiments, with the exception of molecular docking, similarity search, microscale thermophoresis (MST) and mouse experiments. Gunnar Knobloch took part in the design of macroH2A1.2 binding pocket mutant, and taught me and provided his expertise in the design and performance of protein crystallization, isothermal titration calorimetry (ITC), and helped with the design of ADP-ribose pulldown

assay in the lab of Andreas Ladurner. Anke Mager, Julius Perschel, Janik Hedderich, Jonas Kammertöns and Peter Kolb performed the molecular docking, similarity search, and microscale thermophoresis (MST). Jesus Garcia and Marta Frigole taught me how to prepare samples preparation for STD-NMR and JG performed and analyzed STD-NMR experiments in the lab of Xavier Salvatella. Laura Amoedo established the panel of stable macroH2A1 KD P19 cell lines and characterized them under my supervision. Joan Miro Blanch and Oscar Yanes took part in the design of the untargeted approach to detect macroH2A1.2 ligands and performed and analyzed the LC/MS-MS data with the help from Maria Vinaixa and Alexandra Junza. Manja Thorwirth showed me how to perform the immunofluorescence staining of the mouse brain slices and Giacomo Masserdotti imaged them and helped with interpretation, in Magdalena Goetz's lab. Maximilian Lassi and Raffaele Teperino took care of the breeding of the colony of macroH2A1.2 knock-out mice and provided the animals for the phenotyping that was entirely performed and analyzed by the German Mouse Clinic.

4 Discussion

4.1 Discussion I

4.1.1 The origin of a macrodomain-containing histone variant

The macrodomain is the defining feature of all macroH2A histone variants. By focusing our evolutionary analysis on the amino acid sequence of macroH2A macrodomains, we were able to delineate the evolutionary history of this atypical histone variant. MacroH2A first appeared in protists ancestral to modern animals, filastereans and breviate, with the original gene resembling that of macroH2A1.1. The presence of the macroH2A gene was retained and further diversified in vertebrates, while it was sporadically lost in some invertebrates with accelerated evolution, such as *Drosophila*. Interestingly, the loss of macroH2A in these species correlates with a reduced number of PARP genes (Rack et al., 2016). Gene duplication in a common ancestor of vertebrates resulted in the appearance of macroH2A2, an ADP-ribose binding deficient isoform. Consecutively, the alternatively spliced exon encoding for macroH2A1.2 appeared in a common ancestor of jawed vertebrates, adding to a second example of NAD signaling-inert macroH2A histone variant which can be incorporated into chromatin. The presence of macroH2A-encoding genes in haptist and breviate suggest that the fusion of a macrodomain and a histone-fold occurred before the split between fungi and animals, and thus much earlier than previously thought. Due to the scarcity of the data, we cannot fully rule out the possibility that the appearance of macroH2A in the haptist or breviate was caused by independent fusion events in an example of convergent evolution.

Macrodomains are present in all forms of life. Interestingly, it was shown that some viral and bacterial macrodomains bind ADP-ribose (Catara et al., 2019), while the first extensively characterized ADP-ribose-binding macrodomain was archaeal (Karras et al., 2005). However, the histone variant macroH2A, which is presumably a result of the fusion of histone and macrodomain encoding genes, is detected in eukaryotes, which are the product of endosymbiosis involving both archaea and bacteria. Interestingly, histone proteins, including the histone variant macroH2A, have a longer half-life than average cellular proteins (W. Chen et al., 2016; Commerford et al., 1982; Fornasiero et al., 2018; Mathieson et al., 2018).

It is intriguing to speculate that evolution may have selected for the fusion of a macrodomain to a histone in order to provide an increased half-life and a strictly nuclear localization to an ADP-ribose sensor. The ability to influence NAD⁺-dependent reactions in a compartmentalized manner might have provided an advantage to eukaryotes, allowing them to adapt to changes in their environment by adopting different states, consistent with our analysis of *Capsaspora* life stages.

4.1.2 Positive selection of few key residues drives evolutionary change

A major conclusion of our study is that ADP-ribose binding is the most ancient trait of the histone variant macroH2A. The fusion of a macrodomain to a histone provided cells with an abundant and strictly nuclear element for the regulation of NAD-dependent reactions in the nuclear compartment. The comparative analysis of *Capsaspora* and vertebrate macroH2A macrodomains provides us with a better understanding of how the ADP-ribose and NAD metabolism-related functions of macroH2A were shaped through evolution. The *Capsaspora* macrodomain bound ADP-ribose with almost 10 times higher affinity than its mouse counterpart, and as a consequence was a much more potent PARP1 inhibitor. We were able to map this functional difference to the substitution of only two residues that close the binding pocket in the *Capsaspora* macrodomain, glutamine 225 and asparagine 316. We conclude that this happened in a strictly defined order, with the Q225E substitution preceding N316R and determining the decrease in affinity. While N316R is strongly selected for, the physiological reason remains less clear. Interestingly, a similar course of evolution was recently reported for haemoglobin, where only two historical substitutions in the ancestral protein decrease oxygen affinity, while enabling tetramerization and cooperativity (Pillai et al., 2020).

4.1.3 The macroH2A – PARP1 – NAD⁺ axis sustains catabolism in non-proliferative stages

The decreased stringency of macroH2A-dependent regulation was likely selected for along the vertebrate stem lineage. In vertebrates, macroH2A1.1 takes part in cross-compartmental regulation of NAD metabolism by inhibiting PARP1 activity in the nucleus. An important prerequisite for this function is a higher expression of macroH2A1.1 than PARP1, as is the case in terminally differentiated myotubes, but not proliferating myoblasts (Hurtado-Bagès, Knobloch, et al., 2020; Posavec-Marjanović et al., 2017). *Capsaspora* has three different life stages, two proliferative stages and a non-proliferative, spore-like, cystic stage, to which it transitions in unfavorable environmental conditions (Sebé-Pedrós et al., 2013). We found that the ratio between macroH2A and PARP1 was the highest in the cystic stage, suggesting that the putative macroH2A-dependent compartmental regulation of NAD metabolism might mostly operate in the non-proliferative stage of *Capsaspora*, similar to what was observed in vertebrates. Our transcriptional analysis further indicated metabolic differences between the three stages, with the cystic-stage being the most divergent. For example, V-ATPases were highly expressed only in the cystic stage. These proton pumps enable survival through whole-cell acidification in bacteria (Beales, 2004) and are correlated with autophagy, endosome formation and compartmental acidification in metazoans (Mijaljica et al., 2011; Wilms et al., 2017). Spore-like cells show an increased degradation of nucleotides and surprisingly high expression of mitochondrial genes involved in oxidative phosphorylation. The cystic stage was further characterized by high expression of genes involved in biosynthesis of NAD and the amino acids that can serve as NAD precursors, while the overall anabolic pathways were downregulated. Curiously, several reports indicate the importance of ADP-ribose and NAD during sporulation and germination of bacterial spores (Huh et al., 1996; P. Setlow & Kornberg, 1970). Bacteria might rely on their NAD-based redox potential for germination, more specifically the accumulation of the reduced forms may have an important role in the initiation of germination (R. Setlow & Setlow, 1977). It will be interesting to test whether a similar mechanism enables *Capsaspora* to re-enter the proliferative stages of its

life cycle. Taken together, our results indicate that *Capsaspora* uses a combination of both eukaryotic and bacterial mechanisms for survival in nutrient poor environments. Our data indicates that NAD biosynthesis is channelled to life-sustaining catabolic reactions in the cystic stage. This coincides with a high macroH2A-to-PARP1 ratio, which would be predicted to limit the nuclear NAD⁺ consumption.

4.1.4 The need for the compartmental regulation of NAD metabolism

NAD homeostasis is vital for optimal cell function, and by extension, for organismal health (Covarrubias et al., 2021; Rajman et al., 2018). The NAD⁺ pools of independent compartments communicate and are connected through the shared NAMPT reaction of the salvage pathway, and transport of NAD⁺ and its precursors (Cambronne & Kraus, 2020), thus creating a requirement for communicating and regulating local needs. But why is the compartmental regulation of NAD levels essential? NAD⁺-dependent enzymes differ in their K_m as much as 100-fold, from 2-1000 μM (Berger et al., 2005; Canto et al., 2015). Thus, their activities are controlled by the local NAD⁺ concentrations, which differ between cellular compartments (Cambronne et al., 2016; Ryu et al., 2018). This is well illustrated by the example of nuclear NAD⁺ consuming enzymes PARP1, sirtuin 1 (SirT1) and PARP2, which have decreasing NAD⁺ affinity, respectively. They form a regulatory loop, whereby SirT1 can inhibit PARP1, while PARP2 regulates the activity of SirT1, depending on the nuclear NAD⁺ levels (Canto et al., 2015). Furthermore, some of the nuclear NAD⁺ consuming enzymes, such as SirT6, have an even higher NAD⁺ affinity than PARP1. Interestingly, macroH2A1.1 has been shown to interact with ADP-ribosylated SirT7 in an ADP-ribose-binding pocket dependent manner (Simonet et al., 2020). This raises the possibility that macroH2A1.1 may be involved in a more general regulation of nuclear NAD⁺ metabolism.

4.1.5 Relevance of macroH2A for compartmental regulation in the context of evolution

The fine-tuned regulation of NAD⁺ consumption and localization is particularly relevant during the cell differentiation, when the requirements for NAD⁺ compartmentalization change (Ryu et al., 2018). It is conceivable that non-proliferative states require less NAD⁺ for nuclear functions, such as replication-associated DNA repair, and thus benefit from prioritizing NAD usage for life-sustaining functions, such as ATP production through mitochondrial respiration. The function of macroH2A1.1 as a nuclear NAD⁺ regulator has been demonstrated in differentiated myotubes (Posavec-Marjanović et al., 2017). However, macroH2A1.1 is expressed in a wide array of tissues (Kozłowski et al., 2018b). Additionally, its upregulation is also observed during the differentiation of other tissues apart from muscle, such as skin and colon (Creppe, Janich, et al., 2012; Judith C. Sporn & Jung, 2012). This suggests that its function is more widespread and generally related to cell differentiation. While macroH2A is only present in a few protist taxa, it is associated with increased cellular plasticity in animals. In this regard, it is worth noting that *Capsaspora*, which is closely related to animals, has been shown to have a complex life cycle, and shares several mechanisms of spatial cell differentiation with animals (Sebé-Pedrós, Ballaré, et al., 2016; Sebé-Pedrós, Peña, et al., 2016).

Furthermore, the consolidation of macroH2A with decreased ADP-ribose affinity in vertebrates coincided with the diversification of NAD biosynthesis pathways, which provided additional elements of regulation (Bockwoldt et al., 2019). The increased complexity of higher organisms requires intricate fine-tuning of cell processes. This is often achieved by increasing the number of proteins in a regulatory network to allow for efficient sensing of subtle changes in the environment, thus enabling fast response to environmental cues. It has been suggested that changes at the periphery of metabolic networks are more likely to endow the system with the high probability of gaining beneficial changes than the changes in the rigid core of the pathway (Morowitz, 1999). Our data suggest that the emergence of macroH2A in protists could be such a peripheral change in the

network of NAD metabolism and that it has been selected for during the evolution of metazoans.

Here, we described the evolution of a histone variant as an inhibitor of nuclear NAD⁺ consumption, adding a unique mechanism for compartmental metabolic regulation. A better understanding of these regulatory mechanisms will be informative for ongoing development of the therapies targeting NAD metabolism and signaling (Covarrubias et al., 2021; Rajman et al., 2018). Future work will have to further elucidate how the metabolic function of macroH2A integrates with its other molecular functions, such as the regulation of higher-order chromatin architecture (Douet et al., 2017), DNA repair (Sebastian et al., 2020) and transcription (Lavigne et al., 2015).

4.2 Discussion II

The majority of eukaryotic macrodomains characterized up-to-date take part in ADP-ribose binding and processing, and are referred to as ADP-ribose readers and ADP-ribosylation erasers, respectively. Only three macrodomains are known not to interact with ADP-ribose and their ligands remain unknown. These are the macrodomains of macroH2A1.2 and macroH2A2 histone variants, and the macrodomain of GDAP2. Our identification of phospholipids as potential natural ligands of the macroH2A1.2 binding pocket has opened a new avenue for future research. I will start this sub-chapter by discussing what distinguishes macrodomains unable to bind ADP-ribose from macrodomains with ADP-ribose affinity and then briefly discuss the approach that led us to the identification of phospholipids as potential ligands. As the central part of this discussion, I will reflect on how described functions of phospholipids in membranes, cell specification and brain function might be related to observed, yet unexplained phenotypes of macroH2A1.2 loss of function.

4.2.1 The specific features of non-ADP-ribose binding macrodomains

By now, there are many studies addressing the structural features of ADP-ribose binding macrodomains (Rack et al., 2016). As described in previous studies, and reinforced in the here presented study on the evolutionary conservation of macrodomains, the macrodomains are defined by the specific 1276354 order of the β -sheets surrounded by α -helices thus making up the globular macrodomain structure. Furthermore, the ADP-ribose binding is stabilized by several highly conserved amino acid residues, three of which are encoded by the alternatively spliced exon 5 which differentiates macroH2A1.1 and macroH2A1.2 macrodomains, and consequently, renders the macroH2A1.2 macrodomain to bind ADP-ribose. In our study, we focused on characterizing how the variant amino acids encoded by the mutually exclusive exons 5 contribute to the architecture and physicochemical characteristics of the binding pocket. The

comparison of the three non-ADP-ribose binding macrodomains with representatives of ADP-ribose binding macrodomains delineated the features, which differentiate them. It was previously described that the insertion of a loop composed of the three amino acids EIS in macroH2A1.2 disables the part of the binding pocket accommodating the adenosine moiety of ADP-ribose in macroH2A1.1. We show here that macroH2A2 has a similar insertion of DIS in the same position. In contrast, GDAP2 possesses the conserved aspartate residue present in all ADP-ribose binding macrodomains. This indicates that there must be additional features determining the incapacity of orphan macrodomains to bind ADP-ribose. We postulate that one of the key previously undescribed features of non-ADP-ribose binding macrodomains is the highly charged D-rich loop, in place of the flexible G-rich loop of ADP-ribose binding macrodomains. Based on our evolutionary study of macroH2A macrodomains, the amino acids in the D-rich loop are highly conserved suggesting physiological importance. Additionally, our *in silico* docking analyses predict the establishment of relevant salt bridge and H-bond interactions between the top hits and the residues located in the D-rich loop, which is further reinforced by the pH-dependent binding observed *in vitro*. Notably, the D-rich loops of the three non-ADP-ribose binding macrodomains are expected to differ in their rigidity. While the D-rich loops of macroH2A1.2 and macroH2A2 are more similar, the D-rich loop of GDAP2 is the more divergent. Specifically and in contrast to macroH2A1.2 and macroH2A2, GDAP2 has a proline in the D-rich loop, which probably makes it rigid and limits flexibility of the binding pocket affecting its capacity to close as part of an induced fit. MacroH2A1.2 has a charged and the most flexible binding pocket. While macroH2A2 has a flexible D-rich loop, a proline-rich region on the side opposing the D-loop potentially limits the flexibility of the binding-pocket. We speculate that these features might be important for defining the ligand specificity of non-ADP-ribose binding macrodomains. Based on our analysis, we were able to predict which mutations in the binding pocket are likely to abolish the binding capacity for ligands. Such mutants will be instrumental for the future validation of phospholipids as potential macroH2A1.2 ligands. Indeed, two potential configurations suggested by docking phospholipids into the binding pocket involve key interactions with the side chains in the D-loop and R318 on the

opposing side of the binding pocket. Furthermore, we are confident that the identified features will also be informative for the future deorphanization of the macrodomains of macroH2A2 and GDAP2.

4.2.2 Identification of phospholipids as putative natural ligands of macroH2A1.2

As detailed in the results section, we have obtained the most promising results pointing towards the identification of macroH2A1.2 ligands from an untargeted metabolomics approach that combined affinity-based enrichment on immobilized baits with simultaneous monitoring of affinity-based depletion from the input material. As a complex and rich source of natural metabolites, we used metabolic extracts from proliferating P19 cells. It has been previously postulated that the majority of the cell types have significant levels of common metabolites (Yanes et al., 2011). Having this in mind, we used the proliferating P19 cells as a cellular source of metabolites to validate and establish the untargeted metabolomics approach. We considered the possibility that the metabolite of interest could reach physiologically relevant concentration only upon neuronal cell differentiation coinciding with the pronounced upregulation of macroH2A1.2 expression. As such we were initially planning to repeat the experiment also with an extract from differentiated cells in case the first approach would not have delivered any interesting results. A major limitation was the unknown chemical nature of the macroH2A1.2 ligand. To maximize our possibilities to extract and detect the metabolite of interest independently of its chemical properties, we combined two extraction solvents with different polarity with two columns of increasing polarity for the liquid chromatography preceding the mass spectrometric analysis. Following the fragmentation and identification of the macroH2A1.2 specific peaks by mass spectrometry, the results strongly indicated that using this method we were able to detect metabolites specifically interacting with macroH2A1.2 but not with macroH2A1.1 macrodomain or GST, a tag used for the protein purification and immobilization. The annotated peaks assigned to the natural compounds corresponded to phospholipids. More specifically, we identified a phosphatidylcholine and phosphatidylethanolamine, although the

exact identity of the fatty acid chains could not be ascertained. Additionally, the analysis uncovered the peaks of several other molecules with similar weight and properties to PE(35:1) and PC(34:2). However, these peaks could not be annotated due to the lack of available standards in the metabolite databases but are compatible with different ionization states of detected and related phospholipids. Furthermore, we did not detect any other lipids to specifically interact with macroH2A1.2, indicating interaction with the phospholipids occurs via the polar head and not the fatty acid chains. The docking further confirmed the plausibility of the polar head of these phospholipids to bind inside of the binding pocket of macroH2A1.2, specifically by interacting with the D-loop and R318. Taken all these observations together, we postulate that there is a high probability that phospholipids are true natural ligands of the macroH2A1.2 macrodomain.

4.2.3 The implication of phospholipids in the regulation of membrane structure and function and potentially related functions of macroH2A

Phospholipids are a class of lipids characterized by their specific physicochemical properties. The glycerol backbone of the phospholipids carries two chemically diverse fatty acid chains, differing in chain length and number and position of double bonds and hydroxylation. Their name arises from the characteristic head group which contains the phosphate group. The head group substituents define particular subgroups of phospholipids, such as phosphatidylinositol (PI), phosphatidylethanolamine (PE) or phosphatidylcholine (PC) (Harayama & Riezman, 2018). As the key components of membrane lipid bilayers, they contribute to the membrane fluidity and consequently affect various cell processes (Harayama & Riezman, 2018). This is especially relevant for all processes involving the fission and fusion of membranes or the function of membrane embedded proteins. The latter is illustrated by the impaired function of ion channels as a result of membrane remodeling in omega-3 fatty acid deprived *C. elegans* (Caires et al., 2017). Furthermore, impaired calcium

oscillations were observed in hippocampal neurons depleted of PPAR α , a transcription factor implicated in the regulation of fatty acid metabolism (Roy et al., 2016). While the authors have attributed this to the deregulation of the expression of genes encoding for ion channels, it would be interesting to assess if a change in membrane composition contributes to the observed phenotype.

Furthermore, optimal composition and fluidity of cell membranes is necessary for normal fusion and fission processes of cells, organelles and vesicles (Harayama & Riezman, 2018; Lenzi et al., 1996; Zick et al., 2014). Excitingly, a number of phenotypes caused by the loss of total macroH2A1 or the macroH2A1.2 isoform are potentially related with changes in membrane function. First, myogenesis involves the fusion of myoblasts to multinucleated myotubes (Perdiguero et al., 2009). Recently, our group reported a significant increase in myotube fusion upon the differentiation of macroH2A1.2-depleted myoblasts, while depletion of macroH2A1.1 had the opposite effect (Hurtado-Bagès, Posavec Marjanovic, et al., 2020). The significant increase in myotubes with more than 50 nuclei was specifically observed in macroH2A1.2-depleted cells. This phenotype correlated with the upregulation of a subset of genes involved in cell adhesion. Here, we further analyzed this transcriptomic dataset with the focus on the genes specifically affected by macroH2A1.2 depletion. Interestingly, we found a significant downregulation of the genes involved in lipid metabolism. Indeed, eight of the top ten most enriched Gene Ontology terms related to various aspects of lipid metabolism. We hypothesize that the increase in myotube fusion in the absence of macroH2A1.2 might be a result of the combination of (I) increased expression of genes involved in adhesion and (II) altered membrane fluidity as a result of the perturbed lipid metabolism. In its future work the group will analyze data sets from other cell types to see if the role of macroH2A1.2 in the transcriptional regulation of genes related with lipid metabolism is more general. Interestingly, an isoform-specific loss-of-function data set is available in the breast cancer cell line (Dardenne et al., 2012). Furthermore, a number of transcriptomic data sets have been generated after total macroH2A1 knock-down in cancer cells such as HepG2 cells (Lo Re et al., 2018). These data sets can also be used to analyze macroH2A1.2-dependent effects as many of these fast proliferating cancer cells express very little macroH2A1.1 (Novikov et al., 2011).

Another important cell-cell fusion event is the fusion of gametes during fertilization. The plasma membrane of the sperm plays an important role in sperm fertilization capacity. The sperm membrane significantly rearranges due to the active lipid metabolism as spermatozoa passes through epididymis, resulting in increased membrane fluidity (Lenzi et al., 1996). Thus, if macroH2A1.2 took part in the regulation of lipid metabolism during gametogenesis, it would be plausible that it affects composition and fluidity of gamete membranes. This could provide a potential explanation for the observed non-Mendelian distribution (1:3.5:1) of offspring genotypes from heterozygous breeding pairs. Finally, behavioral phenotypes are frequently related with changes in neuronal function. As such it will be interesting if the behavioral phenotype of macroH2A1.2 KO mice is related to a change in synaptic vesicle fusion and neuronal circuit activity.

Taken together, it is plausible that phospholipid function and the poorly understood loss of function phenotypes of macroH2A1.2 converge on the level of membrane functions. Future work in the lab will test this intriguing hypothesis.

