UNIVERSITA' VITA-SALUTE SAN RAFFAELE

CORSO DI DOTTORATO DI RICERCA INTERNAZIONALE
IN MEDICINA MOLECOLARE

Curriculum in Cellular and Molecular Physiopathology

The histone perspective: chromatin regulation of cell identity through manipulation of histone content and specific variants

DoS: Dr.ssa Alessandra Agresti

Second Supervisor: Dr.ssa Raffaella Santoro

Tesi di DOTTORATO di RICERCA di Luca Michetti

matr. 013926

Ciclo di dottorato XXXIV

SSD BIO/11

Anno Accademico 2021/ 2022

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il/la sottoscritto/*I*	Luca Michetti
Matricola / *registration number*	013926
nat_ a/ *born at*	Roma
il/*on*	14/07/1993

autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD Thesis titled*

The histone perspective: chromatin regulation of cell identity through manipulation of histone content and specific variants

ɣ AUTORIZZA la Consultazione della tesi per 12 mesi / AUTHORIZES *the publicrelease of the thesis*

E' fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / *Copyright the contents of the thesis in whole or in part is forbidden*

Data /Date ....13/05/2022....……..... Firma /Signature …….....

DECLARATION

This thesis has been composed by myself and has not been used in any previous application for a degree. Throughout the text I use both ‘I’ and ‘We’ interchangeably.

All the results presented here were obtained by myself, except for:

1. The immunofluorescence staining in FIG. 13 performed in collaboration with Thodoris Karnavas.
2. The SUPER-Silac experiments (FIG. 14-15) performed in collaboration with Roberta Noberini, Tiziana Bonaldi group at IEO (European Institute of Oncology), Milan, Italy.
3. The bioinformatic analysis of RNAseq experiments, performed by Emanuele Monteleone at Università Vita Salute San Raffaele, Milan, Italy.
4. The RT-qPCR experiments on StemHistones KO, performed in collaboration with Erika Di Domenico at Università Vita Salute San Raffaele, Milan, Italy.

All sources of information are acknowledged by means of reference.

ACKNOWLEDGEMENTS

I would like to thank my Second Supervisor Raffaella Santoro for her suggestions during and after my annual seminars, and for useful discussions.

I also want to thank Dr.ssa Roberta Noberini and Dr.ssa Tiziana Bonaldi for the SUPER-Silac experiments and analysis, and for thoughtful discussions on the interpretation of the results.

I am grateful to Dr. Emanuele Monteleone for all the bioinformatic analysis on RNAseq experiments and the Center of Omics Sciences at IRCCS Ospedale San Raffaele (COSR) for technical support in transcriptomic experiments.

I thank Dr. Thodoris Karnavas for starting the Hmgb1 reprogramming project and for useful discussions.

Last, I would like to thank the Fondazione San Raffaele and especially the Fronzaroli Family for providing the funding for my PhD fellowship.

DEDICATION

Alla fine di questi 3 lunghi anni, vorrei ringraziare Alessandra e Marco per avermi affidato questo progetto e per avermi guidato nel corso del dottorato. Grazie a tutte le persone presenti e passate del Lab Bianchi, in particolare a Francesco, Francesca, Liam, Arianna, Samu, Emilie e Davide. Ringrazio Thodoris per l'aiuto "remoto" del primo anno e Ema per avermi aiutato a trovare idee nuove quando non ce ne erano (oltre che per all things computational, l'aiuto giusto al momento giusto..!).

È bello avere idee nuove e che ci sia qualcuno che ti aiuti a metterle in pratica. Grazie a Erika, che per un lungo anno di tesi magistrale e oltre è stata la spalla (e le braccia e le pipette) che mi hanno permesso di fare in 15 mesi il lavoro di tre anni. Sarai per sempre la mia prima studenta!

Grazie alle due persone che hanno condiviso più di tutte con me la quotidianità di questo PhD. Elly e Çise AKA the Power Rangers. Çise thank you for the sarcasm and the reciprocal psychological support, both very much needed helps in the dark PhD times, and for being the best office-mate when I needed a quiet place. Grazie Elly perché molto semplicemente ho trovato una sorella in più, e pure bresciana..! Grazie per l'accoglienza e l'ascolto, per i weekend a cambiare il mezzo, per il nostro ciaccolare ininterrotto da zitelle. E grazie pure perché mi hai fatto conoscere Setfano (non potevo esimermi, scusa), il finto burbero più buono che c'è! Vi meritate un grazie di coppia per le imbucate a cena e la Sardegna.

Grazie ai coinquilini di questi anni, a Manu per il lockdown e la parmigiane, a Cami perché ha portato una ventata di gioia e di spontaneità quando serviva di più. Grazie a Thomas, because as I always say, asking you to come live in the house was the best idea I had during my PhD. I couldn't have hoped for a better housemate and friend, guitar buddy and calcetto mate. Don't lose your religion and shine on your crazy single-cells Tommaso!

Grazie al Prof e a Peppe, che si sono presi cura di me in modi e tempi diversi ma entrambi preziosi.

Grazie al Clan, che è stato la mia via di fuga nei primi due anni quando pensare agli Scout era un altro impegno ma anche un modo di staccare (e che adesso che non c'è più, quanto mi manca...). E grazie a Cola che ha condiviso questa Strada con me.

Grazie al Fantacani, ovviamente per i meme e piano piano per essere diventato un gruppo di amici, completamente pazzo e spesso da remoto, ma che si fa sentire.

Grazie ad Albi, LucaePiero (d'ufficio in coppia), Pippo, Vale, Leti, Sofi, Chiara, Irene, Marta e tutti i 4vecchi&90universitari perché ci sono sempre.

Grazie a Ciukì (e Popoti) e Bob, perché non sempre è facile ma insieme sicuramente lo è di più. Grazie a babbo e mamma e il perché non si scrive.

Alla fine, in fondo in fondo dove stanno le cose più importanti, grazie ad Ale. Per scrivere i perché mi servirebbe un'altra tesi, quindi va bene questa frase, che riassume tutto quello che serve:

"Lei si poggia sul tuo petto, che cos'è che vuoi di più?"

ABSTRACT

Epigenetic control of cell identity is a key factor in stem cell regulation and in how living cells can lose or re-gain their pluripotency in the processes of reprogramming and differentiation. Much is known about the role of histone marks, chaperones and remodelling machinery in this context but the importance of histone content and how it varies during these processes is much less studied. At the same time, very little is known about the regulation of histone gene transcription during cell fate transitions. Here we investigated how the lack of HMGB1, generating a "low-histone" chromatin environment in differentiated cells, impacts on reprogramming and iPSCs. We found that iPSCs can be generated and self-renew in the absence of HMGB1, while further decreasing their histone content. Furthermore, we analysed how the histone PTMs landscape is altered in iPSCs depleted of HMGB1. Eventually, we observed how the lack of HMGB1 does not alter the efficiency of the reprogramming process. As we saw a down-regulation of histone content during the road to iPSCs, we also focused our attention on transcription of histone genes. We found that histone genes transcription is not decreased in iPSCs, but to the contrary specific histone genes are highly upregulated in iPSCs. We named them "StemHistones" and characterized the function in pluripotent cells of three of them, namely H1f6, H2bu2 and H3f4. We showed that these histones are indeed chromatin-bound and that their down-regulation or complete ablation did not impact on mESCs self-renewal, but it did alter differentiation onset. Transcriptomic analysis revealed that StemHistones are involved in the regulation of differentiation, as H1f6 deletion specifically impacts on endodermal gene expression while depletion of H2bu2 or H3f4 causes downregulation of markers of all three germ layers upon differentiation.

In conclusion, we characterized the effects of HMGB1-mediated histone decrease in reprogramming, showing that it alters iPSCs chromatin but does not impact on reprogramming efficiency. We also uncovered the high expression of some unusual histone genes that we named StemHistones and we linked their function to differentiation onset. Our work highlights the importance of histone protein content and histone gene transcriptional regulation as new important epigenetic layers in the control of cell identity.

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

2i	Two inhibitors
AP	Alkaline phosphatase
BMP	Bone morphogenetic protein
bp	Base pair
CAF1	Chromatin assembly factor 1
Cd24a	Cluster of differentiation 24a
CDS	Coding sequence
CENP-A	Centromere protein A
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTCF	CCCTC-binding factor
CTD	C-terminal domain
E14.5	Embryonic day 14.5
EBs	Embryoid bodies
Eomes	Eomesodermin
FISH	Fluorescent in situ hybridization
Gata6	GATA binding protein 6
GD	Globular domain
GO	Gene Ontology
GSK3	Glycogen synthase kinase 3
H1t	Testis-specific histone H1
H3mm	H3 *Mus musculus*
H3K27me2/3	Histone H3 lysine 27 di/trimethylation
H3K36me3	Histone H3 lysine 36 trimethylation

H3K4me3	Histone H3 lysine 4 trimethylation
H3K9ac	Histone H3 lysine 9 acetylation
H3K9me1/2/3	Histone H3 lysine 9 mono/di/trimethylation
H3t	Testis-specific histone H3
Hi-C	High Chromosome Conformation Capture
HIRA	Histone cell cycle regulator
HMGB1	High mobility group box 1
HP1	Heterochromatin protein 1
ICM	Inner cell mass
iPSC	Induced pluripotent stem cell
ISWI	Imitation SWItch
KD	Knock-down
KO	Knock-out
LIF	Leukemia inhibitory factor
Mb	Mega base
MEF	Mouse embryonic fibroblast
mESC	Mouse embryonic stem cells
MNase	Micrococcal nuclease
NTD	N-terminal domain
Oct4	Octamer-binding transcription factor 4
OKSM	OCT4, SOX2, KLF4 and c-MYC
Otx2	Orthodenticle homeobox 2
PCA	Principal component analysis

PTM	Post-translational modification
RC	Replication-coupled
RI	Replication-independent
rtTA	Reverse tetracycline-controlled transactivator
SETDB1	SET Domain Bifurcated Histone Lysine Methyltransferase 1
sgRNA	Single-guide RNA
shRNA	Short hairpin RNA
SILAC	Stable isotope labeling by/with amino acids in cell culture
SLBP	Stem loop-binding protein
Sox17	SRY (sex determining region Y)-box 17
Sox2	SRY (sex determining region Y)-box 2
STAT3	Signal transducer and activator of transcription 3
TAD	Topological associated domain
tH2A	Testis-specific H2A
tH2B	Testis-specific H2B
UTR	Untranslated region
WT	Wild-type

LIST OF FIGURES AND TABLES

INTRODUCTION

1.1 Histones and nucleosomes: dynamic building blocks of chromatin

Eukaryotic cells need to fit approximately 2 meters of non-compacted DNA into their nucleus. Nuclear DNA is packaged into chromatin, a complex of DNA and proteins organized in several compaction levels. Chromatin constrains genetic material in a small volume, forming the substrate for the vital processes of replication, recombination, transcription, DNA repair and chromosome segregation. The fundamental unit of chromatin is the nucleosome, which contains a core made by 146 ± 1 bp of DNA wrapped around an octamer of histone proteins, consisting of a central H3-H4 tetramer flanked by two H2A-H2B dimers (Luger *et al*, 1997) (Kornberg, 1977). Two adjacent nucleosomes are connected through a segment of linker DNA, which can associate with the linker histone H1 (FIG. 1). Therefore, the definition of a nucleosome as the repeating subunit of chromatin comprises a core plus the linker DNA, which may be bound by a molecule of histone H1 (Hao *et al*, 2021).

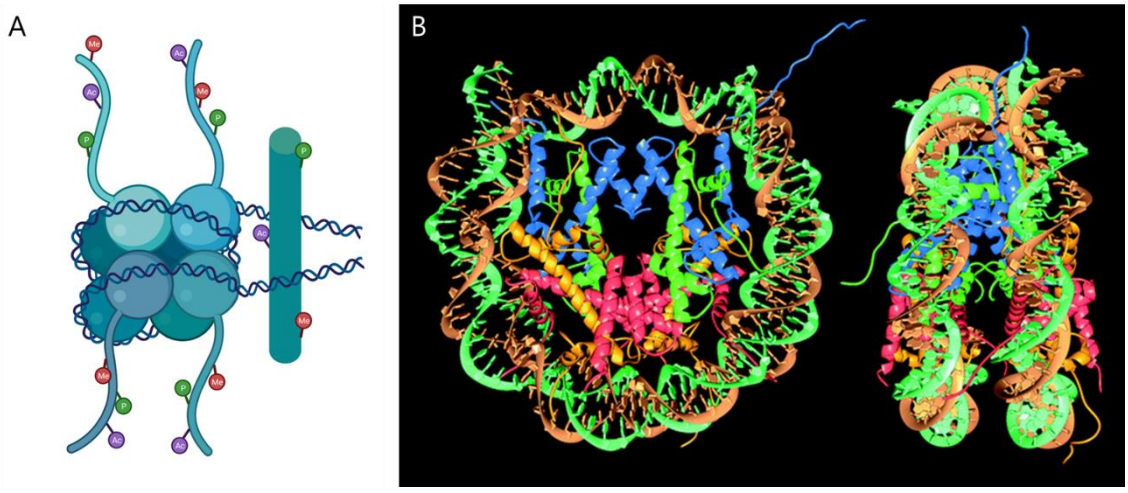

Figure 1 The structure of the nucleosome, (A) Cartoon showing the nucleosome, formed by an histone octamer and the DNA wrapped around it. Binding to linker DNA is histone H1. (B) Crystal structure of the nucleosome as originally resolved by (Luger et al, 1997).