4.2.4 The role of phospholipids in gene regulation, cell specification and behavior and possible links through direct binding to macroH2A1.2

Although the phospholipids are best known for their role as membrane building blocks, they are also well recognized for their role in signaling and gene regulation. As such, phospholipids can be transported to, synthesized or modified in the nucleus where they can affect transcriptional regulation directly through binding of nuclear receptors, or in a more indirect manner by affecting chromatin regulation (Crowder et al., 2017; Musille et al., 2013). The latter includes interaction of phosphatidylinositol-phosphates (PIPs) with various components of chromatin remodeling enzymes such SWI/SNF complexes and the regulation of HDAC activity (Crowder et al., 2017; Toska et al., 2012). On the other hand, the interaction of a variety of different phospholipids with nuclear receptors, such as peroxisome proliferator-activated receptor (PPAR) family of transcription factors, NR5A or LXR, directly takes part in the regulation of the gene expression. PPAR α

is a transcription factor implicated in the regulation of lipid metabolism (Montaigne et al., 2021). The transcriptional regulation by PPAR α depends on the availability of its physiologically relevant ligands, reported to be different phospholipids. Furthermore, the enzymatic activity of fatty acid synthase (FAS) is necessary for the activation of PPAR α both in liver and in brain. However, PPAR α seems to bind different ligands and to regulate gene expression in a tissue-specific manner. More specifically, phosphatidylcholine PC(34:1) was identified as first endogenous ligand in liver that induced PPAR α -mediated gene expression counteracting liver steatosis (M. V. Chakravarthy et al., 2009). In hippocampus, a different set of endogenous ligands controls PPAR α activity. These are 3-hydroxy-(2,2)-dimethyl butyrate, hexadecanamide and 9-octadecenamide that increase synaptic function of hippocampal neurons (Roy et al., 2016). Interestingly, hypothalamus-specific FAS knock-out (FASKO) mice show impaired feeding behavior resulting in decreased food intake, reduction of adipose tissue content and increased physical activity as a consequence of impaired PPAR α activity (M. V. Chakravarthy et al., 2007). In addition, these animals showed decreased anxiety-like behavior that was reflected in an increased number of entries, time spent and distance traveled in the center of an open field. The phenotypes of FASKO mice were successfully rescued by pharmacological activation of PPAR α , indicating the relevance of endogenous ligand binding to PPAR α in hypothalamus for regulating animal behavior (M. V. Chakravarthy et al., 2007).

With respect to our study, the key phenotype of the whole-body knock-out of macroH2A1.2 was also related to anxiety-like behavior. As this might be caused by neuronal changes, we will need to perform more studies to better understand how the macroH2A1.2 knock-out impacts on gene expression, protein interactome and lipidome in brain regions controlling behavior. More specifically, we observed a significant decrease in the reactivity of macroH2A1.2 deficient mice to acoustic stimuli. This phenotype was observed in male and female mice and was strictly isoform-specific, as it was not observed in macroH2A1.1 knock-out mice. Since decreased auditory brain stem response is observed only in female animals while males reacted normally, we postulate this is macroH2A1.2 specific phenotype not occurring due to the deafness of the animals. Moreover,

the acoustic startle response was assessed in young animals at 10 w/o, whereas the hearing disability in the females was observed in animals at 17 w/o. Furthermore, the animals had a normal response to the prepulse, which had lower sound pressure and same frequency as the louder startle. Looking at the reactivity of the animals, we can observe a certain threshold in the reactivity, as if their ability to react is much lower. In addition, we observe a slightly increased time spent in the center of the open field which indicates decreased anxiety, but could also indicate higher risk-taking behavior. We will need to perform additional behavior-focused tests to further explore the observed phenotype. Interestingly, recently another group reported an autism-like behavior in total macroH2A1 knock-out mice that they have related to macroH2A1.2 (Ma et al., 2021). Focusing on the mouse neocortex, the authors further report a switch from differentiation to increased proliferation after knockdown of macroH2A1.2 *in vivo*. Since we observed much higher expression of macroH2A1.2 in the hippocampus and the cerebellum in the brains of adult mice, we plan to focus our efforts in exploring the role of macroH2A1.2 there.

In addition to the behavioral phenotype and lack of response to auditory stimuli, macroH2A1.2 knock-out mice displayed a number of phenotypes that were to some extent sex-dependent. This included a lower body weight and mild metabolic phenotypes such as decreased carbohydrate and increased fat substrate oxidation during the active phase that in male mice was accompanied by slightly lower fasting triglyceride and total cholesterol levels (Annex 8.6 and results not shown). The pathology analysis further indicated tissue alterations in the liver of female macroH2A1.2 knock-out mice that were hepatitis, steatosis or immune infiltrates in 3/5 cases, with no differences observed in the male mice (results not shown). This is in agreement with previously published study in macroH2A1 total knock-out mice which showed a female-specific liver fat accumulation resulting in hepatic steatosis (Boulard et al., 2010). Since the livers of macroH2A1.1 knock-out mice were normal (results not shown), we now have an indication that the liver changes might be a consequence of the lack of the macroH2A1.2 isoform. However, it remains to be seen, which of these phenotypes are binding-pocket dependent and thus potentially linked to the binding of a phospholipid.

We further hypothesize that similarly as observed for PPAR α gene regulation, differences in the concentration of metabolites in different tissues might result in a differential regulation of macroH2A1.2 function. However, PPAR α has a hydrophobic binding pocket accommodating the fatty acid chains in the interior, whereas the polar head interacts with polar outer surface (Kamata et al., 2020). This is similar to the interaction of lipid transfer proteins with their cargo, but opposite to what is expected for macroH2A1.2. The binding pocket of macroH2A1.2 does not seem to be big or hydrophobic enough to accommodate long lipid chains in their entirety, although this still needs to be tested. As discussed before, it is plausible that the putative interaction of macroH2A1.2 with phospholipids occurs mainly with the polar head. We could further speculate that the polar head of the phospholipid could be only the specific docking interaction mediating the interaction with another protein. Many unknowns remain to be elucidated, but we do speculate that macroH2A1.2 might be able to sense changes in lipid metabolism. The binding of phospholipids might mediate or affect the interaction with proteins related with transcriptional regulation or chromatin structure and as such cause changes in gene expression. Even if we assume low affinity of macroH2A1.2 for phospholipids, its high expression in the brain and the fact that brain has a high lipid content of 50% of its dry mass, mostly phospholipids (Bruce et al., 2017), reinforce the possibility that concentrations are high enough to mediate a dynamic regulation and to be relevant for macroH2A1.2 response.

Interestingly, *in vitro* studies show that phospholipids play an important role in fate specification of neural progenitors. More specifically, neural progenitor cells differentiate into neuronal lineage when exposed to phosphatidylcholine (PC), whereas treatment with phosphatidylethanolamine (PE) drives them preferentially towards an astroglial fate (Montaner et al., 2018). This is particularly interesting in the context of the putative binding of both phospholipid specimens to macroH2A1.2. Noteworthy, we observed an impaired differentiation of macroH2A1.2-depleted cells towards glial cells, whereas neuronal cell differentiation seemed to be unaffected. It remains to be elucidated if this is reflected in macroH2A1.2 KO mice. Furthermore, when examining brain slices only neurons expressed macroH2A1.2. If this holds true *in vitro*, macroH2A1.2

might affect the glial cell differentiation in an indirect manner. The normal function of nervous system requires efficient crosstalk between different cell types. As a consequence, neuronal synapses are a result of the dynamic communication between neurons and glial cells (Araque & Navarrete, 2010; Stevens, 2003). Neurons produce trophic factors which trigger the differentiation of glial cells. On the other hand, glial cells regulate survival of neurons. This mutual signaling occurs at various stages of the development of neuronal and glial networks (Lemke, 2001). Therefore, we hypothesize that macroH2A1.2 could have an unknown role in neurons, which could consequently affect the viability and/or survival of glial cells.

4.2.5 A potential lesson from a domain co-occurring with the orphan macrodomain in GDAP2

The complexity of lipid metabolism is reflected in the abundance of different lipid species whereby small chemical changes can result in thousands of different lipid molecules (Crowder et al., 2017), that are generated and distributed by a large number of enzymes and transporters expressed in a tissue specific-manner. Furthermore, the presence of relevant enzymes in different cell compartments indicates compartmentalization of lipid metabolism on an organellar and a suborganellar scale (Chiapparino et al., 2016; Harayama & Riezman, 2018). As such, lipid-transfer proteins represent one of the key components of lipid metabolic pathways as they enable cross-organellar lipid transfer to ensure optimal sensing of intracellular and environmental changes (Chiapparino et al., 2016; Crowder et al., 2017).

Since SEC14-homology domain takes part in intracellular trafficking of phospholipids, SEC14 containing proteins are regarded as *bona fide* lipid transport proteins (Saito et al., 2007). Some transfer proteins are accompanied by additional domains which take part in regulating their interactions with other proteins, specify their subcellular location or provide additional signaling, sensing or enzymatic activities (Chiapparino et al., 2016). Interestingly, GDAP2 has a N-terminal macrodomain and a C-terminal SEC14 domain and is thus considered

to be a lipid transfer protein. GDAP2 is highly expressed in the brain (Eidhof et al., 2018; Liu et al., 1999). This is particularly true for Purkinje cells of the cerebellum that are also highly enriched in macroH2A1.2 (**Figure 36B**). Like macroH2A1.2, GDAP2 does not bind ADP-ribose and a possible interaction with poly(A) has been suggested but not confirmed (Neuvonen & Ahola, 2009). Intriguingly, the initial hypothesis that led to the identification of ADP-ribose and related metabolites as interactors of the macrodomains, was based on the observation that macrodomains occasionally occurred in enzymes generating ADP-ribose as a side-product of their NAD⁺ consuming activity such as Sir2 gene products in fungal and microbial species and some PARP family members (Karras et al., 2005; Kustatscher et al., 2005). As discussed above, the GDAP2 macrodomain shares a similarity in the D-rich loop of the binding pocket with macroH2A1.2, especially with respect to the aspartate predicted to stabilize the phospholipid polar head within the macroH2A1.2 binding pocket. Although the rigidity of the binding pockets significantly differs between the three orphan macrodomains, the D-rich loop on the lower part of the binding pocket remains the common denominator (**Figure 28D**). Thus, it is a compelling fact that all three orphan macrodomain proteins show significant expression in the brain, whose lipid content accounts for the 50% of its dry mass, mostly phospholipids (Bruce et al., 2017). Although our preliminary results indicate that phospholipids might be the physiological ligands of macroH2A1.2, it remains to be seen if all three orphan macrodomains bind ligands with similar or related chemical nature.

5 Conclusions

5.1 Conclusions

In the first study, we aimed to investigate the evolution of macroH2A histone variants and clarify the implication of macroH2A1.1 isoform in the regulation of nuclear NAD metabolism from its evolutionary origin to the vertebrates.

We arrived to the following conclusions:

1. MacroH2A histone variants first appeared in protists. The ancestral macroH2A was the most similar to macroH2A1.1. The whole-genome duplication in the last common ancestor of vertebrates resulted in the appearance of macroH2A2, followed by the appearance of macroH2A1.2.
2. The ancestral macroH2A has a high structural conservation. It binds ADP-ribose stronger and inhibits PARP1 more potently than the vertebrate macroH2A1.1, consequently increasing the mitochondrial respiration more effectively. Thus, macroH2A1.1 is implicated in the regulation of NAD compartmentalization from unicellular organisms to humans.
3. MacroH2A in *Capsaspora* might be necessary for the inhibition of nuclear NAD⁺ consumption in the spore-like stage to enable its survival in poor environmental conditions, through indirectly maintaining the sufficient NAD concentrations for life sustaining processes.

In the second study, we investigated the biophysical, biochemical and physiological aspects of macroH2A1.2 isoform with the goal to find the physiological ligand and the relevant model for the future study of the physiological role of this isoform.

We derive the following conclusions:

1. MacroH2A1.2 macrodomain shows structural hallmarks which might be relevant for the accommodation of ligands. These hallmarks help defining the chemical properties of the putative ligands.
2. Our untargeted approach for the detection of the low affinity ligands interacting with orphan proteins of unknown function via polar interactions, successfully detects phospholipids as putative macroH2A1.2-specific physiological ligands.
3. MacroH2A1.2 deletion in mice results causes significant behavioural changes and mild metabolic effects which might be a consequence of the implication of macroH2A1.2 in the regulation of lipid metabolism.

7 Materials and Methods

7.1 Computational analyses

7.1.1 Molecular data mining

MacroH2A sequences were collected from GenBank database by Blast searches using human sequences as a query. For better representation of species, especially in the transition to the vertebrate lineage, de novo assembly of transcriptomes for jawless fish (hagfish and lamprey; bioproject acc. numbers: PRJDB4902 and PRJNA292033, respectively) and the bowfin (bioproject acc. number: PRJNA292033) were carried out using Trinity software, version 2.2.0 in the Galaxy web platform (Afgan et al., 2016). Briefly, paired-end SRA Fastq files were uploaded from the European Nucleotide Archive (ENA) to the Galaxy platform and their quality analyzed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). All left and right reads were concatenated in two separate files and used as input for Trinity with default parameters. Assembled sequences were used to create local nucleotide databases and Blast searches were performed as described above.

Overall, 467 sequences encompassing 327 metazoan (58 vertebrate and 269 non-vertebrate) and 3 non-metazoan species were retrieved. Three macroH2A variants (macroH2A1.1, macroH2A1.2 and macroH2A2) were collected for all vertebrate species except for species displaying only macroH2A1.1. For all macroH2A proteins, only the globular part of the macrodomain (amino acids 182 to the end in human sequences) was used in the analyses, unless stated otherwise. Manual Blast searches were performed for curating the data of underrepresented species, specifically for filasterean and ichthyosporean species using available information (Brown et al., 2018; Dudin et al., 2019; Grau-Bové et al., 2017; Hehenberger et al., 2017; Richter et al., 2020; Torruella et al., 2015).

7.1.2 Sequence alignments

Multiple sequence alignments were performed using MAFFT version 7 (Katoh & Standley, 2013) and Jalview version 2 (Waterhouse et al., 2009), and edited for potential errors in BioEdit (version 7.2). Logo plots were generated based on the aligned sequences using WebLogo3 (Crooks et al., 2004). Multiple sequence

alignments of the macroH2A1.1-like macrodomains sequences from 305 species with one macroH2A1.1-like isoform was generated using the alignment tool PRALINE (Bawono & Heringa, 2014). Conservation score indicates the conservation of amino acid biochemical properties on the scale from 1 to 6 and is represented in height and the color of the bar.

For pairwise sequence homology analysis, we used blast homology search and macroH2A macrodomain sequences of representative organisms as a query. The resulting percent homology rate is represented in the homology matrix visualized using Morpheus matrix analysis and visualization software (software.broadinstitute.org/morpheus/).

7.1.3 Phylogenetic and evolutionary analyses

If not stated otherwise, phylogenetic and molecular evolutionary analyses were conducted using MEGA version 7 (Kumar et al., 2016). MacroH2A phylogeny was inferred by using the Maximum Likelihood method with the LG substitution model (Le & Gascuel, 2008), and including gamma-distributed variation among sites. Positions with less than 95% of site coverage were eliminated, and thus the analysis involved 121 amino acid sequences and a total of 156 positions in the final dataset. The reliability of the reconstructed topology was contrasted by nonparametric bootstrap (BS) method (1000 replicates).

Nucleotide and protein sequence divergence was estimated using uncorrected differences (p-distances, partial deletion 95%) and the rates of evolution were estimated by correlating pairwise protein divergences between pairs of taxa with their corresponding divergence times as defined by the TimeTree database (Hedges et al., 2006).

Additionally, the number of synonymous (pS) and nonsynonymous (pN) nucleotide differences per site were computed using the modified method of Nei–Gojobori (Zhang et al., 1998) and estimating standard errors by using the BS method (1000 replicates). pS and pN nucleotide differences per site were compared using codon-based Z-tests for selection, setting the alternative hypothesis as $H_A: pN < pS$. The amount of codon usage bias was investigated using the program DnaSP version 6 (Rozas et al., 2017).

7.1.4 Focused phylogenetic analysis of selected key residues

For the focused phylogenetic analysis of the residues at positions 225 and 316 (numbering according to *Capsaspora* macroH2A) the analysis was done on the 305 sequences of the species with one macroH2A1.1-like isoform. The residues were defined as characteristic of protists, vertebrates or other residues separately for the position 225 and the position 316. Then, we calculated the number of species with a particular group of residues on each of the positions in the protein. We discarded the phyla with three or less species per phyla for higher accuracy, with the exception of Filasterea. The percentage of species with either the residues characteristic of protists, vertebrates and other residues per phyla was calculated for both the position 225 and the 316.

For Mollusca, we grouped the species into those with one or with two alternatively spliced macroH2A isoforms. We then calculated the percentage of species per class with an alternative exon 7. Finally, within the species with two macroH2A isoforms as a result of alternative splicing of exon 7, we calculated the percentage of the residues characteristic of protists, vertebrates and other residues in all of the species, separately for exon 7.1 and 7.2.

7.1.5 Metabolism-focused analysis of high-content data from *Capsaspora*

Previously generated raw RNA sequencing data (Sebé-Pedrós et al., 2013) was re-aligned using STAR version 2.7.3a (Dobin et al., 2013). The function genomeGenerate was used to include the mitochondrial transcriptome data in the described assembly of the *Capsaspora* genome (Suga et al., 2013) and quantified using featureCounts software ver. 2.0.1 (Liao et al., 2014). Statistical analysis was performed using DESeq2 (Love et al., 2014) using Likelihood ratio test selecting genes that show an adjusted p-value equal or less than 0.01. Clusterization of the data was obtained using DEGreport software package on Bioconductor.

Functional annotation was performed using EGGnog 5.0 using precomputed clusters and phylogenies (Huerta-Cepas et al., 2019). The curated list of metabolic genes from KEGG (Kanehisa & Goto, 2000) was generated by

retrieving *Capsaspora* genes using EGGnog 5.0 and GHOSTKoala (Kanehisa et al., 2016). This was further complemented by using Orthofinder (Emms & Kelly, 2019) to identify *Capsaspora* orthologs of the human queries. The absence of orthologues was confirmed by Blast search. The resulting list has further been manually curated and genes involved in multiple pathways have been assigned to their parent metabolic pathway and categorized as anabolic, catabolic, context-dependent (anabolic or catabolic), other and unknown.

The relative abundance of proteins was extracted from previously generated proteomics data using raw measurements and averaging three replicates per condition (Sebé-Pedrós, Peña, et al., 2016).

7.1.6 Molecular docking and similarity search

For the docking of natural compounds, the binding pocket of human macroH2A1.2 macrodomain (PDB: 1ZR5) was explored by the program SEED to calculate the preferred positions of small probe molecules inside the binding pocket and PINGUI was used to choose for the energetically best poses. As a source of compounds were used the following databases: HMDB food, HMDB drug metabolites, HMDB plants and KEGG. The docking calculation was performed by DOCK, and the best scored 500 molecules were extracted and evaluated. The molecules were rated from A to D, with A for a very good interaction and D for an impossible pose. DAIM was used for the calculation of statistical significance of the enrichment of molecule fragments in the best scored docking poses.

Similarity searches were carried out using ZINC15 database as a source of molecules. A preselection of molecules was done using the online tranche browser and only those with a size of 200-350 Da and logP from -1 to 5 were considered. The calculations were done by KNIME 3.7.2. by loading the query molecule and the database Simplified Molecular Input Line Entry Specification (SMILES) input files into the workflow followed by ECFPlike fingerprint calculation by RDKit Fingerprint node. The similarity search was carried out by the Analytics Similarity Search node with Tanimoto and Dice coefficients for the molecule comparison with the cut-off set to 0.4 and 0.5, respectively. Both sets were combined and doubles were removed. The crystal structure of human

macroH2A1.2 macrodomain (PDB: 1ZR5) was prepared for docking using WitnotP 2.21. Protonation, assignment of atom types and bond orders and preparation of input files were carried out for the CHARMM minimizations (200 cycles), followed by the placement of ligands into the binding site using MOE 2019. Further minimizations were carried out using the Amber14:EHT force field as implemented in MOE 2019. Four different structures were prepared taking in the consideration the possible poses based on those predicted by docking of 3-OHR and ADP-ribose, while taking into consideration different residue rotamers and distance restrained minimizations. Protomers of the ligands were calculated for a pH value of 7.4 using the DOCK 3.7-2rc1 ligand preparation pipeline and the dockings were done by DOCK 3.7-2rc1. Grids and spheres were generated using the blastermaster.py script with default settings. While the placement of the low dielectric spheres was not modified, the matching spheres were moved and optimized using the shake.py script. The default number of low dielectric (120) and matching (45) spheres was used. For each docking, the 1000 best scored molecules were extracted and evaluated. Molecules that showed intra- or intermolecular clashes or strange conformations were energy minimized using the Amber14:EHT force field as implemented in MOE 2019. For more details see Master Thesis by Janik Bjorn Hedderich, Philipps-Universitat Marburg, 2019.

7.2 Approaches in biophysics and biochemistry

7.2.1 Plasmids

For plasmid construction, we used standard cloning techniques. Capsaspora macroH2A sequence (NCBI sequence ID: XM_004347479.1) was synthesized Life Technologies Inc. (Thermo Fisher) and cloned into pcDNA3.1 backbone with an N-terminal GFP-tag. Capsaspora (G224E, Q225E, N316R) and murine (G224E) macroH2A binding pocket mutants were generated via Stratagene's site-directed mutagenesis QuickChange protocol. Macrodomain sequences (corresponding to amino acids 182-368 for Capsaspora macroH2A and 182-369 for murine macroH2A1.1) were cloned into pETM-11 backbone with an N-terminal 6 His-tag. Capsaspora PARP1 (NCBI sequence ID: XP_004363957.1)

N-terminal sequence (corresponding to amino acids 1-350) was amplified from *Capsaspora* cDNA and cloned into pETM-11 backbone with an N-terminal 6 His-tag.

For mammalian expression constructs, the full-length sequences of *Capsaspora* macroH2A, or mouse macroH2A1.1 or macroH2A1.2 were cloned into pLVX-Puro lentiviral backbone (Clontech) adding an N-terminal EGFP-tag. Mouse-*Capsaspora* chimeras were generated by sequential cloning of fragments. First, histone fold and linker domain sequences from mouse macroH2A1.1 were inserted into the backbone, followed by the insertion of either wildtype or mutant (G224E) *Capsaspora* macroH2A macrodomain. pLVX-Puro with EGFP alone was cloned and kindly provided by Matthew Gamble (Ruiz & Gamble, 2018). All sequences were verified by sequencing.