Histones are a family of small, positively charged proteins that tightly bind the negatively charged DNA. Each of the core histone proteins (H3, H4, H2A, H2B) consists of a structured histone-fold domain involved in histone-histone and histone-DNA interactions and a unstructured N-terminal tails protruding from the nucleosome (Luger & Richmond, 1998). These tails are highly conserved across eukaryotic species and their lack leads to lethality (Ling *et al*, 1996). Their importance is due to the extensive post-translational modifications (PTMs) they can undergo, forming a substrate for numerous protein-protein interaction, changing their affinity for DNA and ultimately finely regulating the accessibility of DNA to the transcriptional machinery (Jenuwein & Allis, 2001; Strahl & Allis, 2000). Core histones have a high number of isoforms with specific functions, some of them with only 1-2 aa of difference compared to canonical proteins (see section 1.1.4).
On the other hand, linker histone H1 has 3 domains, a short N-terminus (NTD), an ~80 aa central globular domain (GD) and a long C-terminal domain (CTD, ~100 aa). The CTD is the most variable part of the protein and is the main responsible for the ability of H1 to condense chromatin (Hendzel *et al*, 2004).

1.1.1 Chromatin structure and 3D organization

Nucleosomes interact with each other to form the first level of chromatin compaction. For a long time *in vitro* electron microscopy studies visualized the first level of nucleosome-to-nucleosome interaction as an ordered structure, the 30 nm chromatin fibre (Schalch *et al*, 2005). Only recently this view was challenged by *in vivo* analysis using super-resolution microscopy techniques. Observing chromatin in living cells demonstrated that nucleosomes are instead more flexible in their organization and associate in groups called “clutches” that are very heterogenous and cell-type specific in size. This analysis confirmed the ability of histone H1 to condense chromatin, as it is associated with larger clutches (Ricci *et al*, 2015). Clutch diameter varies between 5-24 nm and chromatin is defined as a “chain”, something that is more flexible and with possible density variation compared to the classic view of a rigid fiber (Ou *et al*, 2017).

Recent progresses in Chromosome Conformation Capture techniques (3C), and in particular their high-throughput version (Hi-C), allowed to show that chromatin folds in

the 3D space of a nucleus forming "chromatin loops" to favour contact between transcriptional regulatory regions such as enhancer, promoters and gene bodies (Rao *et al*, 2014a). These loops are constrained at their borders by CTCF binding sites; they interact much more within themselves than with other loops, forming separate domains. These domains appear as squares along a diagonal in Hi-C contact maps and are now commonly referred to as topologically associated domains (TADs) (Ge *et al*, 2012). TADs can be of various sizes and vary in a cell type- and species-specific fashion, but they are now widely accepted as higher-order units of chromatin that shape functional genome organization. At the chromosomal scale, TADs tend to associate with each other according to their transcriptional status, forming so-called "A" and "B" compartments, corresponding to the classic definition of euchromatin and heterochromatin. In fact, "A" compartments contain chromatin enriched in actively transcribed genes, while "B" compartments are mostly enriched in repressive chromatin. (Rao *et al*, 2014b). The highest level of chromatin organization is the formation of chromosome territories, defined by the fact that intra-chromosomal interactions are more frequent than inter-chromosomal ones, as can be visualized by Hi-C, confirming early FISH studies (Lichter *et al*, 1988; Lieberman-Aiden *et al*, 2009).

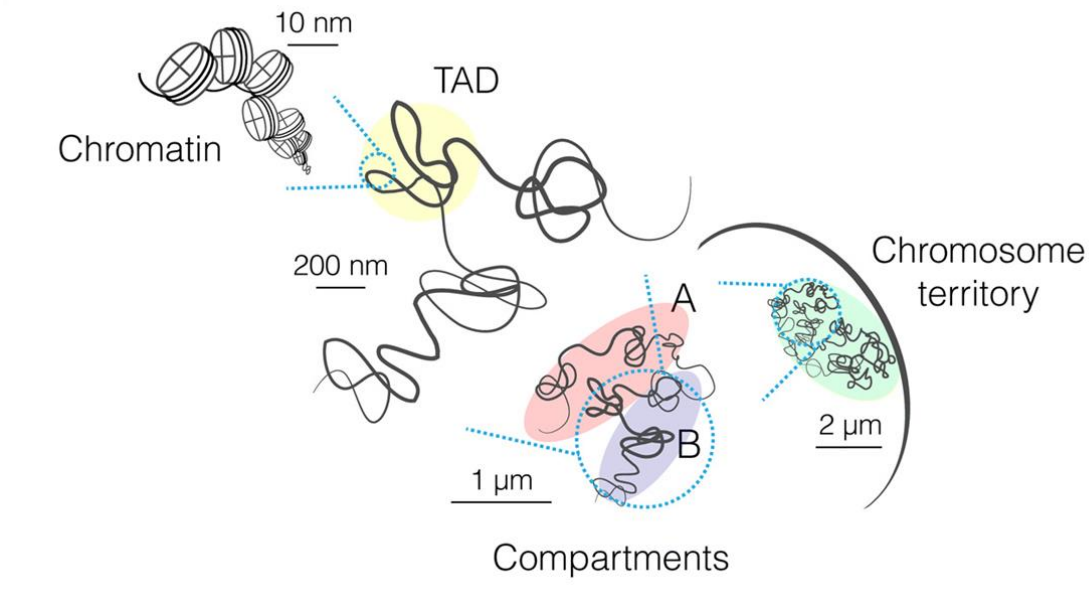

Figure 2 A scheme of chromatin folding inside the nucleus. Chromatin chains of about 10 nm associate into distinct domains called TADs, often bringing together functional elements for transcriptional activation. Driven by the active or inactive transcriptional status, TADs aggregate in "A" and "B" compartments. At even higher scale, chromosomes form distinct territories. Adapted from Szabo et al (2019).

1.1.2 Nucleosome assembly and dynamics

Newly synthesized histones need to be assembled into functional nucleosomes. This function is carried out by proteins called histone chaperones, that immediately bind histones in the cytosol and allow their translocation to the nucleus. There, CAF1 directs replication-dependent nucleosome assembly, while the HIRA complex acts throughout the cell cycle (Tagami *et al*, 2004). CAF1 is the only chaperone that can bind the canonical H3 histone, while HIRA interacts with variant H3.3, a replication-independent histone. Another key chaperone is the FACT complex, that moves together with polymerases during transcription and replication, binding both H3-H4 and H2A-H2B (Yang *et al*, 2016).

Nucleosome position is also influenced by ATP-dependent nucleosome remodelling complexes, such as ISWI, and by whether or not the H1 linker histone is present (Clapier *et al*, 2017). Other proteins interacting with linker DNA, such as high mobility group (HMG) proteins, also affect nucleosomes stability, as they can influence nucleosome spacing and DNA bending (Bonaldi *et al*, 2002).

1.1.3 HMGB1 regulates nucleosome number and occupancy

HMGB1 is a very abundant non-histone chromatin protein. It binds to DNA in the minor groove and interacts with nucleosomes at the dyad axis. It competes with histone H1, but binding of HMGB1 increases chromatin accessibility, opposed to the compaction effect of H1 binding. Moreover, HMGB1 facilitates nucleosome assembly *in vitro* but also interacts with histone H3 and can destabilize its binding to DNA, favouring nucleosome remodelling (Celona *et al*, 2011b; Watson *et al*, 2014). It appears then that HMGB1 plays a complex role in nucleosome dynamics, both stabilizing and destabilizing it. *In vivo* studies showed how cells lacking HMGB1 decrease their histone content by 20% and have a more accessible chromatin, as measured in MNase assays. The global effect on nucleosomes is somehow surprising, leading not to a redistribution of nucleosomes but rather to a reduction in their occupancy. In fact, DNA sequences stably occupied by nucleosomes stay the same, while sequences with lower occupancy in normal conditions are even more nucleosome-depleted in the absence of HMGB1 (Celona *et al*, 2011b).

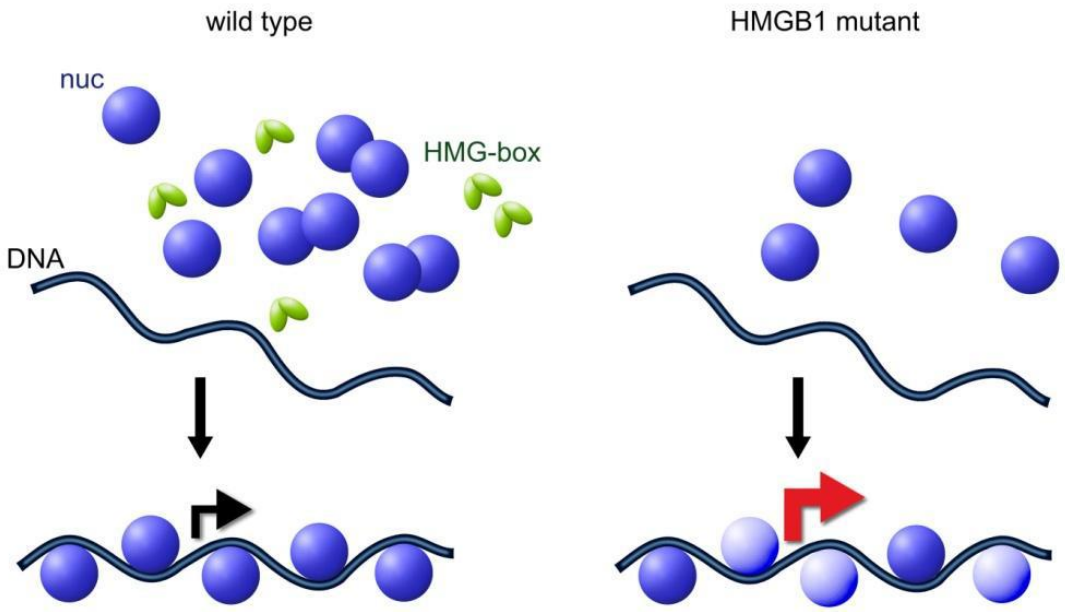

Figure 3 Changes in nucleosome occupancy in the absence of HMGB1. In HMGB1 depleted cells, nucleosomes are reduced by 20% relative to wild type. The positioning is maintained, but the occupancy decreases, resulting in transcriptional increase (red arrow). Nucleosomes are depicted as spheres, dark blue are high occupancy nucleosomes and light blue are low occupancy nucleosomes. HMGB1 protein is depicted in green. DNA is depicted as lines.

1.1.4 Histone gene organization and diversity

Histone proteins are mainly encoded by a highly complex set of genes that are expressed during S-phase and therefore are defined as "replication-coupled" (RC) or canonical histone genes. Their expression during S-phase couples DNA replication with duplication of chromatin, allowing to package the newly synthetized genome. Three different characteristics define RC histone genes in metazoans. First, they are organized in large clusters in the genome and this organization is conserved across species (Marzluff *et al*, 2002). Moreover, their mRNAs are intronless and not polyadenylated, but instead have a 3' stem-loop sequence bound by the stem-loop binding protein (SLBP), which is involved in all phases of the life cycle of histone mRNAs, including transcription termination, transport from the nucleus, translation and degradation (Battle & Doudna, 2001; Marzluff, 2005). There are additional paralogs for histones which are expressed throughout the cell cycle ("replication-independent", RI). These RI genes are not part of clusters and encode so-called "histone variants". These proteins contain several differences in their primary sequence compared to canonical histones and their mRNAs contain polyA tails and introns. Histone variants can have specific functions and their RI-status allows them to substitute canonical histones during the cell cycle, influencing nuclear processes such as transcription, DNA repair and chromatin remodelling (Talbert & Henikoff, 2021).

Here we will focus on the specific characteristics of mouse histone genes and their organization in the genome. The information presented is derived from Mouse Genome Informatics (MGI, http://www.informatics.jax.org/) and follow the recent nomenclature update (Talbert *et al*, 2012). The main reference for the organization of mouse histone genes (Marzluff *et al*, 2002) has been updated in this presentation since in recent years more histone genes have been identified; these are included in figures and analysis presented below.