7.2.2 Protein production and purification

Rosetta (DE3) chemically competent *E. coli* were transformed with bacterial expression vectors and grown in LB medium supplemented with 34 µg/mL chloramphenicol and 50µg/mL kanamycin at 37°C overnight. The culture was used to inoculate 1L terrific broth medium and grown at 37°C and 200rpm until reaching an optical density at 600 nm of 0.4-0.6. The protein expression was then induced with 0.5mM isopropyl-β-d-1-thiogalactopyranoside (IPTG) for 16h at 20°C. The next day, bacteria were pelleted by centrifugation at 10,000g for 15min at 4°C. The bacterial pellet was lysed in 50mM Tris, 300mM NaCl, 10mM imidazole, 1mM DTT (pH 7.4) supplemented with 1mg/mL lysozyme, 10 µg/mL DNase I and protease inhibitors (Roche cOMplete EDTA free) for His-tagged proteins, or in 50mM Tris, 150mM NaCl, 10mM EDTA, 1mM DTT (pH 7.4) supplemented with 1mg/mL lysozyme, 10 µg/mL DNase I and protease inhibitors (Roche cOMplete EDTA free) for GST-tagged proteins. The lysates were cleared by centrifugation at 30,000g for 45min at 4°C. Subsequently, the cleared lysates containing His-tagged proteins were incubated with Ni-NTA beads (Quiagen) for 1 hour at 4°C and passed over a gravity flow column 3 times. After washing the beads with 3 column volumes using 50mM Tris, 1M NaCl, 10mM imidazole, 1mM DTT (pH 7.4), the proteins of interest were eluted with 50mM Tris, 300mM NaCl, 300mM imidazole, 1mM DTT (pH 7.4). The cleared lysates

containing GST-tagged proteins were incubated with glutathione (GSH) beads (Pierce) for 4-6 hours at 4°C and passed over a gravity flow column 3 times. After washing the beads with 3 column volumes using 50mM Tris, 1M NaCl, 1mM DTT (pH 7.4), the proteins of interest were eluted with 50mM Tris, 10mM DTT, 15mM reduced glutathione (GSH). The eluted proteins were dialysed overnight into a phosphate buffer (50mM KH₂PO₄, 1mM DTT, pH 6.5), unless stated otherwise. The purified proteins were concentrated using 3kDa MWCO centrifugal concentrator (Amicon), and then flash frozen in liquid nitrogen and stored at -80°C.

7.2.3 Saturation Transfer Difference Nuclear Magnetic Resonance (STD-NMR)

Saturation Transfer Difference Nuclear Magnetic Resonance (STD-NMR) experiments were performed as described elsewhere (Aretz et al., 2018; Mayer & Meyer, 1999). Briefly, macrodomains were dialysed and prepared at 10µM in a deuterated water buffer (20mM NaP, 1mM TCEP, pH 7.4 and 10µM DSS). ADP-ribose was prepared in the deuterated water buffer, and 3-OHR and compound 4 (Annex) were dissolved in deuterated DMSO and diluted in the deuterated water buffer. STD spectra of 1mM compounds in the presence of 10µM macroH2A macrodomains were obtained at 25°C (298 K) on a Bruker Avance 600 MHz spectrometer equipped with a cryoprobe. A pseudo-2D version of the STD NMR sequence was used for the interleaved acquisition of on-resonance and off-resonance spectra. The on-resonance frequency was set to 0.0ppm and the saturation time was 3s. The STD effect (%) was quantified based on the following equation: $ISTD = 100 \cdot (I_0 - ISAT) / I_0$, where ISAT and I₀ are the intensities of a given signal in the on-resonance and the off-resonance spectra, respectively.

7.2.4 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was performed as previously described (Karras et al., 2005). Before the experiment, proteins were dialyzed overnight against 50mM KH₂PO₄, 1mM DTT (pH 6.5) at 4°C. The dialyzed proteins were then centrifuged for 20min at 20000 g at 4°C and the protein concentration was

determined by absorbance measurements at 280nm wavelength using calculated molar extinction coefficients. The nucleotides and ribose were prepared in the same buffer in the concentration range 1-1.5mM. The concentration of ADP-ribose was additionally confirmed by absorbance measurement at 260nm, using a molar extinction coefficient of 13,500 M⁻¹ cm⁻¹. For the ITC experiments in the presence of DMSO, the proteins were dialyzed into ITC buffer (50mM KH₂PO₄, pH 6.5, 1mM DTT, 1% DMSO). The compounds were resuspended in DMSO and diluted by ITC buffer. Final DMSO concentration in the experiment was 1,01% and the experiment was performed with 65μM macroH2A1.1 macrodomain and 1.4mM ADPR or with 60μM macroH2A1.2 macrodomain and 2mM of 3-OHR. Assays were conducted on the PEAQ-ITC instrument (MicroCal) at 25°C, and experimental data analysis was performed with MicroCal PEAQ-ITC Analysis Software.

7.2.5 Thermal shift assays

Fluorescence-based thermal shift assays were performed as previously described (Singh et al., 2017). Briefly, 5μM protein solutions supplemented with 8 x SYPRO Orange were heated in 50mM KH₂PO₄ (pH 6.5) and 1mM DTT from 5°C to 95°C at a ramp rate of 1%. The assays were conducted in MicroAmp Fast 96-well reaction plates sealed with MicroAmp Optical Adhesive Film (Applied Biosystems). The fluorescence measurements at 554nm were normalized to the lowest value before the transition and the maximum fluorescence.

7.2.6 Crystallization, data collection and processing

All crystallization experiments were conducted at the Crystallization Facility of the Max-Planck-Institute for Biochemistry (Martinsried, Germany). Before setting up the crystallization, the proteins were dialyzed overnight against 20mM Bis-Tris (pH 7.0) at 4°C. The dialyzed proteins were then centrifuged for 20min at 20000 g at 4°C and the protein concentration was determined by absorbance measurements at 280nm wavelength using calculated molar extinction coefficients.

The high-throughput crystallization screening of mouse macroH2A1.2 macrodomain was done in sitting drops vapor diffusion experiments performed

at 20°C by mixing the solution containing 2.5mM of 3-OHR and protein at 25mg/mL in distilled deionized water with the panel of crystallization buffers in 1:1 ratio. Crystals of *Capsaspora* macroH2A macrodomain in apo-form were obtained in sitting drops vapor diffusion experiments performed at 20°C by mixing 100nL 0.1M Bis-Tris (pH 5.5) and 25% (w/v) PEG3350 with 200nL solution containing the protein at 27mg/mL. Crystals of *Capsaspora* macroH2A macrodomain in complex with ADP-ribose were obtained in sitting drops vapor diffusion experiments performed at 20°C by mixing 100nL of 0.2M ammonium tartrate dibasic and 20% (w/v) PEG3350 with 200nL solution containing the protein at 21mg/mL and 3.8mM ADP-ribose. Crystals were cryoprotected by soaking in mother liquor supplemented with 30% ethylene glycol and flash cooled in liquid nitrogen. Diffraction data of proteins were collected on the Swiss Light Source or on the in-house X-ray source of the Crystallization Facility of the Max-Planck-Institute for Biochemistry (Martinsried, Germany). All datasets were processed using XDS (Kabsch, 2010).

The structures were solved by molecular replacement using the human macroH2A1.1 macrodomain (PDB 3IID, Timinszky et al., 2009) as a search model. Model building and real space refinement were performed in COOT (P. Emsley et al., 2010; Paul Emsley & Cowtan, 2004) and the structures refined using PHENIX REFINER (Afonine et al., 2012). Model and restraints for ADP-ribose was prepared using Phenix.Elbow (Moriarty et al., 2009). UCSF Chimera software (Pettersen et al., 2004) and the PyMOL Molecular Graphics System (Schrödinger, LLC) have been used for visualization.

The LigPlot diagram for the crystal structure of *Capsaspora* macroH2A macrodomain in complex with ADP-ribose was generated using an online platform LIGPLOT v.4.5.3 (Wallace et al., 1995). ConSurf bioinformatic tool (Ashkenazy et al., 2016) was used for the projection of evolutionary conservation scores. Briefly, the conservation analysis of amino acid positions was calculated based on the phylogenetic relations between sequences of 305 species with 1 macroH2A1.1-like isoform and using *Capsaspora* apo macroH2A macrodomain structure as a query. UCSF Chimera software has been used for visualization (Pettersen et al., 2004).

Visualization and comparison of the crystal structures of macroH2A1.2 (PDB 1ZR5), macroH2A2 (PDB 6FY5) and GDAP2 (PDB 4UML) and *in silico* mutagenesis of macroH2A1.2 were performed using UCSF Chimera software (Pettersen et al., 2004).

7.2.7 PARP1 activity assay

In vitro PARP-1 activity was measured using auto-PARylation assay as previously described (Posavec-Marjanović et al., 2017). Briefly, 0.3x activated DNA (diluted from 10x activated DNA, Trevigen) and 200 μ M NAD⁺ were added to the buffer containing 50 mM Tris-HCl (pH 8), 50 mM NaCl and 1 mM MgCl₂. Subsequently, purified macrodomains were added in the reaction mix in the defined concentrations (10 and 50 μ M for mouse, and 10, 25 and 50 μ M for *Capsaspora* macroH2A macrodomains). Finally, 0.2 units/ μ L PARP-1 HSA enzyme (Trevigen) was added to all reactions and the reaction mix was incubated at 25°C for 20min. Reactions were stopped by the addition of Laemmli's sample buffer and boiling on 95°C, separated on SDS-PAGE and analyzed by immunoblotting.

7.2.8 Antibodies

We generated specific antibody against *Capsaspora* proteins by immunizing rabbits with purified His-tagged *Capsaspora* macroH2A macrodomain or carrier protein-coupled peptides of *Capsaspora* PARP1. Specifically, we have used a mix of 3 different peptides corresponding to amino acids 103-114, 132-142 and 299-310 of *Capsaspora* PARP1 protein. Sera were collected from terminal bleeds after three to four rounds of inoculation. The animal procedures were carried out by the CID-CSIC Antibody Generation Service (Spain) and UVic-animal care facility (Canada).

7.3 Metabolomics

7.3.1 ADP-ribose pull-down assay

GST-tagged proteins were produced as described and purified by immobilization on the GSH beads for 4 hours at 4°C. The proteins immobilized on the beads were washed three times with 1x PBS, 1mM DTT buffer. ADP-ribose was prepared at 150µM concentration in the same buffer. The ADP-ribose solution was precleared with immobilized GST by mixing 150µM GST (420µL beads with immobilized GST) with 1mL of 150µM ADP-ribose solution. The immobilized GST and ADP-ribose were incubated on a rotator (Stuart SB2) for 30min on room temperature, followed by the centrifugation at 21,000g for 15min at 4°C. The supernatant was collected and referred to pre-cleared ADP-ribose solution.

Pre-cleared ADP-ribose solution was mixed either with immobilized GST, GST-macroH2A1.1 macrodomain or GST-macroH2A1.2 macrodomain. The total amount of proteins in the reaction amounted to 30nmol as calculated by the GSH-bead affinity and verified by Coomassie staining. The ADP-ribose solution was incubated with immobilized proteins on a rotator (Stuart SB2) for 1h on room temperature, followed by the centrifugation at 5,000g for 5min at 4°C. The supernatant carefully collected and used for the absorbance measurement as the “depleted SN”. 100µL of the buffer was added to the beads with immobilized proteins in the pellet and were transferred to 95°C for 10min followed by the centrifugation at 21,000g for 15min at 4°C. The supernatant carefully collected and used for the absorbance measurement as the “enriched SN”.

The ADP-ribose in the solution was determined by absorbance measurement on Nanodrop at 260nm wavelength. The samples measured were: input ADP-ribose solution, pre-cleared ADP-ribose solution, depleted SN either by the beads alone, GST or macroH2A1.1 or macroH2A1.2 macrodomains and enriched SN either in the beads alone, GST or macroH2A1.1 or macroH2A1.2 macrodomains. The 4 measurements per sample in each experiment was performed. Experiments were repeated two times. The average of value of GST and macroH2A1.2 absorbance set to 1 in the depletion experiments and the results were normalized accordingly. For the enrichment experiment the average absorbance of GST and

macroH2A1.2 was set to 0 in the depletion experiments and the results were normalized accordingly.

7.3.2 Untargeted metabolomics for ligand detection

7.3.2.1 Preparation of metabolite extracts

P19 cells were routinely cultured as described. It was ensured cells are in culture for more than 24h before collecting them to avoid for the stress induced metabolome. At the point of collection, they were at ~70% confluence. Prior to collecting the cells, one P10 plate seeded at the same density as the cells for the metabolite extraction was trypsinized and cells were counted. Cells for metabolite extraction were collected on ice. They were quickly washed once with cell culture grade ice cold PBS, followed by scraping with the addition of 200 μ L ice cold PBS and centrifugation at 500g for 5min at 4°C. PBS was removed and cell pellet was snap-frozen in liquid nitrogen.

For the metabolite extraction cells were resuspended in the way to obtain 1x10⁶ cells per 1mL. For the extraction with the organic solvent, we resuspended the cells with acetonitrile, methanol and water solution in the ratio 4:4:2 and vortexed gently to avoid leftover of cells on the walls of the tube and without touching the solvent and immediately snap-froze them in liquid nitrogen in an Eppendorf tube floating rack. We then started the sonication freeze-thaw cycle which was repeated 4 times in this order: transfer floating rack with samples to the sonication bath and sonicate until the solvent is thawed (~30sec), vortex for 10sec keeping in mind the material is in the solvent and transfer to liquid nitrogen until completely frozen, at which point start the cycle again with the sonication. After the freeze-thaw sonication cycle was finished, the samples were transferred to -20°C for 1h to allow for the precipitation of the proteins, followed by the centrifugation at 21,000g for 15min at 4°C. The supernatant containing metabolite extract was transferred to a lyophilization tube and lyophilized over-night. The lyophilized metabolite extract was resuspended in cell culture grade PBS and aliquoted to avoid freeze thaw cycles. For the metabolite extraction with polar solvent, the cells were resuspended in cell culture grade ice cold PBS to obtain an extract from 1x10⁶ cells/1mL, followed by vortexing for 30sec after which the tubes were transferred into boiling water and incubated there for 2min and then directly to

liquid nitrogen to snap freeze. The sonication freeze-thaw cycle was repeated 4 times as described for organic metabolite extraction and the cells were then kept on ice for 30min followed by the centrifugation at 21,000g for 15min at 4°C. The supernatant containing metabolite extract was transferred to 3kDa centrifugal filter and centrifuged at 4,000g and 4°C until the whole supernatant has been filtered through the membrane. The filtered supernatant was aliquoted and snap frozen.

7.3.2.2 Protein immobilization

GST-tagged proteins were produced and bacterial lysates were obtained and cleared as described. The immobilization to the GSH beads was done during 10 hours at 4°C, with the prior addition of 1mM PMSF and protease inhibitors (Roche cOMplete EDTA free) to the protein lysate. The proteins immobilized on the beads were washed twice with 50mM Tris, 150mM NaCl, 10mM EDTA, 1mM DTT, followed by washing with 50mM Tris, 2M NaCl, 1mM DTT for 20min at 4°C, and subsequent washing with 2 column volumes of 50mM Tris, 150mM NaCl, 10mM EDTA, 1mM DTT and then 2 column volumes washes with PBS and the proteins were checked by Coomassie staining.

7.3.2.3 Pre-clearing of metabolite extracts

After preparing the immobilized proteins, the metabolite extracts were precleared with GST, by mixing 3mL of metabolite extract with 800µL of beads saturated with GST on a rotator (Stuart SB2) for 30min on room temperature, followed by the centrifugation at 21,000g for 2min at 4°C. The obtained supernatant from ACN:MeOH:H₂O and PBS metabolite extracts were used as the inputs for the sequential depletion assay.

7.3.2.4 Sequential metabolite pull-down assay

Sequential depletion assay was performed with mixing 700µL of the pre-cleared metabolite mix with 300µL of beads with immobilized proteins (corresponding to ~100nmol of proteins). In the first step, the pre-cleared metabolite mix was incubated with either GST or GST-macroH2A1.1 macrodomains or GST-macroH2A1.2 macrodomain on a rotator (Stuart SB2) for 30min on room temperature, followed by the centrifugation at 5,000g for 5min at 4°C. The

supernatants were carefully collected and transferred to the tubes with corresponding fresh immobilized proteins to be incubated on a rotator for the second and the third round of metabolite depletions on a rotator, each for 30min on room temperature, followed by the centrifugation at 5,000g for 5min at 4°C. The supernatant after 3 rounds of depletion by the corresponding proteins is referred to as “depleted SN”. During the second and the third round of metabolite depletion, 200µL of PBS was added to the pellet remaining after the first depletion step was transferred to 95°C for 10min followed by the centrifugation at 21,000g for 15min at 4°C. The supernatant carefully collected and referred to as the “enriched SN”.

7.3.2.5 LC–MS and MS/MS analysis and data processing

The untargeted LC/MS data was generated using a UHPLC system (1200 series, Agilent Technologies) coupled to a G6550A ESI-qToF MS (Agilent Technologies) operated in positive electrospray ionization (ESI+). Metabolites were separated using two different chromatographic columns: a hydrophilic interactions HILIC column, InfinityLab Poroshell 120 HILIC-Z, 2.1 x 100 mm, 2.7 µm (PEEK lined) (Agilent Technologies); and a reverse phase C18 column, ACQUITY UPLC BEH C18 Column, 130Å, 1.7 µm, 2.1 mm X 150 mm, 1/pk (Waters). For the HILIC column, mobile phase A was water (50 mM ammonium acetate and 5µM medronic acid), and mobile phase B was acetonitrile. For the C18 BEH column, mobile phase A was water (0.1% of formic acid) and phase B acetonitrile (0.1% of formic acid). The separation of the extracts was conducted under the following gradients: 0-2 min 98% of B; 2-8 min decrease to 40% of B; 8-8.5 min raise to 98% of B; 8.5-11 min 98% of B for the HILIC system; and 0-0.5 min 0% of B; 0.5-2 min increase to 50% of B; 2-6 min raise to 100% of B; 6-7 min hold at 100% of B; 7-7.5 decrease to 0% of B and hold at 0% until 11 min in the C18 BEH column. The parameters for the electrospray ionization source (ESI) were as follow for the HILIC system: gas temperature, 200 °C; drying gas, 14 l min⁻¹; nebulizer, 35 psig; fragmentor, 175 V; and skimmer, 65 V. The parameters for the electrospray ionization source (ESI) were as follow for the C18 BEH system: gas temperature, 150 °C; drying gas, 11 l min⁻¹; nebulizer, 35 psig; fragmentor, 120 V; and skimmer, 65 V. The m/z acquisition range was set between 50 and 1500 Da,

acquiring 3 spectra/s. Column temperature was set to 25°C for the HILIC system and to 65°C for the C18 BEH system. To identify compounds, MS/MS data was generated in targeted mode and instrument was set to acquire spectra over the m/z range from 50 to 600, with a narrow width of 1.3 m/z. Several collision energies (CE) were tested between 10 and 40V.

LC-MS raw data files from either HILIC or C18 assays were converted to open standard format mzML using Proteowizard MS-convert (Chambers et al., 2012) and subsequently processed by XCMS software, version 3.10.2 (Smith et al., 2006). XCMS analysis of these data provided a matrix containing the retention time, m/z value, and integrated peak area for aligned features across samples (a feature being defined as molecular entity with a unique m/z and a specific retention time). Depleted and enriched features ($|FC|>5$) were prioritized to MS/MS experiments performed in targeted mode for metabolites identification. Metabolites were identified conforming to Level 2, as specified by Schymanski et al. since their accurate mass and experimental MS/MS spectra coincide with the fragmentation pattern of chemical standards from the METLIN, MassBank and/or NIST14 databases. All data processing was conducted in R version 3.6.1 (R-Foundation for statistical computing, www.Rproject.org)

7.4 Cell culture and gene expression analyses

7.4.1 Standard culture of human and *Capsaspora* cells

Unless stated otherwise, the following conditions were used for cell culture. MacroH2A-depleted HepG2 (DKD) cells (Douet et al., 2017) and HEK293T (ATCC, CRL-3216) were routinely cultured in Dulbecco's modified Eagle medium (DMEM) containing 4.5 g/L glucose (Gibco) supplemented with 10% v/v fetal bovine serum (FBS) (Gibco), 2 mM L-glutamine (Gibco), 50 U/ml penicillin (Gibco) and 50 mg/ml streptomycin (Gibco). P19 mouse embryonal carcinoma cells (ATCC; CRL-1825) were maintained in full growth medium consisting of Minimum Essential Medium Eagle (MEM) (Sigma-Aldrich; M8042) supplemented with 2 mM L-glutamine, 100 units/mL penicillin-streptomycin (P/S) and 10% fetal bovine serum (FBS). Cells were authenticated, incubated at 37°C in 5% CO₂ and

periodically checked for mycoplasma contamination. Cells were collected by scraping, washed with PBS and pelleted. *Capsaspora* was cultured, and cells from all three different stages were harvested as described in Seb e-Pedr s et al., 2013. Before analysis, *Capsaspora* cells were washed with PBS and flash-frozen in liquid nitrogen.

7.4.2 Induction of neuronal differentiation in P19 cells

To induce neuronal differentiation cells were treated with all-trans retinoic acid (RA) (Sigma-Aldrich; R2625). RA was dissolved in DMSO and stored in dark conditions at -20 C. Seed 3×10^6 cells in low adherent P10 plates in neuronal induction medium, NIM consisting of α MEM (Sigma Life Science; M4526) with 5% FBS, 100U/ml P/S and 10 M RA. The formation of cell aggregates (embryoid bodies, EBs) is observed after 2 days of culture in NIM. After 2 days of incubation the EBs were carefully collected to a tube and left to settle by gravity. Aggregates were transferred in a fresh NIM and incubated for 2 more days. After four days in NIM, EBs were collected, dissociated by trypsinization, and cultured on poly-L-lysine coated plates in neuronal differentiation medium (NDM), consisting of DMEM (Gibco; 21331-020) medium with 2% B-27 supplement 50X (Gibco; 17504-044), 8 mM HEPES and 100 units/mL P/S. After two days of culture in NDM, neural progenitor cells are formed which can be differentiated further into the mixed neuronal cell population. To support for the neuronal cell differentiation and eliminate the remaining proliferating cells, we added 10 M cytosine arabinoside (AraC) (Sigma Life Science; C6645) to NDM and left cells in culture for four more days. Ultimately, neurons were collected by scraping after 12 days of differentiation.