There are 3 histone clusters in the *Mus musculus* genome, containing the RC histone genes (FIG. 4). *Hist1*, the major cluster, is located on chromosome 13 (13A3.1) and spans 2.2 Mb. It contains 55 genes, encoding for all the 5 histones, distributed in a non-homogenous fashion along the cluster. Approximately half of the genes are in the first part of the cluster (from H2bc13 to H2bc11) while the other half is 1.48 Mb away. Of note, all RC H1 genes (H1f1 to H1f6) are located in this cluster. Genes are scattered along

the cluster without a precise pattern; however, H2A and H2B genes frequently found as divergent with a common promoter (e.g., H2ac1 and H2bc1). *Hist2* and *Hist3* clusters are instead much smaller in size, containing respectively 10 and 4 genes. They are located on chromosome 3 (3F2.1) and 11 (11B1.3)

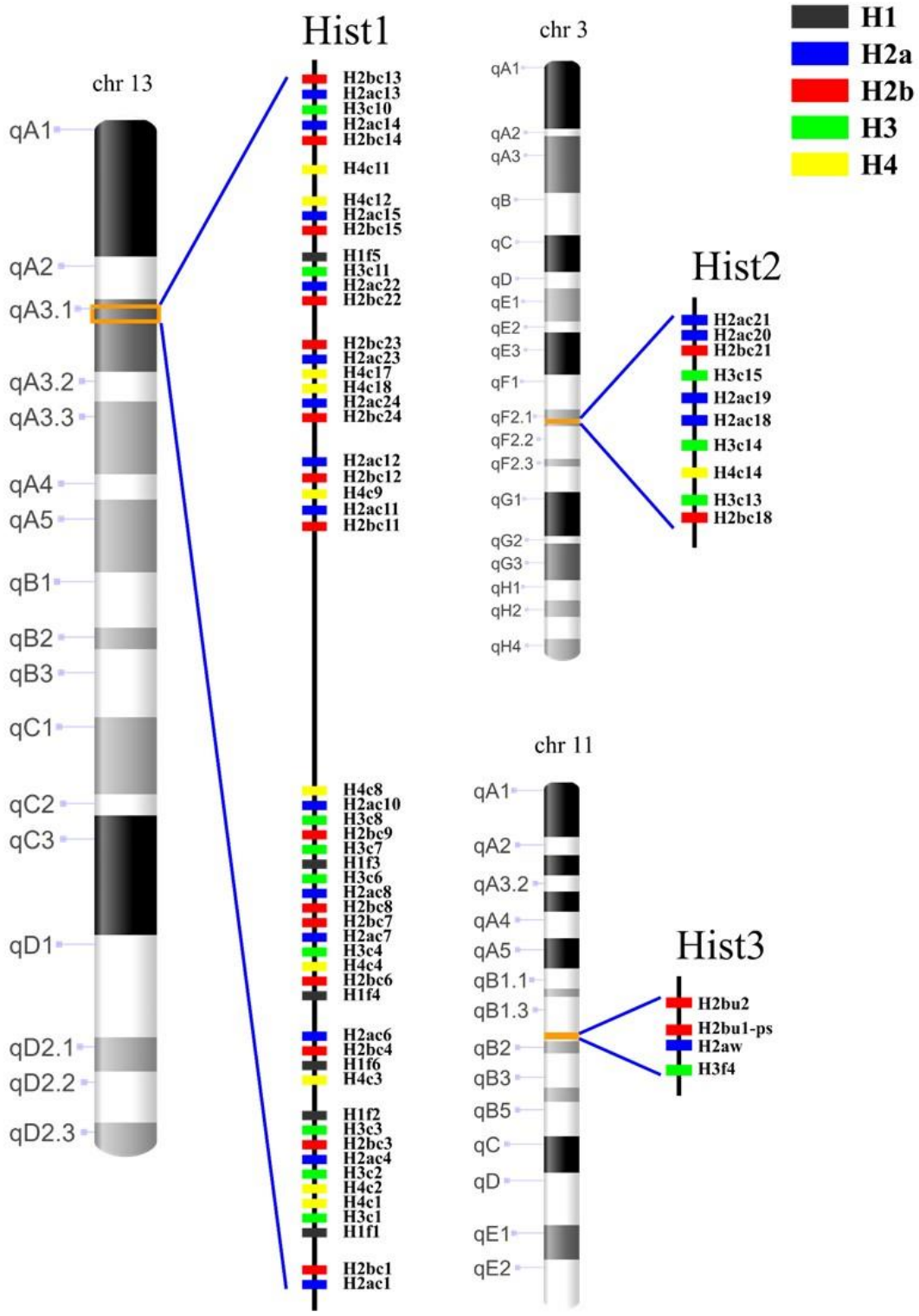

Figure 4 Localization of histone gene clusters on mouse chromosomes. The position and the order of histone genes in the three clusters Hist1, Hist2 and Hist3 in the mouse genome is shown. Genes are ordered and spaced according to their genomic location.

1.1.5 Histone isoforms

Histone proteins constitute a highly diverse group. All RC and RI histone genes can contribute to nucleosomes and chromatin and the slight differences in their aa sequences often dictate a specific function for each protein. Therefore, there are several histone isoforms for each histone type, encoded by both RC and RI histone genes. RI histone variants have been intensely studied (Talbert & Henikoff, 2021), while different isoforms in RC histone genes are less studied and understood. We will compare histone isoforms now, to present their characteristics and differences from a sequence standpoint, both at DNA and aa level.

- Histone H1

H1 isoforms are most diverse group of histones. There are 11 known H1 variants, 6 RC (H1f1, H1f2, H1f3, H1f4, H1f5, H1f6) and 5 RI (H1f0, H1f7, H1f8, H1f9, H1f10). All 5 RC isoforms are widely expressed in somatic tissues, except H1f6 that encodes the H1t variant. H1t was deemed "testis-specific" for its high expression in pachytene spermatocytes (Drabent *et al*, 1993a) but it was recently shown to be expressed in other cell types, including mESCs and cancer cells (Tani *et al*, 2016). RI variants are mostly germ cell-specific (H1f7, H1f8, H1f9), while H1f0 is expressed in differentiated cells (Zlatanova & Doenecke, 1994) but also has important roles in cancer (Torres *et al*, 2016).

There are clear differences in amino acid sequences among H1 histone isoforms, especially between germ cell-specific and somatic variants (FIG. 5). RC isoforms have the highest homology, except for H1f6 that has some key substitutions compared to the other 5. These differences lower its affinity to DNA, making it a poor condenser of chromatin (Ramesh *et al*, 2006; Khadake & Rao, 1995). Interestingly, despite constituting more than half of linker histones in pachytene spermatocytes, H1t KO does not affect germ cell formation and mouse fertility as its absence is compensated by overexpression of other H1 subtypes despite their different primary structures (Lin *et al*, 2000).

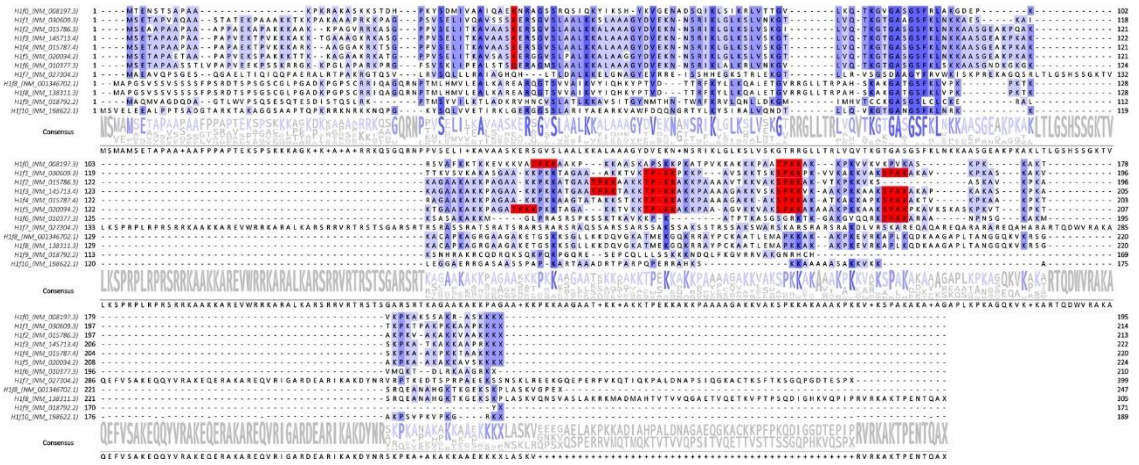

Figure 5 Multiple sequence alignment of histone H1 genes amino acid sequence. Gene names are indicated on the left. Shades of blue mark more or less conserved residues. Red boxes highlight key residues for DNA interaction, notably modified or lacking in H1f6.

- Histone H2A

The H2A family is characterized by a large number of isoforms, mostly diversified in length and amino acid sequence in the C-terminal domain, which can alter the length of DNA wrapped around the nucleosome. The contact with DNA and linker histone H1 makes the C-terminal domain a perfect candidate for structural alterations that cause functional changes in nucleosomes (Bönisch & Hake, 2012).

There are five classes of RI H2A variants: H2A.X, H2A.Z, macroH2A, H2A.W and testis-specific short H2A variants (H2A.B, H2A.L, H2A.P and H2A.Q), with roles ranging from DNA damage to transcriptional regulation and important functions in development and germ cells, especially testes (Celeste *et al*, 2002; Babiarz *et al*, 2006; Chakravarthy *et al*, 2012; Molaro *et al*, 2018).

- Histone H2B

H2B RC histone genes encode several different isoforms with only a few amino acid substitutions, mainly in the globular domain. Their roles are not well studied except for some like H2BE, encoded by H2bc21, which has an alternative, polyadenylated mRNA expressed only in olfactory neurons (Santoro & Dulac, 2012). From a sequence standpoint, DNA sequence alignment of all H2B genes in mouse show very high

similarity in coding sequences (CDS), while 5' and 3' UTRs diverge. This is compatible with possible different post-transcriptional regulation mechanism and, from a practical viewpoint, allows targeting single genes via genome-editing techniques, such as CRISPR-Cas9 (FIG. 6A). Coherently, amino acid sequence is quite homogenous, with some exceptions such as H2bc1, encoding for testis-specific variant tH2B, and H2bu2, which contains some unique changes like a K6R substitution in the N-terminal domain. (FIG. 6B). H2bu2 encodes for an H2B isoforms named H2B.U, which has no characterization in the literature.

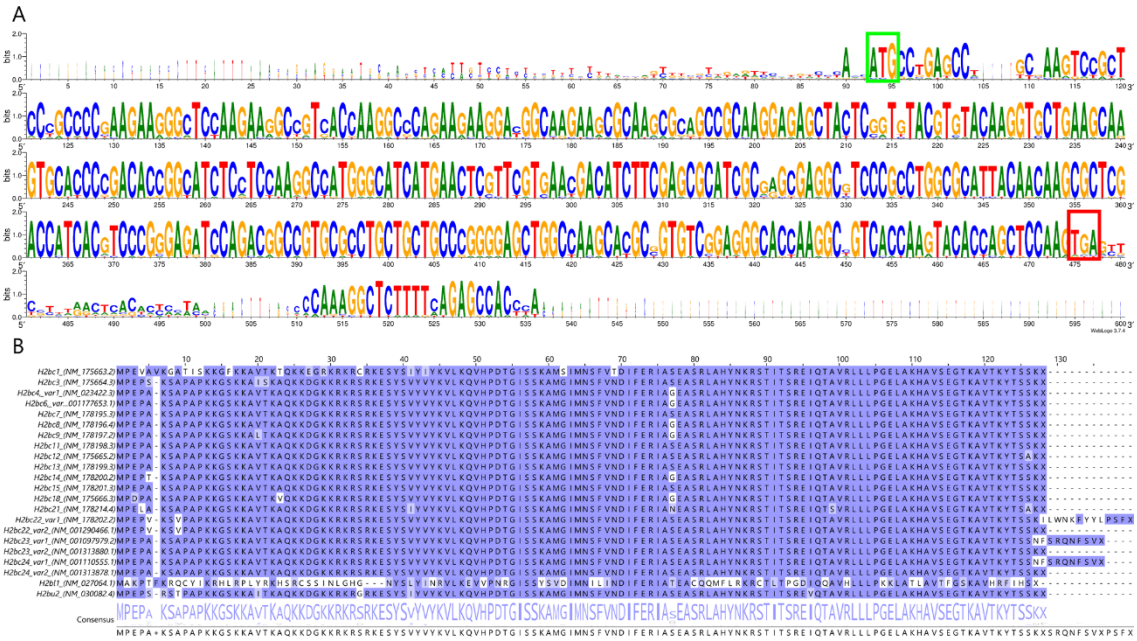

Figure 6 Multiple sequence alignment of histone H2B genes. (A) DNA logo from alignment of all H2B clustered histone genes. Letter size correlate with conservation at each position. Green box highlight start codon, red box stop codon. (B) Multiple sequence alignment of histone H2B genes amino acid sequence. Gene names are indicated on the left. Shades of blue mark more or less conserved residues.

- Histone H3

The histone H3 family consists of three main protein classes: the RC histones H3.1 and H3.2, the RI variant H3.3 and the histone H3-like centromeric protein A (CENP-A). Moreover, there are also a few recently identified isoforms, such as the RC histone H3t and H3.3-related RI variants H3mm, which have tissue specific expression and functions (Maehara *et al*, 2015). The CDS of these genes share high homology, while 5' and 3' UTRs diverge (FIG. 7A). Of note, the only conserved motif in the 3' UTR corresponds to the stem-loop required for histone mRNAs post-transcriptional regulation.

Histones H3.1 and H3.2 are encoded by multiple, clustered RC genes. They have only one amino acid difference (FIG. 7B). The recently identified H3t isoforms, encoded by H3f4, is also RC and contains 4 substitutions compared to H3.1-2. These changes make H3t-containing nucleosomes more flexible, which appears to be important in early stages of spermatogenesis. In fact, mice lacking H3t have azoospermia due to loss of haploid germ cells (Ueda *et al*, 2017).

H3f3a, H3f3b and H3f3c encode for the H3.3 variant, which differs from H3.1 and H3.2 by only 4-6 amino acids. In particular, H3.3 contains a "AAIG" motif in place of the "SAVM" motif in RC H3, specifying a different chromatin assembly pathway mediated by different histone chaperones: CAF1 for RC H3, HIRA or ATRX-DAXX for H3.3 (Ahmad & Henikoff, 2002; Tagami *et al*, 2004; Elsässer *et al*, 2012). In mouse, H3.3 shows HIRA-mediated enrichment at active genes, promoters and enhancers, while it is also enriched in heterochromatic loci such as telomeres and imprinted genes through ATRX-DAXX deposition.

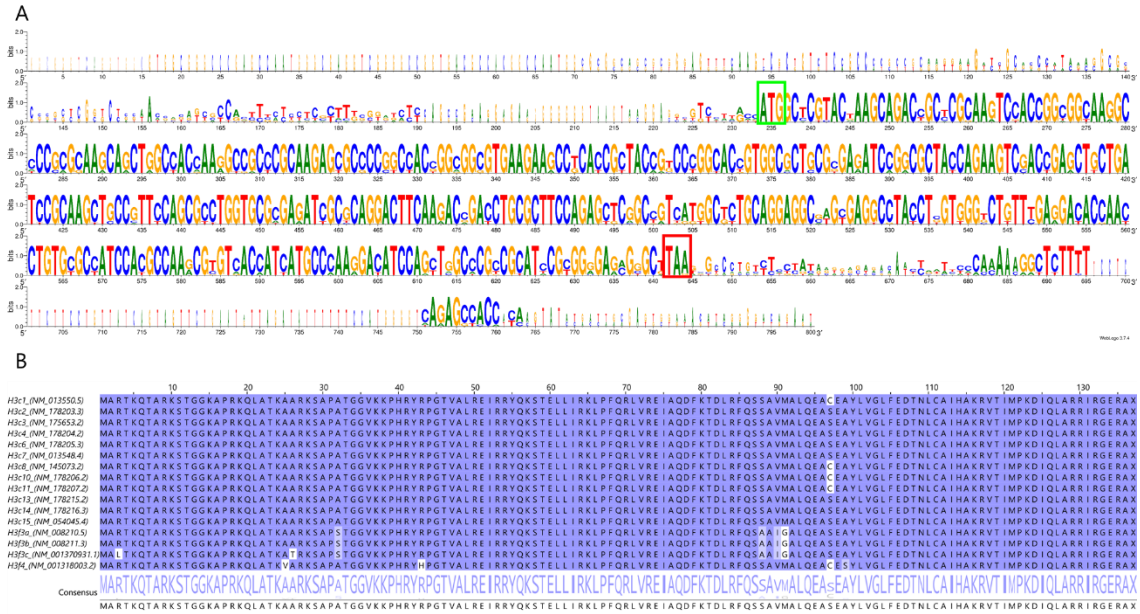

Figure 7 Multiple sequence alignment of histone H3 genes. (A) DNA logo from alignment of all H3 histone genes. Letter size correlate with conservation at each position. Green box highlights start codon, red box stop codon. (B) Multiple sequence alignment of histone H3 genes amino acid sequence. Gene names are indicated on the left. Shades of blue mark more or less conserved residues.

- Histone H4

H4 does not have any known histone variant in mouse, while there are recent reports of a variant in humans named H4G (Long *et al*, 2019). H4 is anyway the most uniform histone protein across species, which probably suggest a strong selective pressure to conserve its sequence and avoid any variation. Like H2B, H4 only contacts other histones, while H3 and H2A also make homodimeric contacts (Luger *et al*, 1997). This differences in core histone positioning inside the nucleosome likely influenced the evolutionary forces and resulted in less or more variety at the protein level.

1.1.6 Post translational modification of histones

Histones and their N-terminal tails undergo extensive PTMs. The most well-known modifications consist in the attachment of small chemical groups, like methylation and acetylation, but there are also PTMs involving proteins like ubiquitin. Two very elegant genetic experiments paved the way for understanding the importance of histone tail modifications: progressive deletions of H4 N-terminus in budding yeast by the Grunstein

lab in the 90s showed the importance of acetylation for promoter activation, while replacement of the whole histone cluster with mutated genes in *Drosophila* proved that Polycomb complex targets H3K27 for methylation (Pengelly *et al*, 2013; Durrin *et al*, 1991). These pioneering studies opened the way and made histone PTMs study a new exciting field to explore. To date, more than 100 different histone modifications have been described, with many having pretty much unknown functions (Tan *et al*, 2011).

Histone PTMs influence chromatin in two different ways. They can act on the interface between DNA and nucleosomes, where modifications that neutralize the positive charge of histones weaken their interaction with negatively charged DNA or hamper histone-histone interaction (Shogren-Knaak *et al*, 2006; Dorigo *et al*, 2003). Histone PTMs can also create platforms for protein-protein interaction, allowing specific factors to recognize modified histone tails. For example, protein containing bromodomains recognize acetylated lysines, while chromodomains bind methylated residues (Yun *et al*, 2011).

Most PTMs happen on the histone tails, which are protruding from the nucleosomes and can interact with other nucleosomes in proximity, therefore influencing chromatin structure and compaction. Recently, PTMs on the globular domain have also been described that can directly impact on nucleosomal stability (Tropberger & Schneider, 2013).

- Acetylation

Histones are acetylated at specific lysines, one of the most abundant residues in their tails. In the tail of histone H4, 4 different lysines can be acetylated, masking their positive charge and therefore weakening the interaction with DNA. In the case of acetylation, the number of modified residues is more impactful than the specific position of the modification, as shown by elegant studies in yeast for both H4 and H3 (Martin *et al*, 2004; Dion *et al*, 2005). Histone hyperacetylation is in fact known to characterize regions of actively transcribed chromatin, which are loosened and "open" and therefore favour the access of the transcriptional machinery to DNA.

- Methylation

Lysines can be mono-, di- or trimethylated without affecting their charge. Interestingly, methylation of different lysines have different effects on transcription,

adding a layer of complexity to the behaviour of histone PTMs, compared to acetylation. In fact, H3K4me3 and H3K36me3 are associated with active transcription, localizing on promoters and gene bodies respectively. On the contrary, H3K9me2/3 and H3K27me3 are associated with heterochromatin and transcriptional silencing (Zentner & Henikoff, 2013). H3K9me3 is an interesting case of how histone PTMs can act as a platform: in fact, it is recognized and bound by the chromodomain-containing proteins HP1, which can oligomerize and recruit histone methyltransferase to "write" new H3K9me3 and spread heterochromatic domains along the DNA (Grewal & Elgin, 2002).

Moreover, different methylations can combine their effects and co-exist even when they have opposite effects. This is the case of bivalent promoters, which are present in mES cells in specific developmental genes that have both H3K4me3 and H3K27me3. Having this double mark, these genes are poised for transcriptional activation, but they can also be rapidly shut down upon specific differentiation signals from the environment (Bernstein *et al*, 2006a).

- Phosphorylation

Phosphorylation is the most common modification for signal transduction in the cell. It is also present on histones, where it can mask lysine positive charge similarly to acetylation. Probably the best studied histone PTM is indeed a phosphorylation event happening at serine 139 of histone H2A.X, known as γH2A.X. This is one of the most precocious DNA damage markers, and it recruits the machinery for DNA repair where double-strand breaks occur (Paull *et al*, 2000). In addition, it can also recruit ATM and ATR kinases which are responsible for its phosphorylation, therefore leading to spreading of the damage signalling.

Another prominent example is phosphorylation of H3S10. This serine residue is immediately downstream of H3K9, responsible for HP1α recruitment upon trimethylation. H3S10 phosphorylation leads to release of HP1α from chromatin during mitosis; this is yet another instance of how histone PTMs form a complex and interconnected network in chromatin regulation (Hirota *et al*, 2005).

1.2 Studying cell identity: pluripotency, OKSM reprogramming and differentiation

1.2.1 Mouse embryonic stem cells as model of in vitro pluripotency

Pluripotency is the capacity of cells to produce all somatic lineages and is a transient state of cells comprising the inner cell mass (ICM) in the pre-implantation embryos at blastocyst stage. These cells can be isolated and expanded *in vitro*, giving rise to mouse embryonic stem cells (mESCs), which represent, in the environment of cell culture, an immortalized status of the naïve epiblast. Interestingly, mESCs maintain a number of features of pluripotent cells, as they can be expanded in culture without genetic transformation but can also be injected back into a blastocyst and contribute to all cell types in an embryo (Bradley *et al*, 1984). Therefore, they represent an invaluable tool to study chromatin regulation and analyse transcriptional network of pluripotency in an easily accessible cell culture system.

mESCs cells can be cultured in vitro, but to prevent their differentiation they were initially cultured on a feeder layer of irradiated fibroblasts, which secrete myeloid leukemia inhibitor factor (LIF) (Williams *et al*, 1988). LIF activates STAT3, which in turn promotes ES cell viability and suppresses differentiation, in combination with either serum or BMP (Matsuda *et al*, 1999; Martello *et al*, 2013). mESCs cultured in this condition are maintained in a pluripotent state but they are quite heterogenous for Nanog expression and can spontaneously exit self-renewal and differentiate (Singh *et al*, 2007). These cells are thus in the condition of “primed” pluripotency, as opposed to “naïve” pluripotency, a more defined and stable mESCs status that can be obtained using defined medium conditions. In particular, intense research in the early 2000s showed that inhibition of Mek/Erk and GSK3 pathways (“2i”, two inhibitors) disfavors mESCs differentiation, without involving the LIF/Stat3 pathway. Furthermore, 2i culturing conditions yield uniform cultures of mESCs, homogenously expressing Nanog, completely abolishing spontaneous differentiation happening in serum+LIF conditions. Therefore, culturing in 2i creates a “ground” state of pluripotency and can also increase the efficiency of differentiated cells reprogramming (Silva *et al*, 2008; Ying *et al*, 2008).

A variety of "pluripotency factors" defines and governs the ES cell state, centered around Oct4, Sox2 and Nanog (Martello & Smith, 2014). The first discovered transcription factor involved in the regulation of pluripotency was Oct4 (Okamoto *et al*, 1990). Interestingly, it has been demonstrated that increasing the expression of Oct4 leads to differentiation, proving that a tight regulation is required to maintain pluripotency (Niwa *et al*, 2000). Oct4 interact with Sox2, acting as a heterodimer that binds DNA and promotes the transcription of genes controlling pluripotency including Oct4 itself (Chen *et al*, 2008). Oct4 and Sox2 expression is not restricted to pluripotent cells, though, since they are also expressed during the first phases of differentiation. On the contrary, the homeodomain protein Nanog has a specific expression in human and mouse pluripotent cells, and it has been demonstrated that elevated levels of Nanog are sufficient to confer to ES cells the ability to self-renew at clonal density in the absence of LIF (Chambers *et al*, 2003). Oct4, Sox2 and Nanog are therefore the master regulators of a complex network of transcription factors, but also non-coding RNAs and microRNAs, that coordinately are required for pluripotency (Guttman *et al*, 2011). Oct4, Sox2 and Nanog are complementary and only partially overlapping in forming an autoregulatory circuitry to maintain the pluripotent state of ES cells (Young, 2011).