For PLL-coated plates, PLL (PLL, Sigma-Aldrich; P1399) was diluted in sterile water to a concentration of 20  g/mL and sterilised using a 0.22  m filter. For coating, cell culture-treated plates were incubated with PLL solution for 3 hours at room temperature. After incubation, plates were washed three times with phosphate-buffered saline (PBS), air dried and stored at +4 C.

7.4.3 Gene transduction and establishment of stable cell lines

HEK293T cells were used as packaging cells to produce viral particles for lentiviral infections. Four million HEK293T cells were seeded in P10 plates and cultured to 60-70% confluency. At that point the cells were transfected with 10 µg of the lentiviral plasmid of interest and 3 µg of the pCMV-VSV-G plasmid and pCMV-dR8.91. Plasmid DNA was mixed in 1x HBS solution (2x HBS: 272 mM NaCl, 2.8 mM Na₂HPO₄, 55mM HEPES, pH 7) containing 125mM CaCl₂ in a total volume of 800 µL. The mix was added onto the cell culture media dropwise and left overnight. Transfection efficiency was controlled using a GFP expression vector. The supernatant containing viral particles produced by HEK293T cells was collected for 24 hours at 48 hours and 72 hours after transfection, filtered using a 0.45µm filter and supplemented with 8µg/mL of polybrene (Sigma-Aldrich). The fresh viral supernatant was added to target cells at 60-70% confluency in six-well plates that were centrifuged 45 min at 1200rpm at 37°C, incubated at 37°C for 45min and then cultured overnight in fresh media. The same process was repeated 24h after the first infection. The transduced cells were selected with 2µg/ml puromycin. The necessary selection time was determined by using a negative control plasmid without resistance. This procedure was used to generate all HepG2 stable cell lines. The efficiency of the cell infection was validated by live cell fluorescence and flow cytometry.

7.4.4 Cell proliferation assays

Cells were seeded at a density of 4,000 cells/cm² and cultured in standard growth conditions during four days. Cells were collected, stained with Trypan blue and manually counted after 24, 48, 72 and 96 hours. This assay was performed in three independent experiments. Values are presented as means of all three experiments and their variability is indicated with the standard deviation (SD).

In the parallel approach, cells were seeded at 1.5x10⁴ cells/well in a standard 96-well microtiter plates. Cell viability was evaluated by methyl-thiazolyl-tetrazolium (MTT) assay at 24, 48, 72 and 96 hours. More specifically, growth media was replaced by media supplemented with MTT and cells were incubated for another 4 hours in the presence of MTT at 37°C. After the incubation period, the

developed formazan crystals were dissolved with DMSO followed by the absorbance measurement at 570nm using a Synergy H1 microplate reader (BioTek). The assay was performed in three independent experiments in eight replicates. Results were analysed by removing the outliers and normalising the data to the day 0. Values are presented as means of all three experiments and their variability is indicated with the standard deviation (SD).

7.4.5 Cell fractionation, immunoprecipitation and immunoblotting

For immunoblotting of total cell material, cell pellets were directly resuspended with Laemmli's sample buffer and sonicated using Bioruptor Plus (Diagenode) and incubated at 95°C for 10min before loading the samples on polyacrylamide gels.

For chromatin lysis, nuclei from mammalian cells were isolated and sonicated to solubilize chromatin, as previously described (Posavec-Marjanović et al., 2017). Briefly, cells were collected by scraping and lysed in sucrose buffer (0.32 M sucrose, 10 mM Tris-HCl, pH 8.0, 3 mM CaCl₂, 2 mM MgOAc, 0.1% Triton buffer, 1:100 PMSF and 1:200 leupeptin). Samples were passed through a syringe to facilitate membrane disruption and then further incubated for 8 min on ice. Intact nuclei were collected by centrifugation for 3 min at 1,000g and washed once more in sucrose buffer. Washed nuclei were resuspended in lysis buffer (50 mM Tris-HCl, pH 7–8, 135 mM NaCl, 0.1% Triton, 1 mM EDTA, 1 mM DTT, 1:100 PMSF and 1:200 leupeptin), and chromatin was solubilized by progressive sonication with a Bioruptor Plus (Diagenode).

For the co-immunoprecipitation of PARP1, the lysis buffer was complemented with PARG and PARP inhibitors (1µM ADP-HPD from CalBioChem and 1µM olaparib from SelleckChem, respectively). Insoluble material was removed by centrifugation, and lysates were pre-cleared with sepharose-beads. At this step, 5% of the total lysate was kept as input material, and the rest of the lysate was incubated for 3 h with anti-GFP nanobodies coupled to magnetic beads (ChromoTek) which were previously blocked with 1%BSA in lysis buffer. Precipitates were washed three times with lysis buffer containing 1% Triton X-

100. For SDS-PAGE and western blotting analysis, typically 1% input and 20% immunoprecipitated material was loaded.

For cell fractionations, nuclei were prepared as described above and the supernatant was kept as the cytosolic fraction. Nuclei were then incubated with high salt buffer (20 mM HEPES, pH 7.9, 410 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 25% glycerol, 0.5% NP-40) for 30 min. Ultracentrifugation at 50,000g was used to separate the chromatin (pellet) and nucleosol (supernatant).

Following the transfer of proteins to the nitrocellulose membrane (GE Healthcare), the membranes were blocked with 5% low fat milk (Nestle) and incubated with primary antibodies overnight at 4°C. The next day, membrane were washed with TBST and incubated with fluorophore-conjugated secondary antibodies for 1 hour on 25°C in dark. After washing with TBST, the dried membranes were scanned with Odyssey® CLx Imager.

7.4.6 RNA and DNA analysis

Total RNA from mammalian cells was isolated by the Maxwell RSC simplyRNA Cells Kit (Promega) using the Maxwell RSC Instrument (Promega) according to the manufacturer's instructions. 1µg of total RNA was used for cDNA synthesis using the First Strand cDNA synthesis kit (Thermo Scientific) using oligo(dT) primers and according to the manufacturer's instructions. Relative cDNA levels were quantified by RT-qPCR (LightCycler 480 II instrument, Roche). All samples were analyzed in technical triplicates. Values were normalized to two reference housekeeping genes (RPLP0 and GAPDH) and are plotted relative to a reference sample set to 1. To measure mitochondrial and genomic DNA, we extracted the total DNA from all cell lines of interest. Briefly, cells were pelleted and the DNA isolation buffer (10mM Tris HCl (pH 8.5), 5mM EDTA, 0,5% SDS, 200mM NaCl, 0.1mg/ml Proteinase K) was added directly on the pellets and the samples were incubated overnight at 37°C while shaking (Thermoblock). Proteinase K was inactivated by incubating the samples for 10min at 99°C. Equal volume of isopropanol was added to the lysates and left incubating for 20min at 25°C under constant shaking to precipitate the DNA. The precipitated DNA was pelleted by centrifuging at 10,000 rpm for 10min at 4°C. Supernatant was removed and precipitate was washed with ice-cold 70% ethanol and centrifugated at 10,000

rpm for 10min at 4°C. After removing the ethanol, the pellet was air-dried and resuspended in an appropriate volume of DNase-free water. The obtained DNA was used to perform quantitative PCR with oligos of mitochondrial (MT-TL1, MT-ND2) and genomic DNA (ACTB, NCOA3). Results were demonstrated as mitochondrial to genomic DNA ratio. The sequences of all primers used are given in the Supplementary Data File S7.

7.4.7 Analysis of mitochondrial oxidative phosphorylation

Mitochondrial respiration was monitored with the XFe96 Cell Bionalyzer (Seahorse Biosciences, XFe96). Optimal cell density and drug concentrations have been previously determined (Posavec-Marjanović et al., 2017). A standard MitoStress assay was performed. Briefly, 20000 cells were plated in an XFe-96 well-plate, and cells were kept for 6 h in DMEM 10% FBS to allow the cells to attach. Then, the medium was changed to 10mM glucose, 2mM glutamine, and 1mM pyruvate XFe DMEM (5mM Hepes, pH7.4), and cells were incubated for 1 h at 37°C without CO₂. Three different modulators of the mitochondrial respiration were sequentially injected. After basal oxygen consumption rate determination, 1.5µM oligomycin, which inhibits ATPase, was injected to determine the amount of oxygen dedicated to ATP production by mitochondria. To determine the maximal respiration rate or spare respiratory capacity, 1.5 µM carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (FCCP) was injected to free the gradient of H⁺ from the mitochondrial intermembrane space and thus to activate maximal respiration. Finally, 0.75µM antimycin A and 0.75µM rotenone were added to completely inhibit the mitochondrial respiration.

7.5 Experimental work on mouse tissues and in mice

7.5.1 Immunofluorescence of mouse brain sections

For the immunofluorescent staining of mouse brains, we used free-floating brain sections prepared from the perfused brains fixed with PFA. Free-floating brain sections were made at 40 µm thickness with freezing microtome (Leica CM 1950) and kept in tubes in the preservation buffer at -20°C. Prior to use, the sections were washed in PBS three times and then blocked for 1h on room temperature

with 10% normal goat serum (NGS) in 1xPBS with the addition of 0.5% Triton X. Following the blocking, the sections were incubated with primary antibodies diluted in blocking solution (NeuN 1:100; S100 1:500; mH2A1.2 1:200). The following day, the primary antibodies were removed and the slices were washed three times prior to the addition of the secondary antibody diluted in the blocking solution (1:500). The slices were incubated with the secondary antibody for 90min at room temperature, washed and incubated with DAPI for 10min, mounted and imaged.

7.5.2 Breeding and phenotyping of macroH2A1.2 knock-out mice

The Dre-recombinase expressing mouse line (Anastassiadis et al., 2009) in C57BL6/J background were received from Francis Stuart and Kostantinos Anastassiadis and genotyped using the primers listed in Annex, using the following PCR conditions: 95°C/5 min, followed by 35 cycles of 95°C/30 sec; 58°C/30 sec and 72°C/1 min to look for the 314bp product. Mating between the animals homozygous for Dre and heterozygous for macroH2A1 Rox locus ($Dre^{+/+}MACROH2A1^{Rox/+}$ x $Dre^{+/+}MACROH2A1^{Rox/+}$) was done to obtain knock-out ($Dre^{+/+}MACROH2A1^{Rox/Rox}$), wild type ($Dre^{+/+}MACROH2A1^{+/+}$) and heterozygous ($Dre^{+/+}MACROH2A1^{Rox/+}$) mice. The obtained mice were submitted to the German Mouse Clinic which performed as outlined here: <https://www.mouseclinic.de/screens/overview/index.html>. More specifically, German Mouse Clinic performed the phenotyping and analyzed the results on their designated specialized platforms for the study of behavior, bone and cartilage development, neurology, clinical chemistry, eye development, immunology, allergy, steroid metabolism, energy metabolism, lung function, vision and pain perception, molecular phenotyping, cardiovascular and pathology and took part in results interpretation.

7.6 Statistical analysis and figure editing

If not indicated otherwise, a two-tailed Student's t-test was used to assess statistical significance. The number of technical replicates or independent cell culture experiments is indicated in the relevant figure legend(s). Figures were edited with Inkscape (inkscape.org).

7.7 Antibodies table

| Antibody | Source | Cat. Number/Ref. | Application |
|--|-----------------------------|------------------|------------------|
| macroH2A1.1 | Home-made | Sporn, 2009 | Western Blot |
| macroH2A1.2 | Home-made | Sporn, 2009 | Western Blot |
| macroH2A1.2 | Cell Signaling | #4827 | Western Blot |
| macroH2A2 | Home-made | Buschbeck, 2009 | Western Blot |
| Histone H3 | Abcam | ab1791 | Western Blot |
| NeuN | Millipore | MAB377 | Western Blot |
| Map2 | Cell Signaling | #8707 | Western Blot |
| β-III tubulin | R&D Systems | MAB1195 | Western Blot |
| S-100 (β-subunit) | Sigma | S 2532 | IF, IHC |
| GFP | Santa Cruz Biotechnology | sc-9996 | Western Blot, IF |
| Poly(ADP-ribose) | Trevigen | 4336-APC-050 | Western Blot |
| Capsaspora macroH2A | Home-made | NA | Western Blot |
| Capsaspora PARP1 | Home-made | NA | Western Blot |
| IRDye 680RD Goat anti-Mouse IgG Secondary Antibody | LI-COR | 926-68070 | Western Blot |
| IRDye 680RD Goat anti-Rabbit IgG Secondary Antibody | LI-COR | 926-68071 | Western Blot |
| IRDye 800CW Goat anti-Mouse IgG Secondary Antibody | LI-COR | 926-32210 | Western Blot |
| IRDye 800CW Goat anti-Rabbit IgG Secondary Antibody | LI-COR | 926-32211 | Western Blot |

7.8 Primers table

| Application | Name | Sequence | Species |
|-------------|----------------------------------|--|---------|
| Cloning | macroH2A macrodomain Fw | ACTGAGAATCTTTATTTTCAGGGCTTTACGATC CTGTCCAAGAAG | Co |
| Cloning | macroH2A macrodomain Rv | AACCGCCGCCGCCATGGCGCCTCAGTTCGTG TTGATGAGCTCGGCG | Co |
| Cloning | macroH2A1.1 macrodomain Fw | ACTGAGAATCTTTATTTTCAGggcttcactgtcctctcc accaaga | Mm |
| Cloning | macroH2A1.1 macrodomain Rv | AACCGCCGCCGCCATGGCGCCctagttggcgtccag ctggccatt | Mm |
| Cloning | macroH2A macrodomain Fw | GGGGgatccAAGAAGCCTGTTCGCGGCTTTAC GA | Co |
| Cloning | macroH2A macrodomain Rv | GGGGgaattcTCAGTTCGTGTTGATGAGCTCGG CG | Co |
| Cloning | macroH2A1.1 macrodomain Fw | gggGAATTCggcacgcctacagacggcttc | Mm |
| Cloning | macroH2A1.1 macrodomain Rv | gggCTCGAGctagttggcgtccagcttggc | Mm |
| Cloning | PARP1_N- term Fw | GGGGgatccatgtctacgactttatcgctgaat | Co |
| Cloning | PARP1_N- term Rv | CCCCgaattcTCAaaagacaatcgatgctgtcg | Co |
| Cloning | GFP Fw | GGGGctcgagATGGTGAGCAAGGGCGAGGAGC TGT | - |
| Cloning | GFP Rv, with STOP codon | CCCCgaattcTCATTACTTGTACAGCTCGTCCA | - |
| Cloning | GFP Rv, without STOP codon | CCCCgaattcCTTGTACAGCTCGTCCATGCC | - |
| Cloning | macroH2A1.1 Fw | TACCGCGGGCCCCGGGATCCcgATGTCGAGCC GCGGCGGGAAG | Mm |
| Cloning | macroH2A1.1 Rv | ttatctagagtcgcgggatccCTAGTTGGCGTCCAGCTT GGC | Mm |

| | | | |
|--------------------|------------------------|---|-------|
| Cloning | wild type and G224E Fw | cgcagacagtagcagcggagggcacgcctacagacggct | Mm-Co |
| Cloning | wild type and G224E Fw | actgacacacattccacagggctcgacctagttggcgctccag | Mm-Co |
| Cloning | wild type and G224E Fw | CGCAGACAGTACGACGGAGAAGAAGCCTGTT CGCGGCTTT | Mm-Co |
| Cloning | wild type and G224E Rv | actgacacacattccacagggctcgacTCAGTTCGTGTTG ATGAGCT | Mm-Co |
| Mutagenesis | macroH2A1.1 G224E Fw | CTGACTTCTACACCGGTGAGGAAGTAGGAAA CACACTG | Mm |
| Mutagenesis | macroH2A1.1 G224E Rv | CAGTGTGTTTCCTACTTCCTCACCGGTGTAGA AGTCAG | Mm |
| Mutagenesis | macroH2A1.1 G224E Fw | ACCATGTCGTTTCGCGGAACAAGTCGGCGGTG CC | Co |
| Mutagenesis | macroH2A1.1 G224E Rv | GGCACCGCCGACTTGTTCCGCGAACGACATG GT | Co |
| Mutagenesis | macroH2A1.1 Q225E Fw | CCATGTCGTTTCGCGGGCgAAGTCGGCGGTGC CATC | Co |
| Mutagenesis | macroH2A1.1 Q225E Rv | GGTACAGCAAGCGCCCGcTTCAGCCGCCACG GTAG | Co |
| Mutagenesis | macroH2A1.1 G314E Fw | CCGTCGATCGGCTCTGAAAACAACCACTTCCC C | Co |
| Mutagenesis | macroH2A1.1 G314E Rv | GGGGAAGTGGTTGTTTTTCAGAGCCGATCGAC GG | Co |
| Mutagenesis | macroH2A1.1 N316R Fw | CCGTCGATCGGCTCTGGCcgCAACCACTTCCC CAAG | Co |
| Mutagenesis | macroH2A1.1 N316R Rv | GGCAGCTAGCCGAGACCGgcGTTGGTGAAGG GGTTC | Co |
| qPCR | Rplp0 Fw | catcaacgggtacaaacgag | Hs |
| qPCR | Rplp0 Rv | agatggatcagccaagaagg | Hs |
| qPCR | GAPDH Fw | CGACCACTTTGTCAAGCTCA | Hs |

| | | | |
|------|-----------|---|----|
| qPCR | GAPDH Rv | TCTTACTCCTTGGAGGCCAT | Hs |
| qPCR | DKK1 Fw | CCTTGAACCTCGGTTCTCAATTCC | Hs |
| qPCR | DKK1 Rv | CAATGGTCTGGTACTTATTCCCG | Hs |
| qPCR | mH2A1 Fw | GACGGTGAAAAACTGCTTGG | Mm |
| qPCR | mH2A1 Rv | GGAGGAGGACATCGTGGAG | Mm |
| qPCR | mH2A1 Fw | CCGCCGTCTGGAATACC | Hs |
| qPCR | mH2A1 Rv | GTTGTCTCTCGCTGCATTGC | Hs |
| qPCR | mH2A2 Fw | CATGGCGGCAGTCATTGAG | Hs |
| qPCR | mH2A2 Rv | ATTGCCGGCCAATTCTAGAA | Hs |
| qPCR | mH2A Fw | GGGGgatccAAGAAGCCTGTTTCGCGGCTTTAC GA | Co |
| qPCR | mH2A Rv | GGGGgaattcTCAGTTCGTGTTGATGAGCTCGG CG | Co |
| qPCR | PARP1 Fw | GGTACCATCCAGGCTGCTTT | Hs |
| qPCR | PARP1 Rv | CTGACTCGCACTGTACTIONGG | Hs |
| qPCR | PARP2 Fw | TGTTGTTGTTGAACTGGAGATTG | Hs |
| qPCR | PARP2 Rv | GGACCCAGAGTGACAGCCA | Hs |
| qPCR | NAMPT Fw | AATGTTCTCTTCACGGTGAAAA | Hs |
| qPCR | NAMPT Rv | ACTGTGATTGGATAACCAGGACT | Hs |
| qPCR | NMNAT1 Fw | TCTCCTTGCTTGTGGTTCATTC | Hs |
| qPCR | NMNAT1 Rv | TGACAACTGTGTACCTTCCTGTT | Hs |
| qPCR | NMNAT2 Fw | TGTCCACGACTCCTATGGAAA | Hs |
| qPCR | NMNAT2 Rv | GTCCGATCACAGGTGTCATGG | Hs |
| qPCR | NMNAT3 Fw | GAGTAGGTCACGACCCAAAAG | Hs |
| qPCR | NMNAT3 Rv | TCGCCTGATGTATGTGGCAC | Hs |

| | | | |
|------|-----------|---------------------------|----|
| qPCR | PARG Fw | AGCTTTTGGAAAGTGAACCTCA | Hs |
| qPCR | PARG Rv | ATCTTCCGTAGTCTGCTTTGC | Hs |
| qPCR | OARD1 Fw | CCAGCAGCCTTAATGAAGATCC | Hs |
| qPCR | OARD1 Rv | ACTGATACAGTGGGCTAAAGAGT | Hs |
| qPCR | SIRT1 Fw | TGTGTCATAGGTTAGGTGGTGA | Hs |
| qPCR | SIRT1 Rv | AGCCAATTCTTTTTGTGTTCTGTG | Hs |
| qPCR | SIRT6 Fw | CCCACGGAGTCTGGACCAT | Hs |
| qPCR | SIRT6 Rv | CTCTGCCAGTTTGTCCCTG | Hs |
| qPCR | NCOA3 Fw | TGCTGTCTCCATGTTTGATGTATCT | Hs |
| qPCR | NCOA3 Rv | TCTCTGCTCCCCACCTCTAAGT | Hs |
| qPCR | ACTB Fw | AGCGGGAATCGTGCGTGAC | Hs |
| qPCR | ACTB Rv | AGGCAGCTCGTAGCTCTTCTC | Hs |
| qPCR | MT-TL1 Fw | CACCCAAGAACAGGGTTTGT | Hs |
| qPCR | MT-TL1 Rv | TGGCCATGGGTATGTTGTTA | Hs |
| qPCR | DNM1L Fw | TCACCCGGAGACCTCTCATTC | Hs |
| qPCR | DNM1L Rv | GGTTCAGGGCTTACTCCCTTAT | Hs |
| qPCR | MFN1 Fw | TGGCTAAGAAGGCGATTACTGC | Hs |
| qPCR | MFN1 Rv | TCTCCGAGATAGCACCTCACC | Hs |
| qPCR | MFN2 Fw | CTCTCGATGCAACTCTATCGTC | Hs |
| qPCR | MFN2 Rv | TCCTGTACGTGTCTTCAAGGAA | Hs |
| qPCR | NDUFV1 Fw | TGACTGGTACAAGACAAAGGAGA | Hs |
| qPCR | NDUFV1 Rv | TGCCATCTGAGGGCTTATTCA | Hs |
| qPCR | MT-ND2 Fw | CGATGGTGCAGCCGCTATTA | Hs |

| | | | |
|------|------------|-------------------------|----|
| qPCR | MT-ND2 Rv | AGATGTGGCGGGTTTTAGGG | Hs |
| qPCR | SDHA Fw | ATGTTGGTGCCTGAGAACGA | Hs |
| qPCR | SDHA Rv | TCACCTAGTAAACCGTCTGCC | Hs |
| qPCR | UQCRB Fw | CCACTTGCCTGATGCTGAAAC | Hs |
| qPCR | UQCRB Rv | TGAGTACTGGTTGCTTGCCTAT | Hs |
| qPCR | ATP5F1A Fw | GTAAGTGCGGCGTAGACTGA | Hs |
| qPCR | ATP5F1A Rv | AGCTGGCATTGTTTAAGCTTGT | Hs |
| qPCR | CD36 Fw | CAGGTCAACCTATTGGTCAAGCC | Hs |
| qPCR | CD36 Rv | GCCTTCTCATCACCAATGGTCC | Hs |
| qPCR | FABP1 Fw | GTGTCGGAAATCGTGCAGAAT | Hs |
| qPCR | FABP1 Rv | GACTTTCTCCCCTGTCATTGTC | Hs |
| qPCR | FABP4 Fw | ACGAGAGGATGATAAACTGGTGG | Hs |
| qPCR | FABP4 Rv | GCGAACTTCAGTCCAGGTCAAC | Hs |
| qPCR | FASN Fw | AAGGACCTGTCTAGGTTTGATGC | Hs |
| qPCR | FASN Rv | TGGCTTCATAGGTGACTTCCA | Hs |
| qPCR | ACSL1 Fw | ATCAGGCTGCTCATGGATGACC | Hs |
| qPCR | ACSL1 Rv | AGTCCAAGAGCCATCGCTTCAG | Hs |
| qPCR | CTP1A Fw | AGTTCTCTTGCCCTGAGACG | Hs |
| qPCR | CTP1A Rv | GTCCATGGTCTCCTCCAAGG | Hs |
| qPCR | ACOX1 Fw | ACCTTGCTTCACCAGGCAAC | Hs |
| qPCR | ACOX1 Rv | CCAAGCCTCGAAGGTGAGTT | Hs |
| qPCR | PPARa Fw | CGGTGACTTATCCTGTGGTCC | Hs |
| qPCR | PPARa Rv | CCGCAGATTCTACATTCGATGTT | Hs |