1.2.2 Going back to pluripotency: OKSM reprogramming

Stem cells have the special ability to both self-renew and differentiate into other cell types, both in the embryo and in the adult organism. Whenever they are committed to a certain lineage, they cannot re-acquire the ability to self-renew or be converted into a different cell type. This paradigm was initially challenged by John Gurdon in 1962 through the establishment of somatic nuclear cell transfer, demonstrating how nuclei of differentiated cells could be reprogrammed when transferred into enucleated frog oocytes (Gurdon, 1962). Moreover, the key discovery of somatic cell reprogramming by overexpression of Oct4, Sox2, Klf4 and cMyc (OKSM factors) by Takahashi and Yamanaka in 2006 opened a completely new field (Takahashi & Yamanaka, 2006). By simply overexpressing these factors, a variety of differentiated cell types can be converted back to pluripotency, establishing induced pluripotent stem cells (iPSCs). Although the reprogramming process is highly inefficient, the last 15 years of research pushed the limit of the initial OKSM cocktail, uncovering new factors, drugs and mechanism to increase

the efficiency of reprogramming (di Stefano *et al*, 2016; Vidal *et al*, 2014; Bar-Nur *et al*, 2014).

In particular, the main obstacle towards an efficient reprogramming is represented by epigenetic barriers. Cell identity is mostly maintained by a complex network of signals stored in chromatin, which result in defined transcriptional outputs characterizing a specific cell type, as we described for pluripotent cells. Forcing the cell to switch identity requires erasure of the epigenetic signature and chromatin rewiring. Therefore, much of the effort has focused on trying to understand the epigenetic barriers that limit reprogramming efficiency. These can be divided into two groups: a first one comprising factors that block transcription and whose inhibition facilitates pluripotency genes activation; and a second one with factors that activate transcription and help somatic cells to define their identity. Repressive factors include a number of heterochromatin-associated proteins like DNA and histone methyltransferases (DNMT3a/b, Polycomb group proteins, Setdb1), HP1, and histone chaperone CAF1 (Sridharan *et al*, 2013; Cheloufi *et al*, 2015). On the other hand, chromatin-associated factors with a positive effect on transcription include DOT1L, histone chaperone FACT and histone SUMOylation (Onder *et al*, 2012; Kolundzic *et al*, 2018; Cossec *et al*, 2018).

Furthermore, analysis of how nucleosomes are positioned in the genome also revealed how the primary structure of chromatin changes when cells change their identity. Indeed, nucleosome profiling of pluripotent and somatic cells showed that regulatory regions of mESCs-associated genes have lower nucleosomal occupancy (Teif *et al*, 2012; West *et al*, 2014). Therefore, nucleosome organization is an additional epigenetic layer contributing to cell type-specific gene regulation.

1.2.3 Going out of pluripotency: differentiation and in vitro models

Pluripotent cells exist in the embryo for less than 24 hours, around the time the embryo is implanted and then proceeds towards germ cell specification and gastrulation.

mESCs require approximately the same time to exit from the pluripotent state upon LIF withdrawal. During this time, they immediately start to lose expression of ground state pluripotency markers such as Rex1 (Zfp42), Esrrb and Nanog, while Oct4 and Sox2

expression is sustained as they are needed for differentiation onset (Trott & Arias, 2013). They play a dual role: both are needed for pluripotency maintenance, where they feed-forward their own expression and inhibit germ layer differentiation; on the other hand, they are needed by cells to transition out of pluripotency, because they can change their binding partners and contribute to activation of differentiation-specific gene sets. For example, Oct4 associates with Otx2, a marker of epiblast induced early in the transition phase both *in vitro* and in the embryo (Buecker *et al*, 2014). At this point, transitioning cells can be induced in different cell lineages in response to extrinsic factors. For example, exposure to retinoic acid primes them for neural differentiation, while Activin induces to endodermal fate (Kubo *et al*, 2004; Gajović *et al*, 1998).

In order to mimic *in vitro* the complexity of the differentiation process, mESCs can be differentiated in 3D structures that resemble some aspects of embryonic development, called embryoid bodies (EBs). In EBs cells start to differentiate but also send signals to each other and self-organize in a spheroid structure with an outer layer of primitive endoderm (Hamazaki *et al*, 2004). Upon exposure to specific conditions, EBs can be further organized and even develop anteroposterior polarity, specify germ layers and undergo gastrulation (van den Brink *et al*, 2014; ten Berge *et al*, 2008). Therefore, EBs are very useful *in vitro* model to test function of specific genes in development and differentiation (Brickman & Serup, 2017).

1.3 Chromatin organization and cell fate identity

We discussed how pluripotent cells maintenance and somatic cell reprogramming are dependent on specific transcriptional networks, tightly interconnected to the chromatin status. Many factors and histone PTMs contribute to cell identity maintenance, with both positive and negative transcriptional outputs. We will focus here on specific examples of particular relevance for my thesis work.

1.3.1 H3K9me3 and CAF1 bridge heterochromatin with nucleosome assembly

A key player in cell identity is H3K9me3, which demarcates so-called "constitutive" heterochromatin and is enriched during differentiation. The main role of H3K9me3 is silencing of repetitive regions in the genome, where it creates large domains of transcriptional repression by oligomerization (Canzio *et al*, 2011). Moreover, H3K9me3 demarcates gene clusters containing key pluripotency genes such as Nanog, Sox2 and Rex1 that are repressed in differentiated cells. Therefore, successful OKSM reprogramming requires erasure of these K9me3 signatures on the somatic cell genome in order to proceed. Indeed, decreasing the expression of histone methyltransferases SUV39H1/2 and SETDB1 accelerates reprogramming (Soufi *et al*, 2012). Low H3K9me3 is by itself a defining element of pluripotent cells, as they require a high plastic and accessible chromatin to be able to maintain their self-renewal and full developmental potential. Disrupting this equilibrium impairs mESC metastate and leads to differentiation (Savić *et al*, 2014).

H3K9me3 repressive role is also due to its ability to create a platform to recruit additional factors and spread heterochromatic domains, in particular through HP1 proteins. These are bound and recruited onto chromatin by specific domains of p150, one of the subunits of histone chaperone CAF1 (Quivy *et al*, 2008). Importantly, mutating the HP1-interacting domains of CAF1 does not impair its ability to assemble nucleosomes.

Recent work has uncovered a particularly interesting role for CAF1 in cell identity maintenance, as it is both a major roadblock on to iPSCs formation and also a general stabilizer of cell identity. CAF1 depletion favours cell transdifferentiation, unlocks totipotent-like features in mESCs and increase OKSM reprogramming efficiency (Cheloufi *et al*, 2015; Ishiuchi *et al*, 2015). This double-sided role of CAF1 implies that

chromatin replication is coupled with conservation of heterochromatic domains. As depletion of CAF1 induces such destabilization in cell identity, it is tempting to speculate that interfering with nucleosome assembly and chromatin repression could be a general mechanism to favour reprogramming and transdifferentiation.

1.3.2 Histone variants

The functional role of histone variants has gained a special focus in recent years (reviewed by Martire & Banaszynski, 2020). There are various mechanisms allowing histone variants to affect chromatin composition. First, the destabilization of the nucleosome particle through incorporation of variants with substitution of key residues for histone-histone or DNA-histone interaction. Second, the specificity of histone variants for distinct chaperones creating novel interaction networks. Third, the diverse array of PTMs that different histone variants can carry. These different factors can be combined and generate a whole set of novel chromatin features that make histone variants key players in chromatin function.

In the context of cell identity and embryonic development in particular, there are very well studied examples such as H3.3, H2A.Z and macroH2A, alongside other emerging or less studied factors (tH2A and tH2B, H3mm) (FIG. 8).

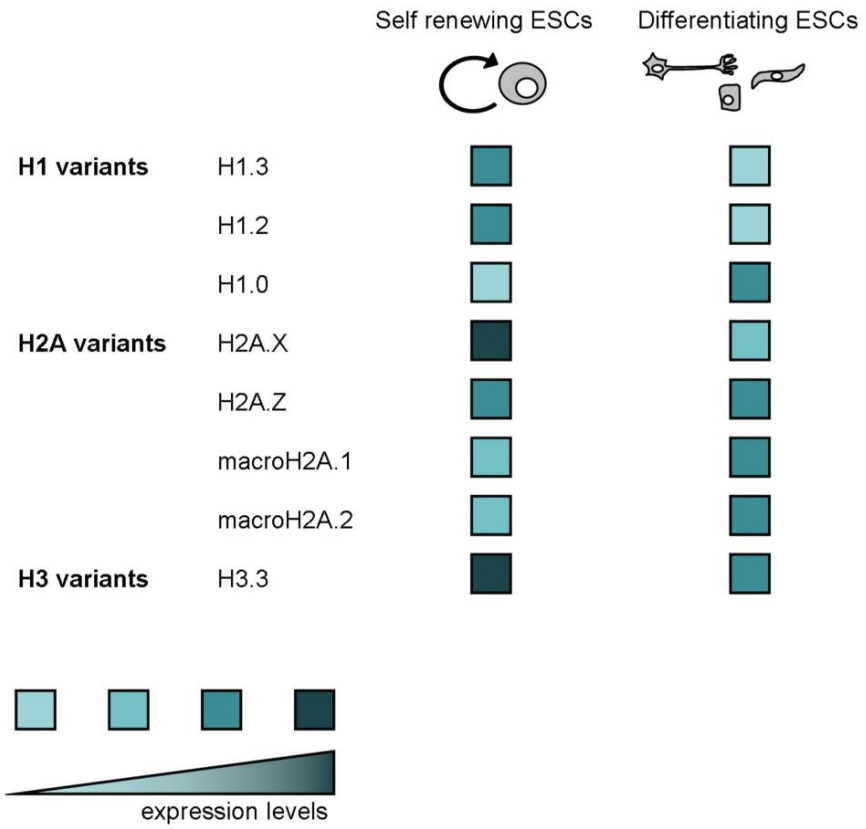

Figure 8 Schematic of histone variants expression levels in self-renewing ESCs and differentiating cells. From Turinetto & Giachino (2015).

They constitute interesting and diverse examples of how incorporation of variants can impact on cell identity:

- H3.3 is generally associated with active transcription, but in mESCs it is also present at bivalent promoters and enhancers (Martire *et al*, 2019; Banaszynski *et al*, 2013). Interestingly, the lack of H3.3 impedes correct mESC differentiation, as a single substitution on its N-terminal tail compared to H3.1 allows an H3.3-specific phosphorylation that stimulates enhancer acetylation and developmental gene activation.
- H2A.Z is incorporated in nucleosomes at actively transcribed genes, but is also needed to repress genes upon differentiation. This double role is due to its ability to generate stable and unstable nucleosomes, depending on its coupling with H3.1 or H3.3 (Park *et al*, 2004; Jin *et al*, 2009). Therefore, its function is strongly context dependent, relying on different binding partners and the surrounding chromatin environment (Hu *et al*, 2013).

- macroH2A is associated with heterochromatin and PRC2 and specifically incorporated into pluripotency genes in differentiated cells. Its depletion during OKSM reprogramming destabilizes silencing via H3K27me3 therefore increasing the efficiency of pluripotency acquisition (Gaspar-Maia *et al*, 2013).
- tH2A and tH2B are testis- and oocyte-specific variants, that when expressed together create an open-chromatin environment. Their overexpression during OKSM reprogramming increased the efficiency of iPSCs generation, through a non-sequence specific mechanism that caused a generally more accessible chromatin environment (Shinagawa *et al*, 2014a).
- H3mm are recently identified mouse-specific H3.3 isoforms (Maehara *et al*, 2015). Interestingly, two isoforms named H3mm7 and H3mm18 were reported to have key roles in muscle regeneration and differentiation. Both generate a somehow unstable nucleosome structure through specific substitutions, impacting on correct transcriptional activation (Harada *et al*, 2018; Hirai *et al*, 2022). These variants represent an extreme example of the importance of small sequence variation in histones, as they are species-specific sub-variants of H3.3, and still they can impact tissue-specific events.

Linker histones H1 represent an exception in this landscape, as they are of course not part of the nucleosome core particle but nonetheless impact on cell identity maintenance. Somatic variants have partially redundant roles, as only a triple-KO of H1.2, H1.3 and H1.4 impacts on mESCs biology by not correctly suppressing pluripotency genes upon differentiation onset (Zhang *et al*, 2012). Deleting single genes do not impact on mice viability nor differentiation capacity of pluripotent cells, suggesting redundancy and compensation mechanisms among the various variants (Fan *et al*, 2001a). Moreover, H1 somatic variants were mapped genome-wide in mESCs and showed clear depletion at TSS and gene-rich regions, while they were enriched in heterochromatic regions, in particular major satellites, confirming their role in promoting chromatin compaction (Cao *et al*, 2013). On the same line, it was recently demonstrated that somatic H1 can be SUMOylated and contribute to silencing of totipotency-related genes (Sheban *et al*, 2022). Lastly, among the other

histone H1 isoforms, testis-specific histone H1t has recently been found expressed in mESCs, demonstrating how accuracy in defining the tissue-specificity of histone variants is still an issue (Tani *et al*, 2016; Hayakawa *et al*, 2020a). H1t apparently targets rDNA repeats, differently from somatic variants, but its mechanism of action remains to be elucidated.

AIM OF THE WORK

Histones and chromatin exert a fundamental role in stabilising cell identity. Histone PTMs and the chromatin assembly machinery are widely studied for their function as roadblocks of somatic cell reprogramming and their ability to fine-tune gene expression in pluripotency and differentiation. On the other hand, the relevance of total histone content and the contribution of specific histone genes to these processes are much less studied.

Here we aimed at providing new insights in the role of histones in pluripotent cells and reprogramming from two different perspectives.

On one side, we tried to bridge the idea that histone protein content is reduced in iPS cells compared to differentiated ones, namely MEFs in our study, with the phenotype of histone reduction that we observed in MEFs lacking HMGB1. We therefore employed Hmgb1 KO MEFs as a "histone-low" differentiated cell type and investigated how this characteristic could impact on iPS cells formation and on their chromatin. Moreover, we analysed the effect of the lack of HMGB1 on the chromatin of pluripotent cells.

On the other side, we were interested in looking at single histone genes and their transcription, because the micro-heterogeneity of these genes and the contribution they give to chromatin has recently emerged as an important layer of chromatin and transcriptional regulation. Recent work highlighted the role that new unusual variants can have in various biological processes, such as spermatogenesis, cancer and muscle biology. On this line, we identified three histone isoforms with a specific upregulation in pluripotent cells ("StemHistones") and decided to characterize them to understand whether their expression is linked to a specific role in mESCs. We studied how StemHistones depletion impact on pluripotency and differentiation by transcriptomic analysis and generated new tools that will allow further dissection of StemHistones function.

Overall this thesis work provided novel insights in histone biology, investigating the role in cell reprogramming of Hmgb1 and of the decrease in histone content and identifying new histone variants involved in differentiation onset.

RESULTS

2.1 Hmgb1 KO and histone protein content in iPS cells and reprogramming

Hmgb1 KO MEFs contain a reduced amount of histone proteins and therefore have a more open chromatin conformation, compared to WT MEF (Celona *et al*, 2011b; Almeida *et al*, 2018). Similarly, also pluripotent cells such as mESC have a more open chromatin conformation compared to differentiated cells, including MEFs (Tonge *et al*, 2014). We therefore decided to investigate whether the accessible chromatin status induced in MEFs lacking Hmgb1 would make them easier to reprogram to iPS cells, given the resemblance of their chromatin to mESC and iPS cells. Moreover, we investigated how chromatin is affected by the lack of Hmgb1 in pluripotent cells.

2.1.1 iPS cells contain less histone protein than MEF but do not downregulate histone transcription

It was previously reported that mESC have a lower histone protein content than their differentiated counterparts, such as MEF or EBs derived from spontaneous differentiation (Karnavas *et al*, 2014).

To confirm this observation in iPS cells, we reprogrammed MEFs derived from OKSM+/+ rtTA+/+ Oct4GFP+ triple transgenic mice ("OKSM triple" mice, from now

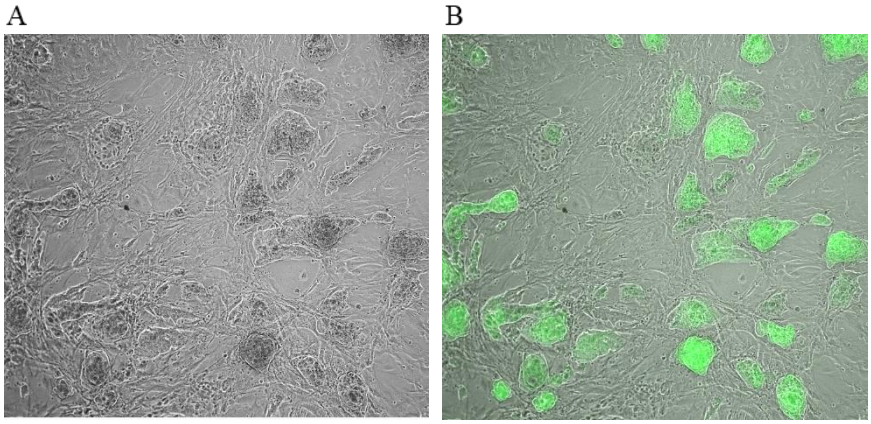

Figure 9 iPSCs colonies express Oct4GFP transgene. Phase-contrast (A) and GFP fluorescent (B) images of reprogrammed iPSCs colonies derived from OKSM triple MEFs

on.) This so called “secondary reprogramming system” allows reprogramming of a population of genetically homogenous cells upon exposure to doxycycline (Carey *et al*, 2009), which induces OKSM expression in all cells. Fully reprogrammed iPS cells colonies can be screened taking advantage of the Oct4GFP transgene, which is switched on only when cells become fully pluripotent.

We therefore derived iPS cell lines by culturing MEFs from E14.5 OKSM triple embryos in presence of 1 µg/ml dox. Fluorescent imaging confirmed the expression of Oct4GFP (FIG. 9) and successful reprogramming. Then, we compared the levels of histone H3 protein in MEF and iPS cells derived from them. Given that core histones protein are essentially equimolar in the cell, we used H3 as a proxy for all core histone content in the

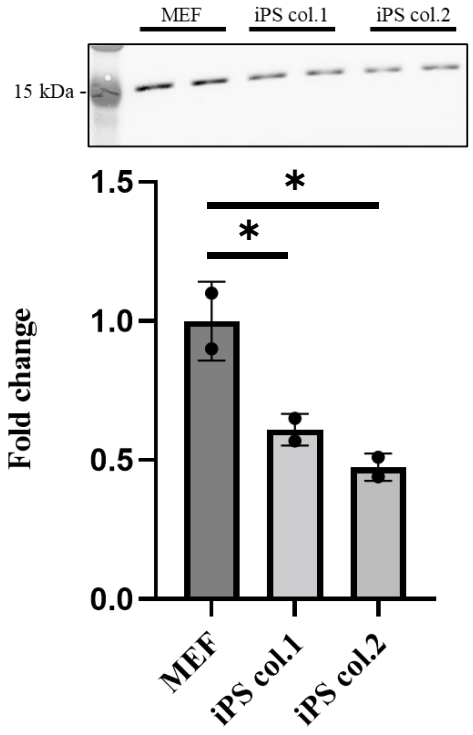

Figure 10 MEF contain 40% more histone H3 than iPSCs colonies derived from them via OKSM reprogramming. A representative western blot of histone H3 in total cell lysates from MEF and 2 independent iPSCs colonies. Lysates contained equal amounts of DNA (200 ng). The histogram shows the values for 2 technical replicates with their standard deviation as error bars. iPSCs have about 40% less histone H3; (*) p-value<0.05, unpaired t test. The experiment was repeated 3 times with similar results.

cell. We observed a large reduction of about 30-40% in two independent iPS cell lines (FIG 10).

On the other hand, we measured histone gene (HG) transcription via RT-qPCR for H3 and H2B, using primer pairs common to all genes for each histone (pan-histone primers), and we could not observe a significant difference between MEF and iPSCs (FIG. 11A).

We confirmed this result performing total RNAseq on MEFs and iPSCs. Differential gene expression analysis revealed that the vast majority of histone genes are upregulated in iPS cells, but there are only minor changes in expression for the most expressed genes for all 4 core histones and histone H1 (FIG. 11B). In particular, most of the genes are slightly upregulated (fold change <2) and only among the less expressed genes (mean of counts $< 10^2$) there are 6 strongly upregulated genes (fold change >2, highlighted in blue). Among highly expressed genes (mean of counts $>10^3$), H1f1 and H1f0 are highlighted, their fold change corresponding to data previously reported in the literature about their expression in pluripotent and differentiated cells (Terme *et al*, 2011). The expression of H1f1 and H1f0 confirms the accuracy of our RNAseq experiment.

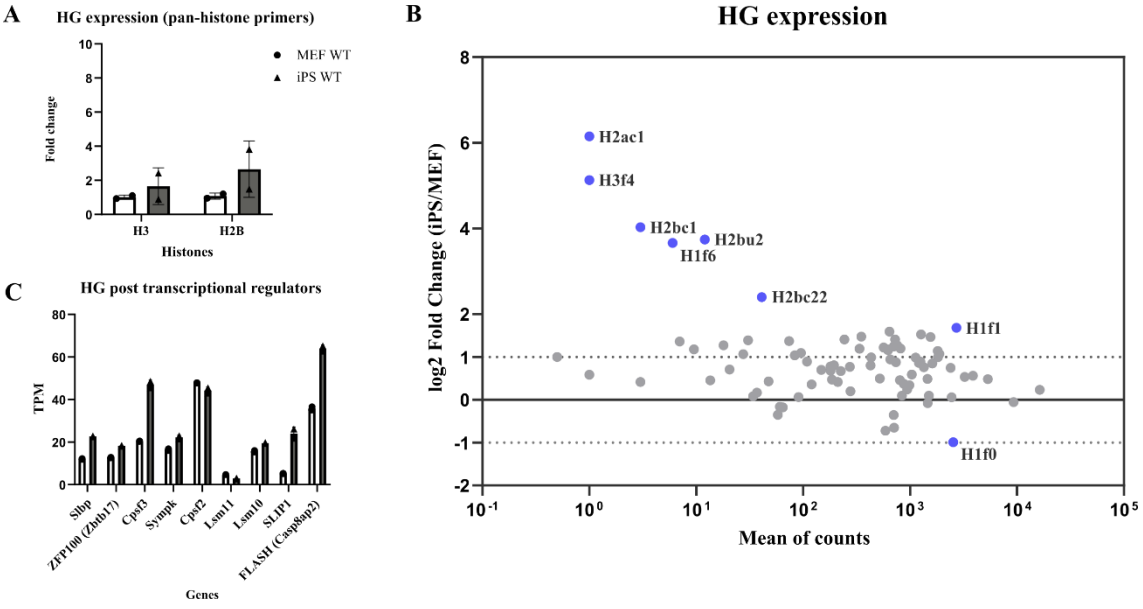

Figure 11 Histone gene (HG) transcripts in MEFs and iPS cells do not mirror the decrease at protein level. (A) RT-qPCR analysis of H3 and H2B mRNA levels detected using pan-histone gene primers in MEF and iPS cells. Histograms represent fold change of mean values over GAPDH with standard deviation, single points are biological replicates. (B) Dot plot of RNAseq data for HG expression in MEFs and iPSCs. Grey points represent HG with log-fold change MEF/iPS between -1 and +2. Blue points depict HG with a fold change below -1 or over +2, plus H1f1 and H1f0 which were flagged for biological relevance. The X axis represents mean of counts for RNAseq reads of MEF samples. (C) Transcript per million (TPM) values from RNAseq for genes involved in HG post transcriptional regulation. Legend as per (A).

Given the specificity of histone genes post-transcriptional regulation, (reviewed in Marzluff et al, 2008) we also checked the expression levels of genes involved in this process, but we could not observe variations explaining the decrease of core histones in iPS cells (FIG. 11C). A few genes where instead upregulated in iPS cells, in agreement with the higher fraction of iPS cells in S-phase compared to MEFs (Fujii-Yamamoto *et al*, 2005). Therefore, we concluded that transcription is not the cause of the histone reduction, which is likely dependent on events happening at the post-transcriptional level.

These results show that iPS cells contain less histones than MEFs, in agreement with results previously reported (Karnavas *et al*, 2014), but the mechanism of this reduction in iPS cells is not dependent on modulations at the transcriptional level.

2.1.2 Hmgb1 KO MEF generate bona fide iPS cells

We just described how iPS cells contain less histone proteins than MEFs and we know that MEFs lacking Hmgb1 contain less histones than WT counterparts (Celona *et al*, 2011b). We speculated that Hmgb1 KO MEFs with a reduced histone content could be "primed" for becoming iPS cells, given the similar decrease in histone levels that we found in iPSCs compared to MEFs. For this reason, we decided to generate transgenic embryos lacking Hmgb1 from our OKSM triple mice. We crossed them with *Hmgb1*+/- mice available in our lab and after two generations we obtained litters of OKSM triple embryos with 2, 1 or 0 WT alleles for Hmgb1 (Hmgb1 WT, Hmgb1 HET, Hmgb1 KO respectively). We obtained MEFs from OKSM triple Hmgb1 WT, Hmgb1 HET and Hmgb1 KO embryos and reprogrammed them as previously described, using 1 µg/ml dox.

We were able to establish multiple independent iPS cell lines that all show GFP signal, indicating successful establishment of the pluripotent status, irrespective of the presence or absence of Hmgb1 (FIG 12).