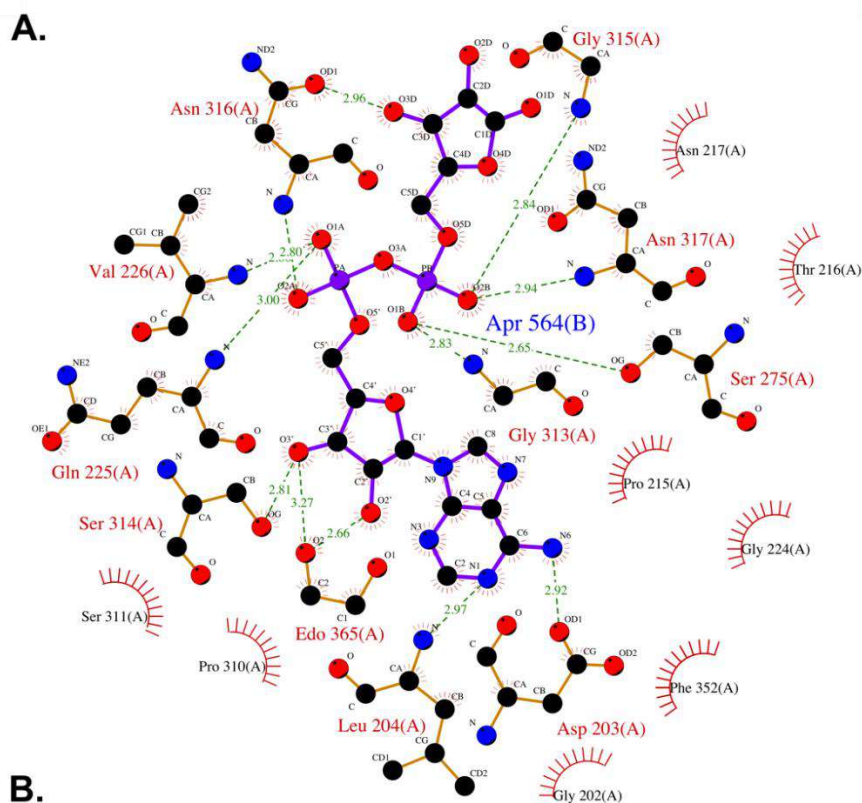
| | | | |
|--------------------|--------------------|-----------------------------|----|
| qPCR | PPARg Fw | AGCCTGCGAAAGCCTTTTGGTG | Hs |
| qPCR | PPARg Rv | GGCTTCACATTCAGCAAACCTGG | Hs |
| qPCR | SLC2A2 Fw | CTAGTCAGGTGCATGTGCCA | Hs |
| qPCR | SLC2A2 Rv | AACTGGAAGGAACCCAGCAC | Hs |
| qPCR | HK1 Fw | CACATGGAGTCCGAGGTTTATG | Hs |
| qPCR | HK1 Rv | CGTGAATCCCACAGGTA ACTTC | Hs |
| qPCR | PFKL Fw | GGCTTCGACACCCCGTGTA | Hs |
| qPCR | PFKL Rv | CGTCAAACCTCTTGTCATCCA | Hs |
| qPCR | LDHA Fw | TTGACCTACGTGGCTTGGAAG | Hs |
| qPCR | LDHA Rv | GGTAACGGAATCGGGCTGAAT | Hs |
| qPCR | CS Fw | TAAGGAGCAGGCCAGAATTAAGA | Hs |
| qPCR | CS Rv | CCACCATACATCATGTCCACAG | Hs |
| qPCR | ACLY Fw | CTGGGGTGCTCCGGATTTT | Hs |
| qPCR | ACLY Rv | CTGAAATTGCCTTGCCGAC | Hs |
| qPCR | IDH1 Fw | ATAATGTTGGCGTCAAATGTGC | Hs |
| qPCR | IDH1 Rv | CTTGA ACTCCTCAACCCTCTTC | Hs |
| qPCR | IDH2 Fw | CGCCACTATGCCGACAAAAG | Hs |
| qPCR | IDH2 Rv | ACTGCCAGATAATACGGGTCA | Hs |
| qPCR | OGDH Fw | TTGGCTGGAAAACCCCAAAG | Hs |
| qPCR | OGDH Rv | TGTGCTTCTACCAGGGACTGT | Hs |
| qPCR | ACY1 Fw | CAGCGCCATGACCAGCA | Hs |
| qPCR | ACY1 Rv | CGGGCTGTCTCCTCAAAGAA | Hs |
| Mutagenesis | mH2A1.2 I202 Fw | CACAGTGAAgagAGTAATTTAGCCGGC | Mm |

| | | | |
|--------------------|------------------|--|----|
| Mutagenesis | mH2A1.2 I202 Rv | GGCTAAATTACTctcTTCACCTGTGAATAAGGTTCAGC | Mm |
| Mutagenesis | mH2A1.2 D225Q Fw | GCTGACATTGACCTTAAAcagGACCTAGGAAACACACTGG | Mm |
| Mutagenesis | mH2A1.2 D225Q Rv | GTTTCCTAGGTcctgTTTAAGGTCAATGTCAGCATTGG | Mm |
| Mutagenesis | mH2A1.2 G315E Fw | CCATTGGCAGCgagAGGAACGGGTTCCTCCG | Mm |
| Mutagenesis | mH2A1.2 G315E Rv | CCCGTTCCTctcGCTGCCAATGGATGGGAAGGCG | Mm |
| Mutagenesis | mH2A1.2 R316I Fw | GGCAGCGGCatcAACGGGTTCCTCCGAAGCAGACAGCGG | Mm |
| Mutagenesis | mH2A1.2 R316I Rv | CGGGAACCCGTTgatGCCGCTGCCAATGGATGGG | Mm |
| qPCR | GAPDH Fw | CGACCACTTTGTCAAGCTCA | Hs |
| qPCR | GAPDH Rv | TCTTACTCCTTGGAGGCCAT | Hs |
| qPCR | RPLP0 Fw | ACATCTCCCCCTTCTCCTTC | Mm |
| qPCR | RPLP0 Rv | TACCCGATCTGCAGACACAC | Mm |
| qPCR | Rpl7 Fw | GAAGCTCATCTATGAGAAGGC | Mm |
| qPCR | Rpl7 Rv | AAGACGAAGGAGCTGCAGAAC | Mm |
| qPCR | H2AFY E5s Fw | CCTACAGACGGCTTCACTGTC | Mm |
| qPCR | H2AFY E6as Rv | GGTCAATGTCAGCATTGGTAGG | Mm |
| qPCR | H2AFY E7as Rv | GTGTAGAAGTCAGTGTGGTTCG | Mm |
| qPCR | mH2A2 Fw | GCTGGAAGAGACCATCAAAAA | Mm |
| qPCR | mH2A2 Rv | CGAAGTGAGCCGAGATGG | Mm |
| shRNA | shRNA scrambled | CCGGTCTCGCTTGGGCGAGAGTAAGCTCGAGCTTACTCTCGCCAAGCGAGATTTTTG | Mm |
| shRNA | sh-mH2A1 #1 | CCGGGCATGCTTCGGTACATCAAGACTCGAGTCTTGATGTACCGAAGCATGCTTTTTG | Mm |

| | | | |
|--|--------------------|---|----|
| shRNA | sh-mH2A1 #2 | CCGGGCTAAAGGGTGTCCACCATAGCCTCGAG GCTATGGTGACACCCTTTAGCTTTTTG | Mm |
| shRNA | sh-mH2A1 #3 | CCGGGCCATAATCAATCCTACCAATCTCGAG ATTGGTAGGATTGATTATGGCTTTTTG | Mm |
| Cloning | GST-mH2A1 MD Fw | GGTCGTGGGATCCCCGAATTCggcttcactgtcctc cacc | Mm |
| Cloning | GST-mH2A1 MD Rv | tgcgccgctcgagtcgaccggCTAGTTGGCGTCCAG CTTGCC | Mm |
| <i>In situ</i> hybridization (Lein et al, 2007) | mH2A1 Fw | CCCGGCACATCCTGTTAG | Mm |
| <i>In situ</i> hybridization (Lein et al, 2007) | mH2A1 Rv | CACTGGCCTAGTTGGCGT | Mm |
| Mouse phenotyping | Dre3 | TGCTGTTCCCTCCTATCCAC | Mm |
| Mouse phenotyping | Dre4 | CGGAGTCCATCAGCCTAGAG | Mm |

8 Annex

8.1 LigPlot interaction network



Number and type of bonds established between macrodomains and ADP-ribose

| | <i>Capsaspora</i> macroH2A | Human macroH2A1.1 |
|----------------|----------------------------|-------------------|
| Hydrogen bonds | 11 | 17 |
| Other bonds | 8 | 6 |

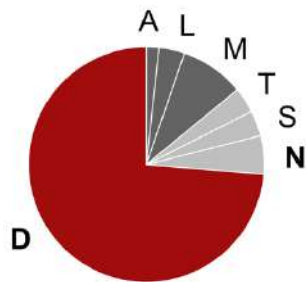
List of residues establishing other bonds than hydrogen bonds

| <i>Capsaspora</i> macroH2A | Human macroH2A1.1 |
|----------------------------|----------------------|
| Gly202 | Pro215 |
| Pro215 | Thr216 |
| Thr216 | Asn217 |
| Asn217 | Ser310 |
| Gly224 | Ile311 |
| Pro310 | Phe351 (pi-stacking) |
| Ser311 | |
| Phe352 (pi-stacking) | |

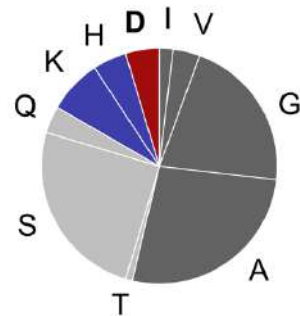
(A) The schematic representation of ADPR interaction with *Capsaspora* macroH2A macrodomain. All non-covalent interactions between ADPR and the macrodomain are summarized in the LigPlot diagram. ADPR ligand is represented in thick purple line and the amino acid residues of *Capsaspora* macroH2A macrodomain in thin orange lines. Hydrogen bonds are represented by the dashed lines between atoms involved, while the circles or semicircles with radiating lines represent atoms or residues involved in hydrophobic contacts between protein and ligand. **(B)** Table indicating the amino acids establishing H-bond or other bonds with ADP-ribose for *Capsaspora* macroH2A (see also Supplementary Figure S2C) and human macroH2A1.1 (Timinszky et al., 2009).

8.2 Other residues at positions 225 and 316 during the evolution of macroH2A(1.1)-like macrodomains

n (species with other at 225) = 57



n (species with other at 316) = 108



Residues: ■ positive ■ negative ■ polar ■ non-polar

The pie charts showing the identity of other residues at positions 225 and 316. Residue colors are based on their biochemical properties: positively charged in blue, negatively charged in red, polar in gray, and non-polar in black.

8.3 Compounds table

| Number | Compound ID | ZINC ID | Figure |
|--------|-----------------|------------------|--------------------------|
| 1 | cmpnd 1 | ZINC21981427 | Figure 30B |
| 2 | cmpnd 2 (3-OHR) | ZINC30731123 | Figure 30B and Figure 32 |
| 3 | cmpnd 2.1 | ZINC000095366474 | Figure 32 |
| 4 | cmpnd 2.2 | ZINC000241587909 | Figure 32 |
| 5 | cmpnd 2.3 | ZINC000097766679 | Figure 32 |
| 6 | cmpnd 2.4 | ZINC000004893738 | Figure 32 |
| 7 | cmpnd 2.5 | ZINC000048922086 | Figure 32 |
| 8 | cmpnd 2.6 | ZINC000216544438 | Figure 32 |
| 9 | cmpnd 2.7 | ZINC000065548609 | Figure 32 |
| 10 | cmpnd 2.8 | ZINC000004961993 | Figure 32 |
| 11 | cmpnd 2.9 | ZINC000000002011 | Figure 32 |
| 12 | cmpnd 2.10 | ZINC000040071651 | Figure 32 |
| 13 | cmpnd 2.11 | ZINC000065562432 | Figure 32 |
| 14 | cmpnd 3 | ZINC30731121 | Figure 30B |
| 15 | cmpnd 4 | ZINC00000507 | Figure 30B |
| 16 | cmpnd 5 | ZINC56874458 | Figure 30B |
| 17 | cmpnd 6 | ZINC13543134 | Figure 30B |
| 18 | cmpnd 7 | ZINC05783970 | Figure 30B |
| 19 | cmpnd 8 | ZINC13458842 | Figure 30B |
| 20 | cmpnd 9 | ZINC04097612 | Figure 30B |

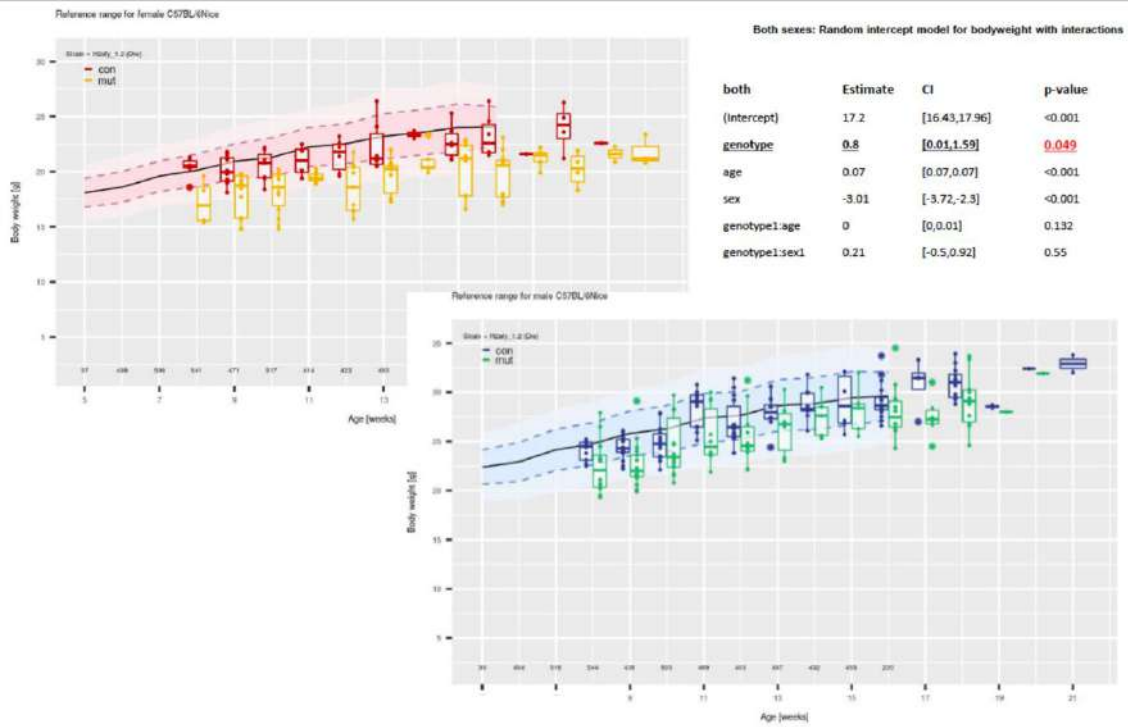
8.4 Detailed Gene Ontology table

| GO ID | GO term | BH adjusted p.value | Input genes |
|------------|---|---------------------|-----------------|
| GO:0016126 | sterol biosynthetic process | 5,87E-06 | KD mH2A1.2 down |
| GO:0055002 | striated muscle cell development | 5,87E-06 | KD mH2A1.2 down |
| GO:0055001 | muscle cell development | 1,10E-05 | KD mH2A1.2 down |
| GO:0006695 | cholesterol biosynthetic process | 1,39E-05 | KD mH2A1.2 down |
| GO:1901617 | organic hydroxy compound biosynthetic process | 1,40E-05 | KD mH2A1.2 down |
| GO:1902653 | secondary alcohol biosynthetic process | 1,40E-05 | KD mH2A1.2 down |
| GO:0008299 | isoprenoid biosynthetic process | 2,42E-05 | KD mH2A1.2 down |
| GO:0046165 | alcohol biosynthetic process | 3,05E-05 | KD mH2A1.2 down |
| GO:0008610 | lipid biosynthetic process | 7,01E-05 | KD mH2A1.2 down |
| GO:0016125 | sterol metabolic process | 8,77E-05 | KD mH2A1.2 down |
| GO:0007155 | cell adhesion | 3,87E-05 | KD mH2A1.2 up |
| GO:0022610 | biological adhesion | 3,87E-05 | KD mH2A1.2 up |
| GO:0001655 | urogenital system development | 0,001573179 | KD mH2A1.2 up |
| GO:0016477 | cell migration | 0,001655517 | KD mH2A1.2 up |
| GO:0030334 | regulation of cell migration | 0,001655517 | KD mH2A1.2 up |
| GO:0051270 | regulation of cellular component movement | 0,001655517 | KD mH2A1.2 up |
| GO:0072001 | renal system development | 0,001655517 | KD mH2A1.2 up |
| GO:0001822 | kidney development | 0,001844137 | KD mH2A1.2 up |
| GO:0072073 | kidney epithelium development | 0,002105186 | KD mH2A1.2 up |
| GO:2000145 | regulation of cell motility | 0,002105186 | KD mH2A1.2 up |
| GO:0001525 | angiogenesis | 0,00226415 | KD mH2A1.1 down |
| GO:0001944 | vasculature development | 0,00226415 | KD mH2A1.1 down |
| GO:0072358 | cardiovascular system development | 0,00226415 | KD mH2A1.1 down |
| GO:0072359 | circulatory system development | 0,00226415 | KD mH2A1.1 down |
| GO:0001568 | blood vessel development | 0,003300619 | KD mH2A1.1 down |
| GO:0001503 | ossification | 0,004556973 | KD mH2A1.1 down |
| GO:0035239 | tube morphogenesis | 0,004556973 | KD mH2A1.1 down |
| GO:0048514 | blood vessel morphogenesis | 0,004556973 | KD mH2A1.1 down |
| GO:0030198 | extracellular matrix organization | 0,00823924 | KD mH2A1.1 down |
| GO:0003197 | endocardial cushion development | 0,008246873 | KD mH2A1.1 down |
| GO:0051607 | defense response to virus | 3,34E-16 | KD mH2A1.1 up |
| GO:0009615 | response to virus | 9,33E-16 | KD mH2A1.1 up |
| GO:0035456 | response to interferon-beta | 2,73E-10 | KD mH2A1.1 up |
| GO:0035458 | cellular response to interferon-beta | 6,73E-10 | KD mH2A1.1 up |
| GO:0098542 | defense response to other organism | 6,04E-08 | KD mH2A1.1 up |

| | | | |
|------------|---|----------|---------------|
| GO:0006952 | defense response | 7,24E-08 | KD mH2A1.1 up |
| GO:0045087 | innate immune response | 7,67E-07 | KD mH2A1.1 up |
| GO:0045071 | negative regulation of viral genome replication | 8,37E-07 | KD mH2A1.1 up |
| GO:0060700 | regulation of ribonuclease activity | 3,70E-06 | KD mH2A1.1 up |
| GO:0048525 | negative regulation of viral process | 6,78E-06 | KD mH2A1.1 up |

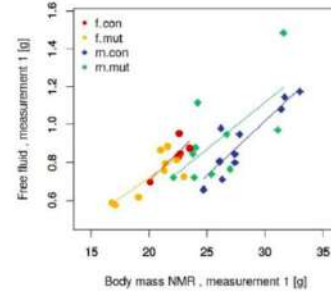
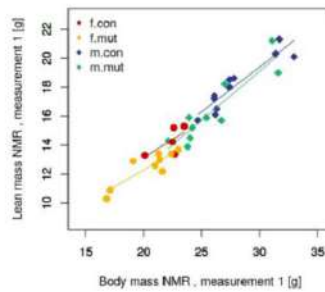
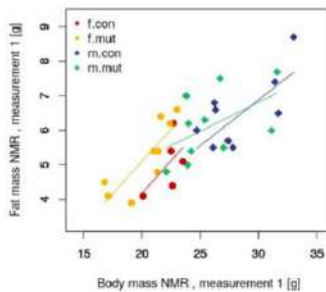
8.5 The measurements of body weight and composition

Body weight



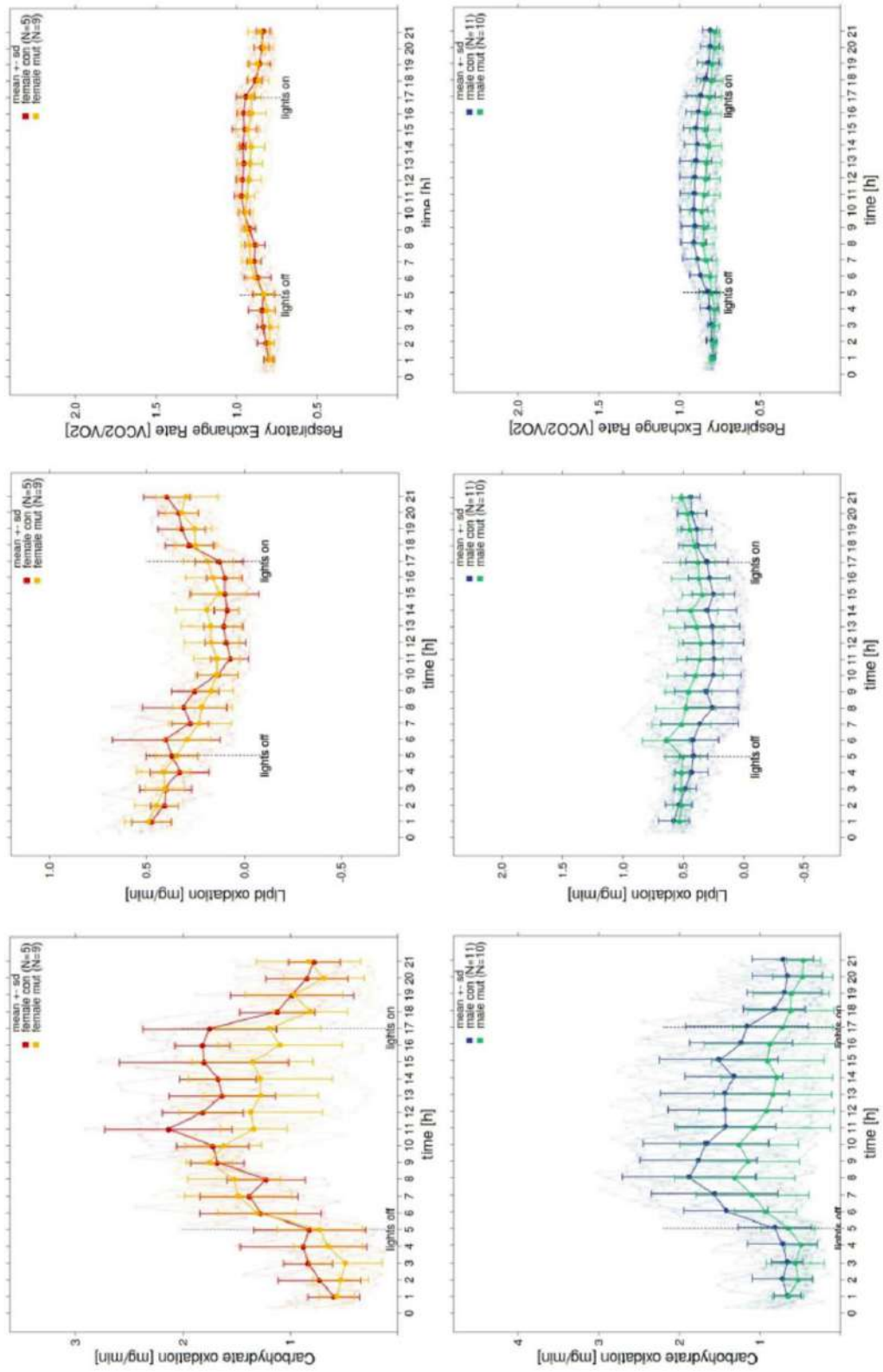
Body composition analysis I (13 weeks)

| | female | | male | | linear model | linear model | linear model | linear model |
|----------------------|------------|------------|------------|------------|--------------|--------------|--------------|--------------|
| | con | mut | con | mut | sex | genotype | Bodymass | sex:genotype |
| | n=5 | n=9 | n=11 | n=10 | | | | pe |
| | mean ± sd | mean ± sd | mean ± sd | mean ± sd | p-value | p-value | p-value | p-value |
| Body mass NMR | 22.3 ± 1.3 | 20.4 ± 2.2 | 28 ± 2.7 | 26 ± 3.2 | < 0.001 | 0.044 | NA | 0.932 |
| Fat mass NMR | 5 ± 0.8 | 5.3 ± 1 | 6.4 ± 1 | 6.1 ± 1 | 0.46 | 0.101 | < 0.001 | 0.469 |
| Lean mass NMR | 14.3 ± 1 | 12.5 ± 1.2 | 18.2 ± 1.8 | 16.4 ± 2.4 | 0.293 | 0.054 | < 0.001 | 0.839 |
| Free fluid | 0.8 ± 0.1 | 0.7 ± 0.1 | 0.9 ± 0.2 | 0.9 ± 0.2 | 0.01 | 0.217 | < 0.001 | 0.131 |



8.6 Indirect calorimetry results

Indirect Calorimetry (13 weeks) – substrate oxidation rates



9 Bibliography

- Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., & Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, *44*(W1), W3–W10. <https://doi.org/10.1093/nar/gkw343>
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*, *68*(4), 352–367. <https://doi.org/10.1107/S0907444912001308>
- Allen, M. D., Buckle, A. M., Cordell, S. C., Löwe, J., & Bycroft, M. (2003). The crystal structure of AF1521 a protein from *Archaeoglobus fulgidus* with homology to the non-histone domain of macroH2A. *Journal of Molecular Biology*, *330*(3), 503–511. [https://doi.org/10.1016/S0022-2836\(03\)00473-X](https://doi.org/10.1016/S0022-2836(03)00473-X)
- Altmeyer, M., & Hottiger, M. O. (2009). Poly(ADP-ribose) polymerase 1 at the crossroad of metabolic stress and inflammation in aging. *Aging*, *1*(5), 458–469. <https://doi.org/10.18632/aging.100052>
- Alva, V., & Lupas, A. N. (2019). Histones predate the split between bacteria and archaea. *Bioinformatics*, *35*(14), 2349–2353. <https://doi.org/10.1093/bioinformatics/bty1000>
- Anastassiadis, K., Fu, J., Patsch, C., Hu, S., Weidlich, S., Duerschke, K., Buchholz, F., Edenhofer, F., & Stewart, A. F. (2009). Dre recombinase, like Cre, is a highly efficient site-specific recombinase in *E. coli*, mammalian cells and mice. *DMM Disease Models and Mechanisms*, *2*(9–10), 508–515. <https://doi.org/10.1242/dmm.003087>
- Angelov, D., Molla, A., Perche, P. Y., Hans, F., Côté, J., Khochbin, S., Bouvet, P., & Dimitrov, S. (2003). The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. *Molecular Cell*, *11*(4), 1033–1041. [https://doi.org/10.1016/S1097-2765\(03\)00100-X](https://doi.org/10.1016/S1097-2765(03)00100-X)
- Araque, A., & Navarrete, M. (2010). Glial cells in neuronal network function. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 365, Issue 1551, pp. 2375–2381). Royal Society. <https://doi.org/10.1098/rstb.2009.0313>
- Aretz, J., Anumala, U. R., Fuchsberger, F. F., Molavi, N., Ziebart, N., Zhang, H., Nazaré, M., & Rademacher, C. (2018). Allosteric Inhibition of a Mammalian Lectin. *Journal of the American Chemical Society*, *140*(44), 14915–14925. <https://doi.org/10.1021/jacs.8b08644>
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*(W1), W344–W350. <https://doi.org/10.1093/nar/gkw408>
- Bai, P., Cantó, C., Oudart, H., Brunyánszki, A., Cen, Y., Thomas, C., Yamamoto, H., Huber, A., Kiss, B., Houtkooper, R. H., Schoonjans, K., Schreiber, V., Sauve, A. A., Menissier-De Murcia, J., & Auwerx, J. (2011). PARP-1 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metabolism*, *13*(4), 461–468. <https://doi.org/10.1016/j.cmet.2011.03.004>
- Bawono, P., & Heringa, J. (2014). PRALINE: A versatile multiple sequence alignment toolkit. *Methods in Molecular Biology*, *1079*, 245–262. https://doi.org/10.1007/978-1-62703-646-7_16
- Beales, N. (2004). Adaptation of Microorganisms to Cold Temperatures, Weak Acid Preservatives, Low pH, and Osmotic Stress: A Review. *Comprehensive Reviews in Food Science and Food Safety*, *3*(1), 1–20. <https://doi.org/10.1111/j.1541-4337.2004.tb00057.x>
- Berger, F., Lau, C., Dahlmann, M., & Ziegler, M. (2005). Subcellular compartmentation and differential catalytic properties of the three human nicotinamide mononucleotide adenylyltransferase isoforms. *Journal of Biological Chemistry*, *280*(43), 36334–36341.

<https://doi.org/10.1074/jbc.M508660200>

- Bockwoldt, M., Houry, D., Niere, M., Gossman, T. I., Reinartz, I., Schug, A., Ziegler, M., & Heiland, I. (2019). Identification of evolutionary and kinetic drivers of NAD-dependent signaling. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(32), 15957–15966. <https://doi.org/10.1073/pnas.1902346116>
- Boulard, M., Storck, S., Cong, R., Pinto, R., Delage, H., & Bouvet, P. (2010). Histone variant macroH2A1 deletion in mice causes female-specific steatosis. *Epigenetics and Chromatin*, *3*(1), 1–13. <https://doi.org/10.1186/1756-8935-3-8>
- Brown, M. W., Heiss, A. A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A. K., Shiratori, T., Ishida, K. I., Hashimoto, T., Simpson, A. G. B., & Roger, A. J. (2018). Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. *Genome Biology and Evolution*, *10*(2), 427–433. <https://doi.org/10.1093/gbe/evy014>
- Bruce, K. D., Zsombok, A., & Eckel, R. H. (2017). Lipid processing in the brain: A key regulator of systemic metabolism. *Frontiers in Endocrinology*, *8*(APR), 1–11. <https://doi.org/10.3389/fendo.2017.00060>
- Brunk, C. F., & Martin, W. F. (2019). Archaeal Histone Contributions to the Origin of Eukaryotes. *Trends in Microbiology*, *27*(8), 703–714. <https://doi.org/10.1016/j.tim.2019.04.002>
- Burki, F., Roger, A. J., Brown, M. W., & Simpson, A. G. B. (2020). The New Tree of Eukaryotes. In *Trends in Ecology and Evolution* (Vol. 35, Issue 1, pp. 43–55). Elsevier Ltd. <https://doi.org/10.1016/j.tree.2019.08.008>
- Buschbeck, M., & di Croce, L. (2010). Approaching the molecular and physiological function of macroH2A variants. *Epigenetics*, *5*(2), 118–123. <https://doi.org/10.4161/epi.5.2.11076>
- Buschbeck, M., & Hake, S. B. (2017). Variants of core histones and their roles in development , stem cells and cancer. *Nature Publishing Group*. <https://doi.org/10.1038/nrm.2016.166>
- Buschbeck, M., Uribealago, I., Wibowo, I., Rué, P., Martin, D., Gutierrez, A., Morey, L., Guigó, R., López-Schier, H., & Di Croce, L. (2009). The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nature Structural & Molecular Biology*, *16*(10), 1074–1079. <https://doi.org/10.1038/nsmb.1665>
- Caires, R., Sierra-Valdez, F. J., Millet, J. R. M., Herwig, J. D., Roan, E., Vásquez, V., & Cordero-Morales, J. F. (2017). Omega-3 Fatty Acids Modulate TRPV4 Function through Plasma Membrane Remodeling. *Cell Reports*, *21*(1), 246–258. <https://doi.org/10.1016/j.celrep.2017.09.029>
- Cambronne, X. A., & Kraus, W. L. (2020). Location, Location, Location: Compartmentalization of NAD⁺ Synthesis and Functions in Mammalian Cells. *Trends in Biochemical Sciences*, *45*(10), 858–873. <https://doi.org/10.1016/j.tibs.2020.05.010>
- Cambronne, X. A., Stewart, M. L., Kim, D., Jones-Brunette, A. M., Morgan, R. K., Farrens, D. L., Cohen, M. S., & Goodman, R. H. (2016). Biosensor reveals multiple sources for mitochondrial NAD⁺. *Science*, *352*(6292), 1474–1477. <https://doi.org/10.1126/science.aad5168>
- Cantariño, N., Douet, J., & Buschbeck, M. (2013). MacroH2A - An epigenetic regulator of cancer. *Cancer Letters*, *336*(2), 247–252. <https://doi.org/10.1016/j.canlet.2013.03.022>
- Canto, C., Menzies, K. J., & Auwerx, J. (2015). NAD⁺ Metabolism and the Control of Energy Homeostasis: A Balancing Act between Mitochondria and the Nucleus. *Cell Metabolism*, *22*(1), 31–53. <https://doi.org/10.1016/j.cmet.2015.05.023>
- Catara, G., Corteggio, A., Valente, C., Grimaldi, G., & Palazzo, L. (2019). Targeting ADP-ribosylation as an antimicrobial strategy. *Biochemical Pharmacology*, *167*(June), 13–26. <https://doi.org/10.1016/j.bcp.2019.06.001>
- Chadwick, B. P., & Willard, H. F. (2001). Histone H2A variants and the inactive X chromosome: Identification of a second macroH2A variant. *Human Molecular Genetics*, *10*(10), 1101–1113. <https://doi.org/10.1093/hmg/10.10.1101>

- Chakravarthy, S., Gundimella, S. K. Y., Caron, C., Perche, P., Pehrson, J. R., & Khochbin, S. (2005). Structural Characterization of the Histone Variant macroH2A. *Molecular and Cellular Biology*, 25(17), 7616–7624. <https://doi.org/10.1128/MCB.25.17.7616>
- Chakravarthy, S., Patel, A., & Bowman, G. D. (2012). The basic linker of macroH2A stabilizes DNA at the entry/exit site of the nucleosome. *Nucleic Acids Research*, 40(17), 8285–8295. <https://doi.org/10.1093/nar/gks645>
- Chakravarthy, M. V., Lodhi, I. J., Yin, L., Malapaka, R. R. V., Xu, H. E., Turk, J., & Semenkovich, C. F. (2009). Identification of a Physiologically Relevant Endogenous Ligand for PPAR α in Liver. *Cell*, 138(3), 476–488. <https://doi.org/10.1016/j.cell.2009.05.036>
- Chakravarthy, M. V., Zhu, Y., López, M., Yin, L., Wozniak, D. F., Coleman, T., Hu, Z., Wolfgang, M., Vidal-Puig, A., Lane, M. D., & Semenkovich, C. F. (2007). Brain fatty acid synthase activates PPAR α to maintain energy homeostasis. *Journal of Clinical Investigation*, 117(9), 2539–2552. <https://doi.org/10.1172/JCI31183>
- Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T. A., Brusniak, M. Y., Paulse, C., Creasy, D., ... Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. In *Nature Biotechnology* (Vol. 30, Issue 10, pp. 918–920). Nat Biotechnol. <https://doi.org/10.1038/nbt.2377>
- Chang, E. Y., Ferreira, H., Somers, J., Nusinow, D. A., Owen-Hughes, T., & Narlikar, G. J. (2008). MacroH2A allows ATP-dependent chromatin remodeling by SWI/SNF and ACF complexes but specifically reduces recruitment of SWI/SNF. *Biochemistry*, 47(51), 13726–13732. <https://doi.org/10.1021/bi8016944>
- Changolkar, L. N., Costanzi, C., Leu, N. A., Chen, D., McLaughlin, K. J., & Pehrson, J. R. (2007). Developmental Changes in Histone macroH2A1-Mediated Gene Regulation. *Molecular and Cellular Biology*, 27(7), 2758–2764. <https://doi.org/10.1128/mcb.02334-06>
- Changolkar, L. N., Singh, G., Cui, K., Berletch, J. B., Zhao, K., Disteche, C. M., & Pehrson, J. R. (2010). Genome-Wide Distribution of MacroH2A1 Histone Variants in Mouse Liver Chromatin. *Molecular and Cellular Biology*, 30(23), 5473–5483. <https://doi.org/10.1128/mcb.00518-10>
- Chen, H., Ruiz, P. D., McKimpson, W. M., Novikov, L., Kitsis, R. N., & Gamble, M. J. (2015). MacroH2A1 and ATM Play Opposing Roles in Paracrine Senescence and the Senescence-Associated Secretory Phenotype. *Molecular Cell*, 59(5), 719–731. <https://doi.org/10.1016/j.molcel.2015.07.011>
- Chen, H., Ruiz, P. D., Novikov, L., Casill, A. D., Park, J. W., & Gamble, M. J. (2014). MacroH2A1.1 and PARP-1 cooperate to regulate transcription by promoting CBP-mediated H2B acetylation. *Nature Structural & Molecular Biology*, 21, 981–989. <https://doi.org/doi:10.1038/nsmb.2903>
- Chen, W., Smeekens, J. M., & Wu, R. (2016). Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chemical Science*, 7(2), 1393–1400. <https://doi.org/10.1039/c5sc03826j>
- Chiapparino, A., Maeda, K., Turei, D., Saez-Rodriguez, J., & Gavin, A. C. (2016). The orchestra of lipid-transfer proteins at the crossroads between metabolism and signaling. *Progress in Lipid Research*, 61, 30–39. <https://doi.org/10.1016/j.plipres.2015.10.004>
- Commerford, S. L., Carsten, A. L., & Cronkite, E. P. (1982). Histone turnover within nonproliferating cells. *Proceedings of the National Academy of Sciences of the United States of America*, 79(4 I), 1163–1165. <https://doi.org/10.1073/pnas.79.4.1163>
- Costanzi, C., & Pehrson, J. R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature*, 393, 599–601.
- Costanzi, C., & Pehrson, J. R. (2001). MACROH2A2, a New Member of the MACROH2A Core Histone Family. *Journal of Biological Chemistry*, 276(24), 21776–21784. <https://doi.org/10.1074/jbc.M010919200>

- Covarrubias, A. J., Perrone, R., Grozio, A., & Verdin, E. (2021). NAD⁺ metabolism and its roles in cellular processes during ageing. *Nature Reviews Molecular Cell Biology*, 22(2), 119–141. <https://doi.org/10.1038/s41580-020-00313-x>
- Creppe, C., Janich, P., Cantariño, N., Noguera, M., Valero, V., Musulén, E., Douet, J., Posavec, M., Martín-Caballero, J., Sumoy, L., Di Croce, L., Benitah, S. A., & Buschbeck, M. (2012). MacroH2A1 regulates the balance between self-renewal and differentiation commitment in embryonic and adult stem cells. *Molecular and Cellular Biology*, 32(8), 1442–1452. <https://doi.org/10.1128/MCB.06323-11>
- Creppe, C., Posavec, M., Douet, J., & Buschbeck, M. (2012). MacroH2A in stem cells: a story beyond gene repression. *Epigenomics*, 4(2), 221–227. <https://doi.org/10.2217/epi.12.8>
- Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, 14(6), 1188–1190. <https://doi.org/10.1101/gr.849004>
- Crowder, M. K., Seacrist, C. D., & Blind, R. D. (2017). Phospholipid regulation of the nuclear receptor superfamily. *Advances in Biological Regulation*, 63, 6–14. <https://doi.org/10.1016/j.jbior.2016.10.006>
- Dardenne, E., Pierredon, S., Driouch, K., Gratadou, L., Lacroix-Triki, M., Polay Espinoza, M., Zonta, E., Germann, S., Mortada, H., Villemain, J.-P., Dutertre, M., Lidereau, R., Vagner, S., & Auboeuf, D. (2012). Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. *Nature Structural & Molecular Biology*, 19, 1139–1146. <https://doi.org/doi:10.1038/nsmb.2390>
- Deans, C., & Maggert, K. A. (2015). What do you mean, “Epigenetic”? *Genetics*, 199(4), 887–896. <https://doi.org/10.1534/genetics.114.173492>
- Dell’Orso, S., Wang, A. H., Shih, H. Y., Saso, K., Berghella, L., Gutierrez-Cruz, G., Ladurner, A. G., O’Shea, J. J., Sartorelli, V., & Zare, H. (2016). The Histone Variant MacroH2A1.2 Is Necessary for the Activation of Muscle Enhancers and Recruitment of the Transcription Factor Pbx1. *Cell Reports*, 14(5), 1156–1168. <https://doi.org/10.1016/j.celrep.2015.12.103>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dosselaere, F., & Vanderleyden, J. (2001). A metabolic node in action: Chorismate-utilizing enzymes in microorganisms. In *Critical Reviews in Microbiology* (Vol. 27, Issue 2, pp. 75–131). CRC Press LLC. <https://doi.org/10.1080/20014091096710>
- Douet, J., Corujo, D., Malinverni, R., Renauld, J., Sansoni, V., Marjanović, M. P., Cantariño, N., Valero, V., Mongelard, F., Bouvet, P., Imhof, A., Thiry, M., & Buschbeck, M. (2017). MacroH2A histone variants maintain nuclear organization and heterochromatin architecture. *Journal of Cell Science*, 130(9), 1570–1582. <https://doi.org/10.1242/jcs.199216>
- Doyen, C.-M., An, W., Angelov, D., Bondarenko, V., Mietton, F., Studitsky, V. M., Hamiche, A., Roeder, R. G., Bouvet, P., & Dimitrov, S. (2006). Mechanism of Polymerase II Transcription Repression by the Histone Variant macroH2A. *Molecular and Cellular Biology*, 26(3), 1156–1164. <https://doi.org/10.1128/MCB.26.3.1156>
- Dudin, O., Ondracka, A., Grau-Bové, X., Haraldsen, A. A. B., Toyoda, A., Suga, H., Bråte, J., & Ruiz-Trillo, I. (2019). A unicellular relative of animals generates a layer of polarized cells by actomyosin-dependent cellularization. *ELife*, 8. <https://doi.org/10.7554/eLife.49801>
- Eidhof, I., Baets, J., Kamsteeg, E. J., Deconinck, T., Van Ninhuijs, L., Martin, J. J., Schüle, R., Züchner, S., De Jonghe, P., Schenck, A., & Van De Warrenburg, B. P. (2018). GDAP2 mutations implicate susceptibility to cellular stress in a new form of cerebellar ataxia. *Brain*, 141(9), 2592–2604. <https://doi.org/10.1093/brain/awy198>
- Elia, J., Laracy, S., Allen, J., Nissley-Tsiopinis, J., & Borgmann-Winter, K. (2012). Epigenetics: Genetics versus life experiences. *Current Topics in Behavioral Neurosciences*, 9, 317–340. https://doi.org/10.1007/7854_2011_144
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative

- genomics. *Genome Biology*, 20(1). <https://doi.org/10.1186/s13059-019-1832-y>
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), 486–501. <https://doi.org/10.1107/S0907444910007493>
- Emsley, Paul, & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I), 2126–2132. <https://doi.org/10.1107/S0907444904019158>
- Fornasiero, E. F., Mandad, S., Wildhagen, H., Alevra, M., Rammner, B., Keihani, S., Opazo, F., Urban, I., Ischebeck, T., Sakib, M. S., Fard, M. K., Kirli, K., Centeno, T. P., Vidal, R. O., Rahman, R. U., Benito, E., Fischer, A., Dennerlein, S., Rehling, P., ... Rizzoli, S. O. (2018). Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-06519-0>
- Frederick, D. W., Loro, E., Liu, L., Davila, A., Chellappa, K., Silverman, I. M., Quinn, W. J., Gosai, S. J., Tichy, E. D., Davis, J. G., Mourkioti, F., Gregory, B. D., Dellinger, R. W., Redpath, P., Migaud, M. E., Nakamaru-Ogiso, E., Rabinowitz, J. D., Khurana, T. S., & Baur, J. A. (2016). Loss of NAD Homeostasis Leads to Progressive and Reversible Degeneration of Skeletal Muscle. *Cell Metabolism*, 24(2), 269–282. <https://doi.org/10.1016/j.cmet.2016.07.005>
- Fu, Y., Lv, P., Yan, G., Fan, H., Cheng, L., Zhang, F., Dang, Y., Wu, H., & Wen, B. (2015). MacroH2A1 associates with nuclear lamina and maintains chromatin architecture in mouse liver cells. *Scientific Reports*, 5(August), 1–12. <https://doi.org/10.1038/srep17186>
- Gabaldón, T., & Pittis, A. A. (2015). Origin and evolution of metabolic sub-cellular compartmentalization in eukaryotes. *Biochimie*, 119, 262–268. <https://doi.org/10.1016/j.biochi.2015.03.021>
- Gamble, M. J., Frizzell, K. M., Yang, C., Krishnakumar, R., & Kraus, W. L. (2010). The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. *Genes and Development*, 24(1), 21–32. <https://doi.org/10.1101/gad.1876110>
- Gaspar-Maia, A., Qadeer, Z. A., Hasson, D., Ratnakumar, K., Adrian Leu, N., Leroy, G., Liu, S., Costanzi, C., Valle-Garcia, D., Schaniel, C., Lemischka, I., Garcia, B., Pehrson, J. R., & Bernstein, E. (2013). MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. *Nature Communications*, 4(1), 1–13. <https://doi.org/10.1038/ncomms2582>
- Gibson, B. A., & Kraus, W. L. (2012). New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nature Reviews Molecular Cell Biology*, 13(7), 411–424. <https://doi.org/10.1038/nrm3376>
- Girardi, E., Agrimi, G., Goldmann, U., Fiume, G., Lindinger, S., Sedlyarov, V., Srndic, I., Gürtl, B., Agerer, B., Kartnig, F., Scarcia, P., Noia, M. A. Di, Liñeiro, E., Rebsamen, M., Wiedmer, T., Bergthaler, A., Palmieri, L., & Superti-Furga, G. (2020). Epistasis-driven identification of SLC25A51 as a regulator of human mitochondrial NAD import. *Nature Communications*, 11(1), 1–9. <https://doi.org/10.1038/s41467-020-19871-x>
- Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. *Cell*, 128(4), 635–638. <https://doi.org/10.1016/j.cell.2007.02.006>
- Grau-Bové, X., Torruella, G., Donachie, S., Suga, H., Leonard, G., Richards, T. A., & Ruiz-Trillo, I. (2017). Dynamics of genomic innovation in the unicellular ancestry of animals. *ELife*, 6, 1–35. <https://doi.org/10.7554/eLife.26036>
- Grolla, A. A., Miggianno, R., Marino, D. Di, Bianchi, M., Gori, A., Orsomando, G., Gaudino, F., Galli, U., Grosso, E. Del, Mazzola, F., Angeletti, C., Guarneri, M., Torretta, S., Calabrò, M., Boumya, S., Fan, X., Colombo, G., Travelli, C., Rocchio, F., ... Garavaglia, S. (2020). A nicotinamide phosphoribosyltransferase–GAPDH interaction sustains the stress-induced NMN/NAD⁺ salvage pathway in the nucleus. *Journal of Biological Chemistry*, 295(11), 3635–3651. <https://doi.org/10.1074/jbc.RA119.010571>

- Gupte, R., Liu, Z., & Kraus, W. L. (2017). PARPs and ADP-ribosylation: recent advances linking molecular functions to biological outcomes. *Genes and Development*, 31(2), 101–126. <https://doi.org/10.1101/gad.291518.116>
- Harayama, T., & Riezman, H. (2018). Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology*, 19(5), 281–296. <https://doi.org/10.1038/nrm.2017.138>
- Hedges, S. B., Dudley, J., & Kumar, S. (2006). TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics*, 22(23), 2971–2972. <https://doi.org/10.1093/bioinformatics/btl505>
- Hehenberger, E., Tikhonenkov, D. V., Kolisko, M., del Campo, J., Esaulov, A. S., Mylnikov, A. P., & Keeling, P. J. (2017). Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals. *Current Biology*, 27(13), 2043–2050.e6. <https://doi.org/10.1016/j.cub.2017.06.006>
- 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>
- Hsu, C. J., Meers, O., Buschbeck, M., & Heidel, F. . (2021). The Role of MacroH2A Histone Variants in Cancer. *Cancers*, 13(3003). <https://doi.org/10.3390/cancers13123003>
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., Von Mering, C., & Bork, P. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), D309–D314. <https://doi.org/10.1093/nar/gky1085>
- Huh, J. W., Shima, J., & Ochi, K. (1996). ADP-ribosylation of proteins in *Bacillus subtilis* and its possible importance in sporulation. *Journal of Bacteriology*, 178(16), 4935–4941. <https://doi.org/10.1128/jb.178.16.4935-4941.1996>
- Hurtado-Bagès, S., Guberovic, I., & Buschbeck, M. (2018). The MacroH2A1.1 - PARP1 axis at the intersection between stress response and metabolism. *Frontiers in Genetics*, 9(OCT). <https://doi.org/10.3389/fgene.2018.00417>
- Hurtado-Bagès, S., Knobloch, G., Ladurner, A. G., & Buschbeck, M. (2020). The taming of PARP1 and its impact on NAD⁺ metabolism. *Molecular Metabolism*, xxx, 100950. <https://doi.org/10.1016/j.molmet.2020.01.014>
- Hurtado-Bagès, S., Posavec Marjanovic, M., Valero, V., Malinverni, R., Corujo, D., Bouvet, P., Lavigne, A. C., Bystricky, K., & Buschbeck, M. (2020). The Histone Variant MacroH2A1 Regulates Key Genes for Myogenic Cell Fusion in a Splice-Isoform Dependent Manner. *Cells*, 9(5). <https://doi.org/10.3390/cells9051109>
- Jankevicius, G., Hassler, M., Golia, B., Rybin, V., Zacharias, M., Timinszky, G., & Ladurner, A. G. (2013). A family of macrodomain proteins reverses cellular mono-ADP-ribosylation. *Nature Structural & Molecular Biology*, 20(4), 508–514. <https://doi.org/10.1038/nsmb.2523>
- Jones-Villeneuve, E. M. V., McBurney, M. W., Rogers, K. A., & Kalnins, V. I. (1982). Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *Journal of Cell Biology*, 94(2), 253–262. <https://doi.org/10.1083/jcb.94.2.253>
- Kabsch, W. (2010). XDS. *Acta Crystallographica Section D Biological Crystallography*, 66(2), 125–132. <https://doi.org/10.1107/S0907444909047337>
- Kamata, S., Oyama, T., Saito, K., Honda, A., Yamamoto, Y., Suda, K., Ishikawa, R., Itoh, T., Watanabe, Y., Shibata, T., Uchida, K., Suematsu, M., & Ishii, I. (2020). PPAR α Ligand-Binding Domain Structures with Endogenous Fatty Acids and Fibrates. *IScience*, 23(11), 101727. <https://doi.org/10.1016/j.isci.2020.101727>
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. In *Nucleic Acids Research* (Vol. 28, Issue 1, pp. 27–30). Oxford University Press. <https://doi.org/10.1093/nar/28.1.27>

- Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, 428(4), 726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>
- Kapoor, A., Goldberg, M. S., Cumberland, L. K., Ratnakumar, K., Segura, M. F., Emanuel, P. O., Menendez, S., Vardabasso, C., LeRoy, G., Vidal, C. I., Polsky, D., Osman, I., Garcia, B. A., Hernando, E., & Bernstein, E. (2010). The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. *Nature*, 468, 1105–1109. <https://doi.org/doi:10.1038/nature09590>
- Karras, G. I., Kustatscher, G., Buhecha, H. R., Allen, M. D., Pugieux, C., Sait, F., Bycroft, M., & Ladurner, A. G. (2005). The macro domain is an ADP-ribose binding module. *The EMBO Journal*, 24(11), 1911–1920. <https://doi.org/10.1038/sj.emboj.7600664>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Khochbin, S. (2001). Histone H1 diversity: Bridging regulatory signals to linker histone function. In *Gene* (Vol. 271, Issue 1, pp. 1–12). Elsevier. [https://doi.org/10.1016/S0378-1119\(01\)00495-4](https://doi.org/10.1016/S0378-1119(01)00495-4)
- Khurana, S., Kruhlak, M. J., Kim, J., Tran, A. D., Liu, J., Nyswaner, K., Shi, L., Jailwala, P., Sung, M. H., Hakim, O., & Oberdoerffer, P. (2014). A macrohistone variant links dynamic chromatin compaction to BRCA1-dependent genome maintenance. *Cell Reports*, 8(4), 1049–1062. <https://doi.org/10.1016/j.celrep.2014.07.024>
- Kim, J. M., Shin, Y., Lee, S., Kim, M. Y., Punj, V., Shin, H. I., Kim, K., Koh, J. M., Jeong, D., & An, W. (2018). MacroH2A1.2 inhibits prostate cancer-induced osteoclastogenesis through cooperation with HP1 α and H1.2: Supplementary materials. *Oncogene*.
- Kim, Jeongkyu, Sun, C., Tran, A. D., Chin, P. J., Ruiz, P. D., Wang, K., Gibbons, R. J., Gamble, M. J., Liu, Y., & Oberdoerffer, P. (2019). The macroH2A1.2 histone variant links ATRX loss to alternative telomere lengthening. *Nature Structural and Molecular Biology*, 26(3), 213–219. <https://doi.org/10.1038/s41594-019-0192-3>
- Kim, Jinman, Shin, Y., Lee, S., Kim, M., Punj, V., Lu, J. F., Shin, H., Kim, K., Ulmer, T. S., Koh, J., Jeong, D., & An, W. (2018). Regulation of Breast Cancer-Induced Osteoclastogenesis by MacroH2A1.2 Involving EZH2-Mediated H3K27me3. *Cell Reports*, 24(1), 224–237. <https://doi.org/10.1016/j.celrep.2018.06.020>
- Kornberg, R. D. (1977). Structure of chromatin. In *Annual review of biochemistry* (Vol. 46, pp. 931–954). Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA . <https://doi.org/10.1146/annurev.bi.46.070177.004435>
- Kory, N., de Bos, J. uit, van der Rijt, S., Jankovic, N., Güra, M., Arp, N., Pena, I. A., Prakash, G., Chan, S. H., Kunchok, T., Lewis, C. A., & Sabatini, D. M. (2020). MCART1/SLC25A51 is required for mitochondrial NAD transport. *Science Advances*, 6(43). <https://doi.org/10.1126/sciadv.abe5310>
- Kozlowski, M., Corujo, D., Hothorn, M., Guberovic, I., Mandemaker, I. K., Blessing, C., Sporn, J., Gutierrez-Triana, A., Smith, R., Portmann, T., Treier, M., Scheffzek, K., Huet, S., Timinszky, G., Buschbeck, M., & Ladurner, A. G. (2018a). MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms: Supplementary materials. *EMBO Reports*.
- Kozlowski, M., Corujo, D., Hothorn, M., Guberovic, I., Mandemaker, I. K., Blessing, C., Sporn, J., Gutierrez-Triana, A., Smith, R., Portmann, T., Treier, M., Scheffzek, K., Huet, S., Timinszky, G., Buschbeck, M., & Ladurner, A. G. (2018b). MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms. *EMBO Reports*, 19(10), 1–13. <https://doi.org/10.15252/embr.201744445>
- Kruger, A. N., Brogley, M. A., Huizinga, J. L., Kidd, J. M., de Rooij, D. G., Hu, Y. C., & Mueller, J. L. (2019). A Neofunctionalized X-Linked Ampliconic Gene Family Is Essential for Male Fertility and Equal Sex Ratio in Mice. *Current Biology*, 29(21), 3699–3706.e5. <https://doi.org/10.1016/j.cub.2019.08.057>

- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K., & Ladurner, A. G. (2005). Splicing regulates NAD metabolite binding to histone macroH2A. *Nature Structural & Molecular Biology*, 12(7), 624–625. <https://doi.org/10.1038/nsmb956>
- Lane, N., & Martin, W. (2010). The energetics of genome complexity. *Nature*, 467(7318), 929–934. <https://doi.org/10.1038/nature09486>
- Lavigne, M. D., Agelopoulos, M., & Thanos, D. (2015). Composite macroH2A/NRF-1 Nucleosomes Suppress Noise and Generate Robustness in Gene Expression Accession Numbers GSE53133 GSE53103 Lavigne et al. *Cell Reports*, 11. <https://doi.org/10.1016/j.celrep.2015.04.022>
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7), 1307–1320. <https://doi.org/10.1093/molbev/msn067>
- Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., Chen, L., Chen, L., Chen, T. M., Chin, M. C., Chong, J., Crook, B. E., Czaplinska, A., Dang, C. N., Datta, S., ... Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445(7124), 168–176. <https://doi.org/10.1038/nature05453>
- Lemke, G. (2001). Glial control of neuronal development. In *Annual Review of Neuroscience* (Vol. 24, pp. 87–105). Annu Rev Neurosci. <https://doi.org/10.1146/annurev.neuro.24.1.87>
- Lenzi, A., Picardo, M., Gandini, L., & Dondero, F. (1996). Lipids of the sperm plasma membrane: From polyunsaturated fatty acids considered as markers of sperm function to possible scavenger therapy. *Human Reproduction Update*, 2(3), 246–256. <https://doi.org/10.1093/humupd/2.3.246>
- Li, F., Yi, P., Pi, J., Li, L., Hui, J., Wang, F., Liang, A., & Yu, J. (2016). QKI5-mediated alternative splicing of the histone variant macroH2A1 regulates gastric carcinogenesis. *Oncotarget*, 7(22), 32821–32834. <https://doi.org/10.18632/oncotarget.8739>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Liu, H., Nakagawa, T., Kanematsu, T., Uchida, T., & Tsuji, S. (1999). Isolation of 10 differentially expressed cDNAs in differentiated Neuro2a cells induced through controlled expression of the GD3 synthase gene. *Journal of Neurochemistry*, 72(5), 1781–1790. <https://doi.org/10.1046/j.1471-4159.1999.0721781.x>
- Lo Re, O., Fusilli, C., Rappa, F., Van Haele, M., Douet, J., Pindjakova, J., Rocha, S. W., Pata, I., Valčíková, B., Uldrijan, S., Yeung, R. S., Peixoto, C. A., Roskams, T., Buschbeck, M., Mazza, T., & Vinciguerra, M. (2018). Induction of cancer cell stemness by depletion of macrohistone H2A1 in hepatocellular carcinoma. *Hepatology*, 67(2), 636–650. <https://doi.org/10.1002/hep.29519>
- Long, M., Sun, X., Shi, W., Yanru, A., Leung, S. T. C., Ding, D., Cheema, M. S., MacPherson, N., Nelson, C. J., Ausio, J., Yan, Y., & Ishibashi, T. (2019). A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Research*, 47(16), 8399–8409. <https://doi.org/10.1093/nar/gkz547>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/s13059-014-0550-8>
- Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), 251–260. <https://doi.org/10.1038/38444>
- Luo, X., Ryu, K. W., Kim, D. S., Nandu, T., Medina, C. J., Gupte, R., Gibson, B. A., Soccio, R. E.,

- Yu, Y., Gupta, R. K., & Kraus, W. L. (2017). PARP-1 Controls the Adipogenic Transcriptional Program by PARylating C/EBP β and Modulating Its Transcriptional Activity. *Molecular Cell*, 65(2), 260–271. <https://doi.org/10.1016/j.molcel.2016.11.015>
- Luongo, T. S., Eller, J. M., Lu, M. J., Niere, M., Raith, F., Perry, C., Bornstein, M. R., Oliphint, P., Wang, L., McReynolds, M. R., Migaud, M. E., Rabinowitz, J. D., Johnson, F. B., Johnson, K., Ziegler, M., Cambronne, X. A., & Baur, J. A. (2020). SLC25A51 is a mammalian mitochondrial NAD⁺ transporter. *Nature*, 588(December 2019). <https://doi.org/10.1038/s41586-020-2741-7>
- Ma, H., Su, L., Xia, W., Wang, W., Tan, G., & Jiao, J. (2021). MacroH2A1.2 deficiency leads to neural stem cell differentiation defects and autism-like behaviors. *EMBO Reports*, 1–19. <https://doi.org/10.15252/embr.202052150>
- Martire, S., & Banaszynski, L. A. (2020). The roles of histone variants in fine-tuning chromatin organization and function. *Nature Reviews Molecular Cell Biology*, 21, 522–541. <https://doi.org/10.1038/s41580-020-0262-8>
- Mathieson, T., Franken, H., Kosinski, J., Kurzawa, N., Zinn, N., Sweetman, G., Poeckel, D., Ratnu, V. S., Schramm, M., Becher, I., Steidel, M., Noh, K. M., Bergamini, G., Beck, M., Bantscheff, M., & Savitski, M. M. (2018). Systematic analysis of protein turnover in primary cells. *Nature Communications*, 9(1), 1–10. <https://doi.org/10.1038/s41467-018-03106-1>
- Mayer, M., & Meyer, B. (1999). Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie - International Edition*, 38(12), 1784–1788. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990614\)38:12<1784::AID-ANIE1784>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q)
- McGhee, J. D., & Felsenfeld, G. (1980). Nucleosome structure. *Annual Review of Biochemistry*, 49, 1115–1156. <https://doi.org/10.1146/annurev.bi.49.070180.005343>
- Mermoud, J. E., Costanzi, C., Pehrson, J. R., & Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *Journal of Cell Biology*, 147(7), 1399–1408. <https://doi.org/10.1083/jcb.147.7.1399>
- Mijaljica, D., Prescott, M., & Devenish, R. J. (2011). Microautophagy in mammalian cells: Revisiting a 40-year-old conundrum. In *Autophagy* (Vol. 7, Issue 7, pp. 673–682). Taylor and Francis Inc. <https://doi.org/10.4161/auto.7.7.14733>
- Montaigne, D., Butruille, L., & Staels, B. (2021). PPAR control of metabolism and cardiovascular functions. *Nature Reviews Cardiology*. <https://doi.org/10.1038/s41569-021-00569-6>
- Montaner, A., Da Silva Santana, T. T., Schroeder, T., Einiker-Lamas, M., Girardini, J., Costa, M. R., & Banchio, C. (2018). Specific Phospholipids Regulate the Acquisition of Neuronal and Astroglial Identities in Post-Mitotic Cells. *Scientific Reports*, 8(1), 1–13. <https://doi.org/10.1038/s41598-017-18700-4>
- Moriarty, N. W., Grosse-Kunstleve, R. W., & Adams, P. D. (2009). Electronic ligand builder and optimization workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallographica Section D: Biological Crystallography*, 65(10), 1074–1080. <https://doi.org/10.1107/S0907444409029436>
- Morowitz, H. J. (1999). A theory of biochemical organization, metabolic pathways, and evolution. *Complexity*, 4(6), 39–53. [https://doi.org/10.1002/\(SICI\)1099-0526\(199907/08\)4:6<39::AID-CPLX8>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1099-0526(199907/08)4:6<39::AID-CPLX8>3.0.CO;2-2)
- Mortusewicz, O., Amé, J. C., Schreiber, V., & Leonhardt, H. (2007). Feedback-regulated poly(ADP-ribosylation) by PARP-1 is required for rapid response to DNA damage in living cells. *Nucleic Acids Research*, 35(22), 7665–7675. <https://doi.org/10.1093/nar/gkm933>
- Murr, R. (2010). Interplay between different epigenetic modifications and mechanisms. In *Advances in Genetics* (Vol. 70, Issue C). Academic Press. <https://doi.org/10.1016/B978-0-12-380866-0.60005-8>
- Musille, P. M., Kohn, J. A., & Ortlund, E. A. (2013). *Phospholipid – Driven gene regulation*. 587, 1238–1246.