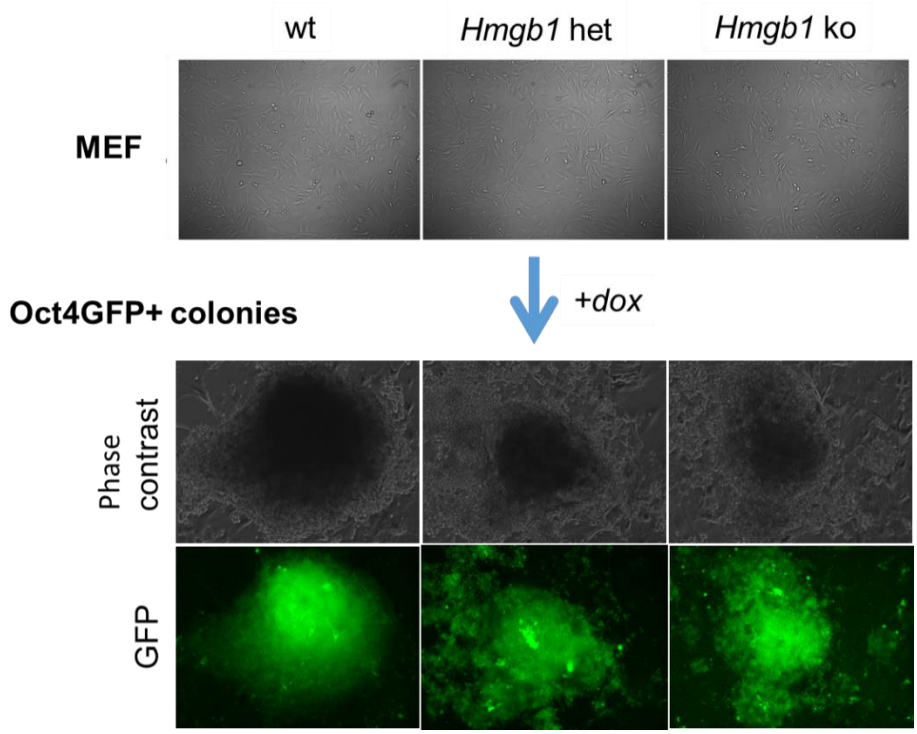

Figure 12 OKSM triple MEF can generate iPSCs colonies positive for Oct4GFP irrespective of *Hmgb1* genotype. (A) Phase-contrast image of MEF at day 0 of reprogramming. (B) Phase-contrast and (C) GFP fluorescent images of single iPSCs colonies after 15 days of culture in dox, at the end of reprogramming.

Transcriptional analysis via RT-qPCR confirmed the expression of pluripotency gene Nanog in iPS cell lines (FIG. 13A) and immunofluorescence staining showed homogenous expression of Oct4 and Nanog at protein level (FIG. 13B), irrespective of Hmgb1 genotype.

We concluded that lack of Hmgb1 does not alter the activation of the Oct4GFP reporter nor the expression of pluripotency factor Nanog, resembling correct iPS cells reprogramming. .

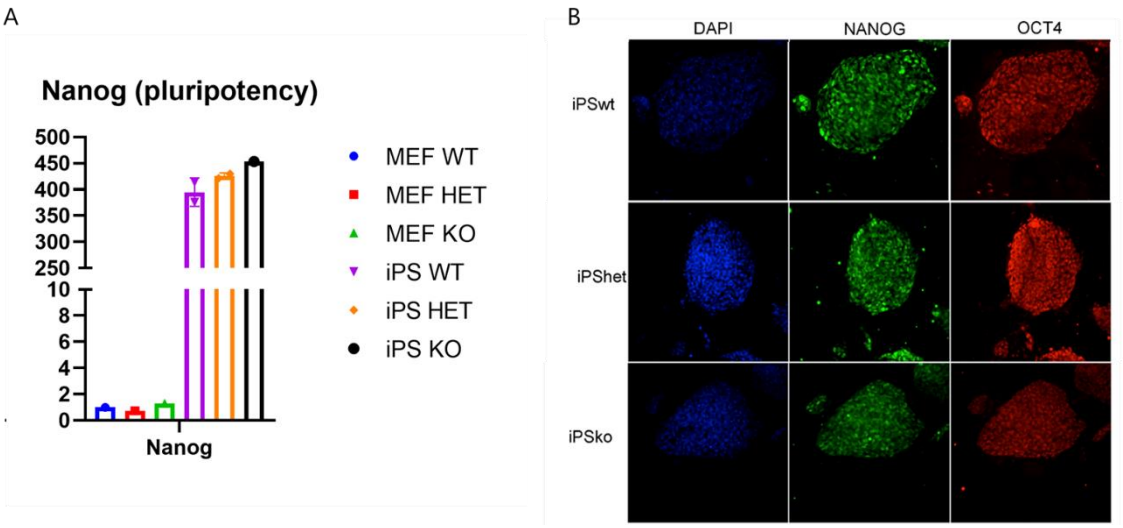

Figure 13 *Hmgb1* genotype does not influence iPSCs pluripotency. (A) RT-qPCR analysis of mRNA levels of Nanog in MEF and iPSCs with the indicated Hmgb1 genotype. Histograms represent mean ΔCt (Ct gene of interest – Ct housekeeping ATP5F1) and standard deviation of 2 biological replicates. (B) Immunofluorescent staining of iPSCs colonies, with the indicated *Hmgb1* genotype. DAPI staining of nuclei in blue, Nanog staining in green and Oct4 staining in red.

2.1.3 Hmgb1 KO iPS cells contain less histones and histone PTM profile is altered

To characterize our Hmgb1 KO iPS cells at the chromatin level, we wanted to precisely measure their histone content and analyse variations in histone post-translational modifications (PTMs).

Super-SILAC is a technique developed for accurate quantification of proteomes, based on isotopical labelling (Geiger *et al*, 2010). Recently, it was adapted for the quantification of histone PTMs in murine tumor tissues and cell lines (Noberini *et al*, 2018). Since the technique involves universal labelled spike-in proteins, it is possible to use it for accurate quantification of histones in any kind of cell type of mouse origin, taking advantage of its higher accuracy and reproducibility compared to classic proteomics approach. We then quantified the total amount of core histone proteins in our iPS cell lines, WT and KO for Hmgb1. The analysis revealed a decrease in the amount of H3 and H4, with a non-significative tendency for H2A and H2B, in Hmgb1 KO iPSCs (FIG. 14A). Considering all core histones together, there is a significant difference between Hmgb1 WT iPSCs and Hmgb1 KO iPSCs, which reduced their histone content by about 20% (FIG. 14B).

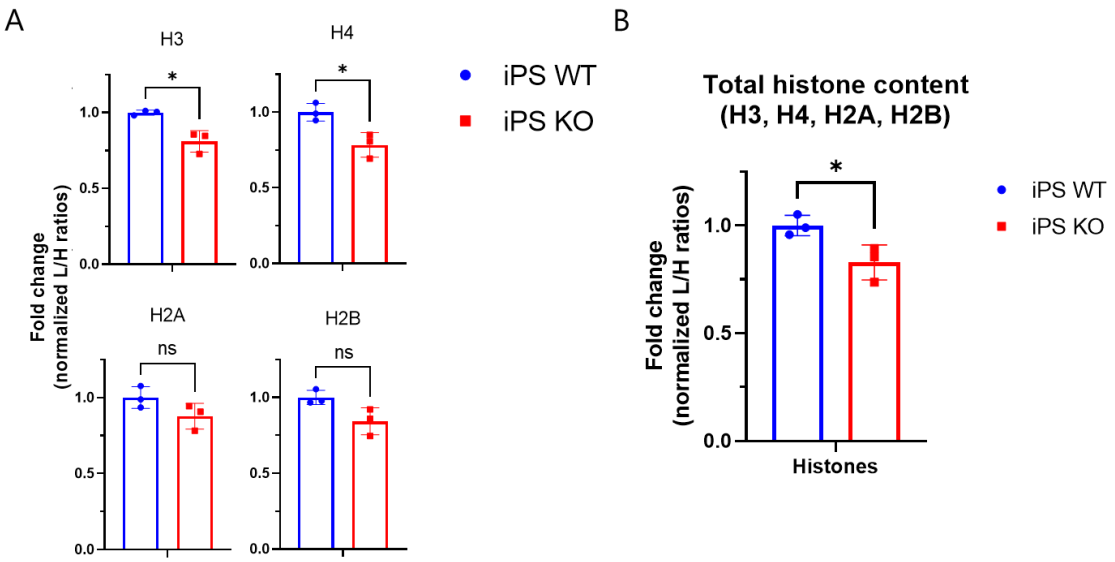

Figure 14 Super SILAC analysis of histone levels show reduction of core histones in iPSCs *Hmgb1* KO. (A) Fold change of iPSCs KO over iPSCs WT of normalized light/heavy ratios quantified using Super SILAC. For each histone 2 or 3 peptides were chosen and analysed. Bar plot represents mean values with standard deviation. (*) $p<0.05$, (ns) $p>0.05$, unpaired t-test. (B) Fold change of iPSCs KO over iPSCs WT for all histones together. Bar plot represents mean values with standard deviation.

To further investigate how the lack of Hmgb1 affected chromatin composition, we also employed Super-SILAC to profile some of the most well-studied histone PTMs (FIG. 15). In particular, we analysed the distribution of modifications occurring at histone H3 lysines 4 (H3K4), 9 (H3K9) and 27 (H3K27), together with histone H4 acetylation at 4 different lysines (K5, K8, K12, K16).

H3K4 PTMs (FIG. 15A) change their relative abundance, with a slight increase in mono-methylation (H3K4me1) and a substantial drop in tri-methylation (H3K4me3), normally associated with promoters of actively transcribed genes (Bernstein *et al*, 2006b). The unmethylated and di-methylated forms are unchanged. Among other marks of active chromatin, H4 acetylation also drops in Hmgb1 KO iPSCs (FIG.1B), with all grades of aceylation (1ac, 2ac, 3ac, 4ac) markedly decreased compared to Hmgb1 WT iPSCs. The unmodified form is instead more present, compensating for the lack of acetylated peptides.

H3K9 PTMs are also deeply affected in Hmgb1 KO iPSCs (FIG. 15C): H3K9me3, marking constitutive heterochromatin (Lachner *et al*, 2001), is strongly depleted, while both H3K9me1 and H3K9me2 are increased. The latter is associated with transcriptional

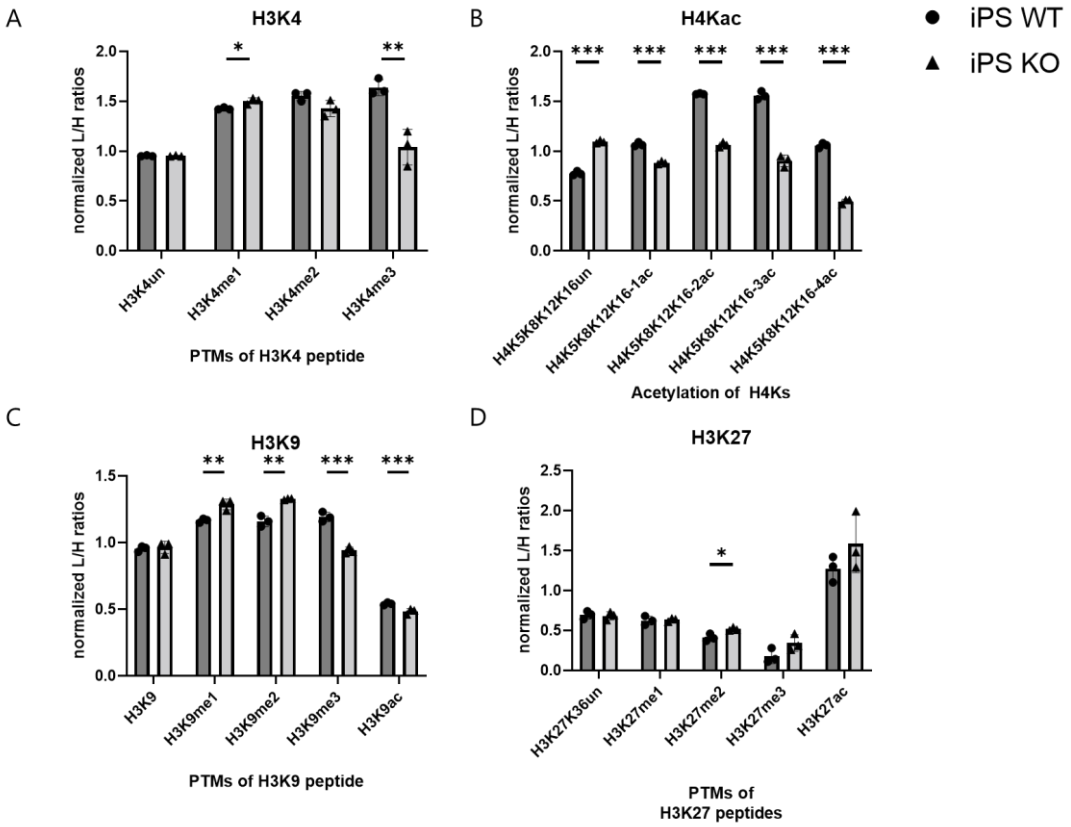

Figure 15 Super SILAC analysis of histone PTMs in iPSCs. All histograms represent normalized light/heavy ratios quantified using Super SILAC, each bar is a single peptide in iPSCs WT (circles) or iPSCs KO (triangles). Bars are mean of biological replicates depicted as circles or triangles, with standard deviation. (***) $p<0.001$ (**) $p<0.01$ (*) $p<0.05$, unpaired t-test. (A) Quantification of H3K4 peptides with indicated methylation statuses. (B) Quantification of acetylation status for H4 lysines between aa 4-16. (C) Quantification of H3K9 peptides with indicated methylation status. (D) Quantification of H3K27 peptides with indicated methylation status.

repression of key genes involved in various biological processes, including cell fate transitions and reprogramming (Rodriguez-Madoz *et al*, 2017). H3K27 PTMs are instead more stable (FIG. 15D): H3K27me2 shows a mild upregulation but overall there is no significant change between Hmgb1 WT and KO iPSCs.