- Muthurajan, U. M., McBryant, S. J., Lu, X., Hansen, J. C., & Luger, K. (2011). The linker region of macroH2A promotes self-association of nucleosomal arrays. *Journal of Biological Chemistry*, 286(27), 23852–23864. <https://doi.org/10.1074/jbc.M111.244871>
- Nagatsu, T. (2006). The catecholamine system in health and disease - Relation to tyrosine 3-monoxygenase and other catecholamine-synthesizing enzymes. In *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* (Vol. 82, Issue 10, pp. 388–415). The Japan Academy. <https://doi.org/10.2183/pjab.82.388>
- Nashun, B., Yukawa, M., Liu, H., Akiyama, T., & Aoki, F. (2010). Changes in the nuclear deposition of histone H2A variants during pre-implantation development in mice. *Development*, 137(22), 3785–3794. <https://doi.org/10.1242/dev.051805>
- Nelson, B. R., Hodge, R. D., Daza, R. A. M., Tripathi, P. P., Arnold, S. J., Millen, K. J., & Hevner, R. F. (2020). Intermediate progenitors support migration of neural stem cells into dentate gyrus outer neurogenic niches. *ELife*, 9. <https://doi.org/10.7554/eLife.53777>
- Neuvonen, M., & Ahola, T. (2009). Differential Activities of Cellular and Viral Macro Domain Proteins in Binding of ADP-Ribose Metabolites. *Journal of Molecular Biology*, 385(1), 212–225. <https://doi.org/10.1016/j.jmb.2008.10.045>
- Ni, K., Ren, J., Xu, X., He, Y., Finney, R., Braun, S. M. G., Hathaway, N. A., Crabtree, G. R., & Muegge, K. (2020). LSH mediates gene repression through macroH2A deposition. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-19159-0>
- Novikov, L., Park, J. W., Chen, H., Klerman, H., Jalloh, A. S., & Gamble, M. J. (2011). QKI-Mediated Alternative Splicing of the Histone Variant MAcroH2A1 Regulated Cancer Cell Proliferation. *Molecular and Cellular Biology*, 31, 4244–4255. <https://doi.org/doi:10.1128/MCB.05244-11>
- Nusinow, D. A., Hernández-Muñoz, I., Fazzio, T. G., Shah, G. M., Kraus, W. L., & Panning, B. (2007). Poly(ADP-ribose) polymerase 1 is inhibited by a histone H2A variant, macroH2A, and contributes to silencing of the inactive X chromosome. *Journal of Biological Chemistry*, 282(17), 12851–12859. <https://doi.org/10.1074/jbc.M610502200>
- Oláh, G., Szczyzny, B., Brunyánszki, A., Lopez-García, I. A., Gerö, D., Radák, Z., & Szabo, C. (2015). Differentiation-associated downregulation of poly(ADP-Ribose) polymerase-1 expression in myoblasts serves to increase their resistance to oxidative stress. *PLoS ONE*, 10(7), 1–22. <https://doi.org/10.1371/journal.pone.0134227>
- Orsak, T., Smith, T. L., Eckert, D., Lindsley, J. E., Borges, C. R., & Rutter, J. (2012). Revealing the allosterome: Systematic identification of metabolite-protein interactions. *Biochemistry*, 51(1), 225–232. <https://doi.org/10.1021/bi201313s>
- Ouararhni, K., Hadj-Slimane, R., Ait-Si-Ali, S., Robin, P., Mietton, F., Harel-Bellan, A., Dimitrov, S., & Hamiche, A. (2006). The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity. *Genes and Development*, 20(23), 3324–3336. <https://doi.org/10.1101/gad.396106>
- Palazzo, L., Mikočević, P., Mikoč, A., & Ahel, I. (2019). ADP-ribosylation signalling and human disease. In *Open Biology* (Vol. 9, Issue 4). Royal Society Publishing. <https://doi.org/10.1098/rsob.190041>
- Pasque, V., Gillich, A., Garrett, N., & Gurdon, J. B. (2011). Histone variant macroH2A confers resistance to nuclear reprogramming. *EMBO Journal*, 30(12), 2373–2387. <https://doi.org/10.1038/emboj.2011.144>
- Paterson, I., & Lam, N. Y. S. (2018). Challenges and discoveries in the total synthesis of complex polyketide natural products. In *Journal of Antibiotics* (Vol. 71, Issue 2, pp. 215–233). Nature Publishing Group. <https://doi.org/10.1038/ja.2017.111>
- Pazienza, V., Borghesan, M., Mazza, T., Sheedfar, F., Panebianco, C., Williams, R., Mazzoccoli, G., Andriulli, A., Nakanishi, T., & Vinciguerra, M. (2014). SIRT1-metabolite binding histone macroH2A1.1 protects hepatocytes against lipid accumulation. *Aging*, 6(1), 35–47. <https://doi.org/10.18632/aging.100632>

- Pazienza, V., Panebianco, C., Rappa, F., Memoli, D., Borghesan, M., Cannito, S., Oji, A., Mazza, G., Tamburrino, D., Fusai, G., Barone, R., Bolasco, G., Villarroya, F., Villarroya, J., Hatsuzawa, K., Cappello, F., Tarallo, R., Nakanishi, T., & Vinciguerra, M. (2016). Histone macroH2A1.2 promotes metabolic health and leanness by inhibiting adipogenesis. *Epigenetics & Chromatin*, *9*, 45. <https://doi.org/10.1186/s13072-016-0098-9>
- Pehrson, J. R., Changolkar, L. N., Costanzi, C., & Leu, N. A. (2014). Mice without macroH2A histone variants. *Molecular and Cellular Biology*, *34*(24), 4523–4533. <https://doi.org/10.1128/MCB.00794-14>
- Pehrson, J. R., & Fried, V. A. (1992). MacroH2A, a core histone containing a large nonhistone region. *Science*, *257*(5075), 1398–1400. <https://doi.org/10.1126/science.1529340>
- Perdiguerro, E., Sousa-Victor, P., Ballestar, E., & Muñoz-Cánoves, P. (2009). Epigenetic regulation of myogenesis. In *Epigenetics* (Vol. 4, Issue 8, pp. 541–550). Taylor and Francis Inc. <https://doi.org/10.4161/epi.4.8.10258>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Pillai, A. S., Chandler, S. A., Liu, Y., Signore, A. V., Cortez-Romero, C. R., Benesch, J. L. P., Laganowsky, A., Storz, J. F., Hochberg, G. K. A., & Thornton, J. W. (2020). Origin of complexity in haemoglobin evolution. *Nature*, *581*(7809), 480–485. <https://doi.org/10.1038/s41586-020-2292-y>
- Pirinen, E., Cantó, C., Jo, Y. S., Morato, L., Zhang, H., Menzies, K. J., Williams, E. G., Mouchiroud, L., Moullan, N., Hagberg, C., Li, W., Timmers, S., Imhof, R., Verbeek, J., Pujol, A., Van Loon, B., Viscomi, C., Zeviani, M., Schrauwen, P., ... Auwerx, J. (2014). Pharmacological inhibition of poly(ADP-ribose) polymerases improves fitness and mitochondrial function in skeletal muscle. *Cell Metabolism*, *19*(6), 1034–1041. <https://doi.org/10.1016/j.cmet.2014.04.002>
- Posavec-Marjanović, M., Hurtado-Bagès, S., Lassi, M., Valero, V., Malinverni, R., Delage, H., Navarro, M., Corujo, D., Guberovic, I., Douet, J., Gama-Perez, P., Garcia-Roves, P. M., Ahel, I., Ladurner, A. G., Yanes, O., Bouvet, P., Suelves, M., Teperino, R., Pospisilik, J. A., & Buschbeck, M. (2017). MacroH2A1.1 regulates mitochondrial respiration by limiting nuclear NAD + consumption. *Nature Structural and Molecular Biology*, *24*(11), 902–910. <https://doi.org/10.1038/nsmb.3481>
- Rack, J. G. M., Perina, D., & Ahel, I. (2016). Macrodomains: Structure, Function, Evolution, and Catalytic Activities. *Annual Review of Biochemistry*, *85*(1), 431–454. <https://doi.org/10.1146/annurev-biochem-060815-014935>
- Rajman, L., Chwalek, K., & Sinclair, D. A. (2018). Therapeutic Potential of NAD-Boosting Molecules: The In Vivo Evidence. In *Cell Metabolism* (Vol. 27, Issue 3, pp. 529–547). Cell Press. <https://doi.org/10.1016/j.cmet.2018.02.011>
- Richter, D., Berney, C., Strassert, J., Burki, F., & de Vargas, C. (2020). EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotic life. *BioRxiv*, 2020.06.30.180687. <https://doi.org/10.1101/2020.06.30.180687>
- Rivera-Casas, C., Gonzalez-Romero, R., Cheema, M. S., Ausió, J., & Eirín-López, J. M. (2016). The characterization of macroH2A beyond vertebrates supports an ancestral origin and conserved role for histone variants in chromatin. *Epigenetics*, *11*(6), 415–425. <https://doi.org/10.1080/15592294.2016.1172161>
- Rosenthal, F., Feijs, K. L. H., Frugier, E., Bonalli, M., Forst, A. H., Imhof, R., Winkler, H. C., Fischer, D., Caflisch, A., Hassa, P. O., Lüscher, B., & Hottiger, M. O. (2013). Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases. *Nature Structural and Molecular Biology*, *20*(4), 502–507. <https://doi.org/10.1038/nsmb.2521>
- Roy, A., Kundu, M., Jana, M., Mishra, R. K., Yung, Y., Luan, C. H., Gonzalez, F. J., & Pahan, K. (2016). Identification and characterization of PPAR α ligands in the hippocampus. *Nature*

Chemical Biology, 12(12), 1075–1083. <https://doi.org/10.1038/nchembio.2204>

- Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Ruiz, P. D., & Gamble, M. J. (2018). MacroH2A1 chromatin specification requires its docking domain and acetylation of H2B lysine 20. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-07189-8>
- Ruiz, P. D., Hamilton, G. A., Park, J. W., & Gamble, M. J. (2020). MacroH2A1 Regulation of Poly(ADP-Ribose) Stability Prevents Necrosis and Promotes DNA Repair. *Molecular and Cellular Biology*, 40(1), 1–17. <https://doi.org/10.1128/MCB.00230-19>
- Ryu, K. W., Nandu, T., Kim, J., Challa, S., DeBerardinis, R. J., & Lee Kraus, W. (2018). Metabolic regulation of transcription through compartmentalized NAD⁺ biosynthesis. *Science*, 360(6389). <https://doi.org/10.1126/science.aan5780>
- Saito, K., Tautz, L., & Mustelin, T. (2007). The lipid-binding SEC14 domain. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1771(6), 719–726. <https://doi.org/10.1016/j.bbalip.2007.02.010>
- Sasaki, Y., Hackett, A. R., Kim, S., Strickland, A., & Milbrandt, J. (2018). Dysregulation of NAD⁺ metabolism induces a Schwann cell dedifferentiation program. *Journal of Neuroscience*, 38(29), 6546–6562. <https://doi.org/10.1523/JNEUROSCI.3304-17.2018>
- Schreiber, V., Dantzer, F., Amé, J. C., & De Murcia, G. (2006). Poly(ADP-ribose): Novel functions for an old molecule. In *Nature Reviews Molecular Cell Biology* (Vol. 7, Issue 7, pp. 517–528). Nat Rev Mol Cell Biol. <https://doi.org/10.1038/nrm1963>
- Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. In *Environmental Science and Technology* (Vol. 48, Issue 4, pp. 2097–2098). American Chemical Society. <https://doi.org/10.1021/es5002105>
- Sebastian, R., Hosogane, E. K., Sun, E. G., Tran, A. D., Reinhold, W. C., Burkett, S., Sturgill, D. M., Gudla, P. R., Pommier, Y., Aladjem, M. I., & Oberdoerffer, P. (2020). Epigenetic Regulation of DNA Repair Pathway Choice by MacroH2A1 Splice Variants Ensures Genome Stability. *Molecular Cell*, 79(5), 836–845.e7. <https://doi.org/10.1016/j.molcel.2020.06.028>
- Sebé-Pedrós, A., Ballaré, C., Parra-Acero, H., Chiva, C., Tena, J. J., Sabidó, E., Gómez-Skarmeta, J. L., Di Croce, L., & Ruiz-Trillo, I. (2016). The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity. *Cell*, 165(5), 1224–1237. <https://doi.org/10.1016/j.cell.2016.03.034>
- Sebé-Pedrós, A., Irimia, M., del Campo, J., Parra-Acero, H., Russ, C., Nusbaum, C., Blencowe, B. J., & Ruiz-Trillo, I. (2013). Regulated aggregative multicellularity in a close unicellular relative of metazoa. *eLife*, 2013(2), 1–20. <https://doi.org/10.7554/eLife.01287>
- Sebé-Pedrós, A., Peña, M. I., Capella-Gutiérrez, S., Antó, M., Gabaldón, T., Ruiz-Trillo, I., & Sabidó, E. (2016). High-Throughput Proteomics Reveals the Unicellular Roots of Animal Phosphosignaling and Cell Differentiation. *Developmental Cell*, 39(2), 186–197. <https://doi.org/10.1016/j.devcel.2016.09.019>
- Setlow, P., & Kornberg, A. (1970). Biochemical studies of bacterial sporulation and germination. *The Journal of Biological Chemistry*, 245(14), 3637–3644.
- Setlow, R., & Setlow, P. (1977). Levels of oxidized and reduced pyridine nucleotides in dormant spores and during growth, sporulation, and spore germination of *Bacillus megaterium*. *Journal of Bacteriology*, 129(2), 857–865. <https://doi.org/10.1128/jb.129.2.857-865.1977>
- Sheedfar, F., Vermeer, M., Paziienza, V., Villarroya, J., Rappa, F., Cappello, F., Mazzocchi, G., Villarroya, F., Van Der Molen, H., Hofker, M. H., Koonen, D. P., & Vinciguerra, M. (2015). Genetic ablation of macrohistone H2A1 leads to increased leanness, glucose tolerance and

- energy expenditure in mice fed a high-fat diet. *International Journal of Obesity*, 39(2), 331–338. <https://doi.org/10.1038/ijo.2014.91>
- Simonet, N. G., Thackray, J. K., Vazquez, B. N., Ianni, A., Espinosa-Alcantud, M., Morales-Sanfrutos, J., Hurtado-Bagès, S., Sabidó, E., Buschbeck, M., Tischfield, J., de la Torre, C., Esteller, M., Braun, T., Olivella, M., Serrano, L., & Vaquero, A. (2020). SirT7 auto-ADP-ribosylation regulates glucose starvation response through mH2A1. *Science Advances*, 6(30). <https://doi.org/10.1126/sciadv.aaz2590>
- Singh, H. R., Nardozza, A. P., Möller, I. R., Knobloch, G., Kistemaker, H. A. V., Hassler, M., Harrer, N., Blessing, C., Eustermann, S., Kotthoff, C., Huet, S., Mueller-Planitz, F., Filippov, D. V., Timinszky, G., Rand, K. D., & Ladurner, A. G. (2017). A Poly-ADP-Ribose Trigger Releases the Auto-Inhibition of a Chromatin Remodeling Oncogene. *Molecular Cell*, 68(5), 860–871.e7. <https://doi.org/10.1016/j.molcel.2017.11.019>
- Slade, D. (2020). PARP and PARG inhibitors in cancer treatment. In *Genes and Development* (Vol. 34, Issue 5, pp. 360–394). Cold Spring Harbor Laboratory Press. <https://doi.org/10.1101/gad.334516.119>
- Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3), 779–787. <https://doi.org/10.1021/ac051437y>
- Sporn, J. C., Kustatscher, G., Hothorn, T., Collado, M., Serrano, M., Muley, T., Schnabel, P., & Ladurner, A. G. (2009). Histone macroH2A isoforms predict the risk of lung cancer recurrence. *Oncogene*, 28(38), 3423–3428. <https://doi.org/10.1038/onc.2009.26>
- Sporn, Judith C., & Jung, B. (2012). Differential regulation and predictive potential of macroH2A1 isoforms in colon cancer. *American Journal of Pathology*, 180(6), 2516–2526. <https://doi.org/10.1016/j.ajpath.2012.02.027>
- Stevens, B. (2003). Glia: much more than the neuron's side-kick. In *Current biology : CB* (Vol. 13, Issue 12). Curr Biol. [https://doi.org/10.1016/s0960-9822\(03\)00404-4](https://doi.org/10.1016/s0960-9822(03)00404-4)
- Strømland, Ø., Niere, M., Nikiforov, A. A., VanLinden, M. R., Heiland, I., & Ziegler, M. (2019). Keeping the balance in NAD metabolism. *Biochemical Society Transactions*, 47(1), 119–130. <https://doi.org/10.1042/BST20180417>
- Suga, H., Chen, Z., De Mendoza, A., Sebé-Pedrós, A., Brown, M. W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N., Torruella, G., Derelle, R., Manning, G., Lang, B. F., Russ, C., Haas, B. J., Roger, A. J., Nusbaum, C., & Ruiz-Trillo, I. (2013). The Capsaspora genome reveals a complex unicellular prehistory of animals. *Nature Communications*, 4, 1–9. <https://doi.org/10.1038/ncomms3325>
- Sun, Z., Filipescu, D., Andrade, J., Gaspar-Maia, A., Ueberheide, B., & Bernstein, E. (2018). Transcription-associated histone pruning demarcates macroH2A chromatin domains. *Nature Structural and Molecular Biology*, 25(10), 958–970. <https://doi.org/10.1038/s41594-018-0134-5>
- Timinszky, G., Till, S., Hassa, P. O., Hothorn, M., Kustatscher, G., Nijmeijer, B., Colombelli, J., Altmeyer, M., Stelzer, E. H. K., Scheffzek, K., Hottiger, M. O., & Ladurner, A. G. (2009a). A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nature Structural and Molecular Biology*, 16(9), 923–929. <https://doi.org/10.1038/nsmb.1664>
- Timinszky, G., Till, S., Hassa, P. O., Hothorn, M., Kustatscher, G., Nijmeijer, B., Colombelli, J., Altmeyer, M., Stelzer, E. H. K., Scheffzek, K., Hottiger, M. O., & Ladurner, A. G. (2009b). A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nature Structural & Molecular Biology*, 16(9), 923–929. <https://doi.org/10.1038/nsmb.1664>
- Torruella, G., De Mendoza, A., Grau-Bové, X., Antó, M., Chaplin, M. A., Del Campo, J., Eme, L., Pérez-Cordón, G., Whipps, C. M., Nichols, K. M., Paley, R., Roger, A. J., Sitjà-Bobadilla, A., Donachie, S., & Ruiz-Trillo, I. (2015). Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Current Biology*, 25(18), 2404–2410. <https://doi.org/10.1016/j.cub.2015.07.053>

- Toska, E., Campbell, H. A., Shandilya, J., Goodfellow, S. J., Shore, P., Medler, K. F., & Roberts, S. G. E. (2012). Repression of Transcription by WT1-BASP1 Requires the Myristoylation of BASP1 and the PIP2-Dependent Recruitment of Histone Deacetylase. *Cell Reports*, 2(3), 462–469. <https://doi.org/10.1016/j.celrep.2012.08.005>
- Vieira-Silva, T. S., Monteiro-Reis, S., Barros-Silva, D., Ramalho-Carvalho, J., Graça, I., Carneiro, I., Martins, A. T., Oliveira, J., Antunes, L., Hurtado-Bagès, S., Buschbeck, M., Henrique, R., & Jerónimo, C. (2019). Histone variant MacroH2A1 is downregulated in prostate cancer and influences malignant cell phenotype. *Cancer Cell International*, 19(1), 1–13. <https://doi.org/10.1186/s12935-019-0835-9>
- Vuignier, K., Schappler, J., Veuthey, J. L., Carrupt, P. A., & Martel, S. (2010). Drug-protein binding: A critical review of analytical tools. In *Analytical and Bioanalytical Chemistry* (Vol. 398, Issue 1, pp. 53–66). Anal Bioanal Chem. <https://doi.org/10.1007/s00216-010-3737-1>
- Waddington, C. H. (2012). The epigenotype. 1942. *International Journal of Epidemiology*, 41(1), 10–13. <https://doi.org/10.1093/ije/dyr184>
- Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). Ligplot: A program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering, Design and Selection*, 8(2), 127–134. <https://doi.org/10.1093/protein/8.2.127>
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Weinhold, B. (2006). Epigenetics: the science of change. In *Environmental health perspectives*. (Vol. 114, Issue 3). <https://doi.org/10.1289/ehp.114-a160>
- Wilms, T., Swinnen, E., Eskes, E., Dolz-Edo, L., Uwineza, A., Van Essche, R., Rosseels, J., Zabrocki, P., Cameroni, E., Franssens, V., De Virgilio, C., Smits, G. J., & Winderickx, J. (2017). The yeast protein kinase Sch9 adjusts V-ATPase assembly/disassembly to control pH homeostasis and longevity in response to glucose availability. *PLoS Genetics*, 13(6). <https://doi.org/10.1371/journal.pgen.1006835>
- Xiao, W., Wang, R. S., Handy, D. E., & Loscalzo, J. (2018). NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism. *Antioxidants and Redox Signaling*, 28(3), 251–272. <https://doi.org/10.1089/ars.2017.7216>
- Xu, X., Ni, K., He, Y., Ren, J., Sun, C., Liu, Y., Aladjem, M. I., Burkett, S., Finney, R., Ding, X., Sharan, S. K., & Muegge, K. (2021). The epigenetic regulator LSH maintains fork protection and genomic stability via MacroH2A deposition and RAD51 filament formation. *Nature Communications*, 12(1), 3520. <https://doi.org/10.1038/s41467-021-23809-2>
- Yanes, O., Tautenhahn, R., Patti, G. J., & Siuzdak, G. (2011). Expanding coverage of the metabolome for global metabolite profiling. *Analytical Chemistry*, 83(6), 2152–2161. <https://doi.org/10.1021/ac102981k>
- Zecchin, A., Stapor, P. C., Goveia, J., & Carmeliet, P. (2015). Metabolic pathway compartmentalization: An underappreciated opportunity? *Current Opinion in Biotechnology*, 34, 73–81. <https://doi.org/10.1016/j.copbio.2014.11.022>
- Zhang, J., Rosenberg, H. F., & Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proceedings of the National Academy of Sciences of the United States of America*, 95(7), 3708–3713. <https://doi.org/10.1073/pnas.95.7.3708>
- Zhou, K., Gaullier, G., & Luger, K. (2019). Nucleosome structure and dynamics are coming of age. *Nature Structural & Molecular Biology*, 26, 3–13. <https://doi.org/10.1038/s41594-018-0166-x>
- Zick, M., Stroupe, C., Orr, A., Douville, D., & Wickner, W. T. (2014). Membranes linked by trans-SNARE complexes require lipids prone to non-bilayer structure for progression to fusion. *ELife*, 2014(3), 1–13. <https://doi.org/10.7554/eLife.01879>