Therefore, in Hmgb1 KO iPSCs we observed a decrease in the total level of core histones as we expected from previous observations (Celona *et al*, 2011b). Moreover, we could detect a general decrease in marks for active chromatin (H3K4me3 and H4

acetylation), while repressive marks presented a more complex picture. H3K27 PTMs are essentially

unaltered, but H3K9 PTMs are rearranged: we found a strong depletion of H3K9me3 and to a lesser extent of H3K9ac, while H3K9me1 and H3K9me2 are increased.

Overall, these changes do not point to a single coherent direction, but they clearly show that the absence of Hmgb1 impacts on the normal assembly and regulation of chromatin in iPS cells.

2.1.4 Lack of Hmgb1 does not alter reprogramming efficiency

Given the observations that we presented in this paragraph on how both HMGB1 and differentiation impact on chromatin and especially on total histone levels, we decided to investigate how MEFs lacking HMGB1 would reprogram to iPS cells. In particular, we asked whether lack of HMGB1, causing a decrease in histone levels, could increase the efficiency of the reprogramming process. To address this question, we reprogrammed in parallel OKSM triple MEF WT or KO for *Hmgb1*, in multiple replicates to improve reproducibility. We also included MEF heterozygous (HET) for *Hmgb1* to verify how the lack of one allele would impact on the process. We cultured MEFs at low density in the presence of 1 µg/ml doxycycline for 15 days and then we assessed reprogramming efficiency by looking at alkaline phosphatase staining or the activation of the Oct4GFP reporter. The staining for alkaline phosphatase gave a more homogenous signal and allowed a better quantification (FIG. 16). We could not detect any significant difference in reprogramming efficiency between Hmgb1 WT and KO OKSM triple MEFs, as we observed some experiment showing a higher efficiency in Hmgb1 KO MEFs but others giving a higher AP staining colony count for Hmgb1 WT MEFs instead.

Therefore, we concluded that the lack of HMGB1, despite decreasing histone content, did not alter the number of AP positive colonies. Of course, we cannot rule out the presence of partially reprogrammed cells that could impact on the overall reprogramming efficiency..

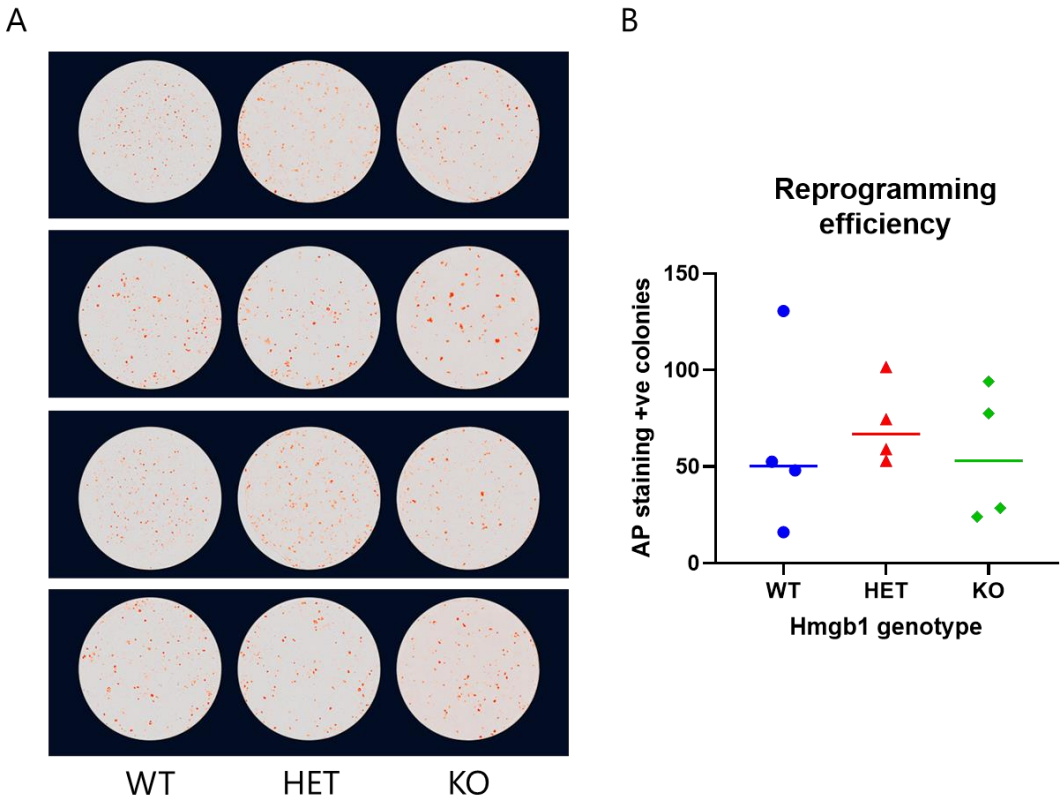

Figure 16 Alkaline phosphatase staining of MEFs lacking HMGB1 and reprogrammed to iPSCs cells. (A) Fluorescent images of alkaline phosphatase staining on Hmgb1 WT, HET and KO reprogramming experiments. (B) Quantification of reprogramming efficiency by counting of positive colonies in (A). Counting was performed using ImageJ, each dot in the dot plot represents a biological replicate.

2.2 Histone transcription is altered during OKSM reprogramming

Histone proteins show a downregulation over the course of OKSM reprogramming and histone gene transcription seems to be also affected. Histone transcription at the single gene level has not been thoroughly investigated, especially for canonical histones, because the general idea in the field is that these genes are all expressed at the same time during cell division, to supply the daughter cells with enough histones to package their newly-replicated DNA. The high degree of identity of canonical histone proteins suggests that they are functionally redundant, and therefore possible differences between individual genes have been largely overlooked. Given that in our total RNAseq in FIG. 3 a shift in histone transcription between MEFs and iPSCs is apparent, and in particular some individual genes are strongly upregulated, we decided to further investigate the matter.

2.2.1 Transcriptomic analysis: histones genes are differentially expressed during reprogramming

We started off by analysing publicly available total RNAseq datasets from studies on OKSM reprogramming. We indeed needed datasets of total RNAseq and not just polyA-enriched RNAseq since canonical histone genes are not polyadenylated. We found two RNAseq datasets (Amlani *et al*, 2018; Knaupp *et al*, 2017) available through Gene Expression Omnibus under accessions GSE106332 and GSE101905, respectively. We analysed them to investigate histone genes transcription changes in MEFs and iPSCs.

In particular, we first performed a PCA analysis on dataset GSE106332 based only on histone gene expression (FIG. 17) to assess whether it could allow us to separate MEFs from iPSCs. Indeed, we observed clear clustering of iPSCs samples (pink dots), which are also closer to Stage 1 and Stage 2 samples (two intermediate statuses of OKSM reprogramming) and well distinguished from MEFs (green dots). This confirms the idea that MEFs and iPSCs have a different “histone signature” from a transcriptional standpoint, meaning that the overall transcription of histone genes is different in the two cell types.

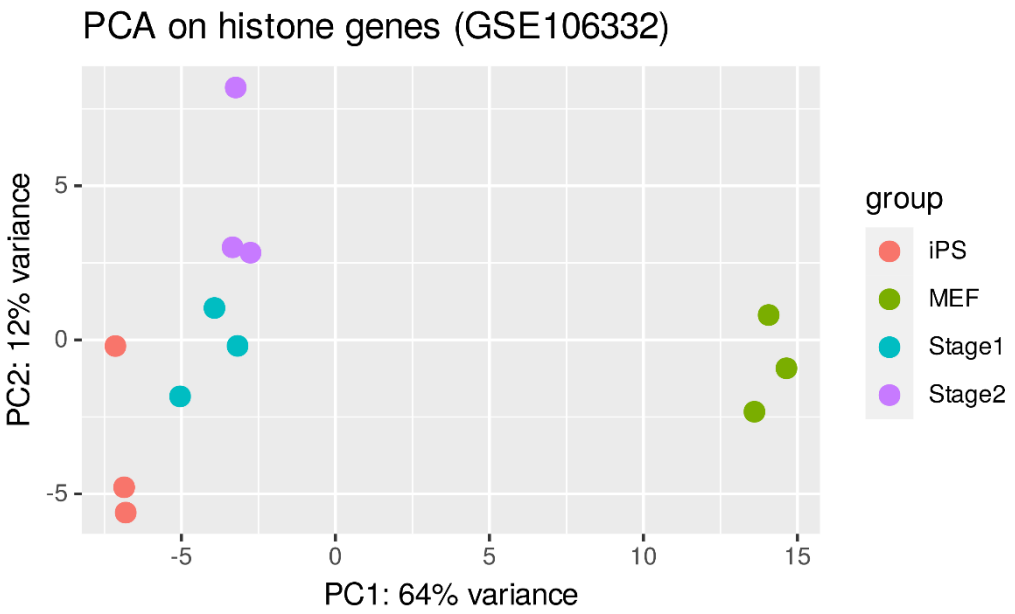

Figure 17 MEF and iPSCs have different "histone signatures". PCA analysis on histone gene expression identify two clearly separated clusters corresponding to MEF and iPSCs and the first two principal components explained about 76% of the total variation.

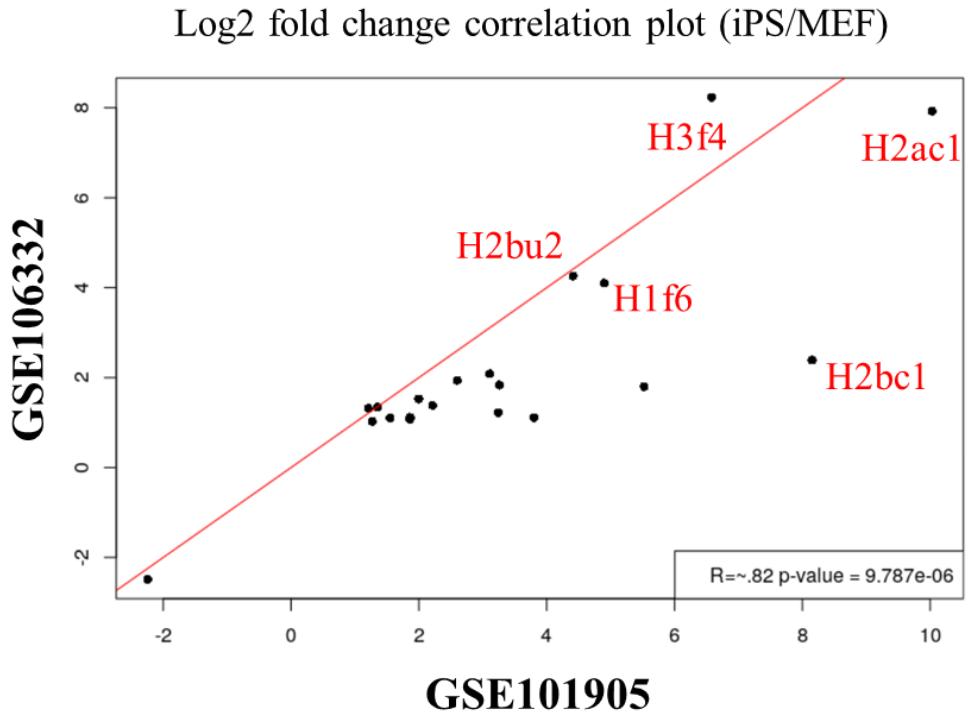

Figure 18 Five histone genes are differentially overexpressed in iPSCs. Correlation plot of log2 fold change of iPSCs over MEF of two public RNA-seq datasets identified 5 histone genes overexpressed in iPSCs. The Y axis represents log2 fold change for dataset GSE106332, the X axis for dataset GSE101905.

We then asked which histone genes were upregulated in iPS cells compared to MEFs and we took in account both datasets to get more robust results. Therefore, we calculated the log2 fold change value for each histone gene in iPS cells compared to MEFs in both datasets and we compared these values to find histone genes that are coherently down- or up-regulated (FIG. 18). We found 5 genes to be consistently and strongly upregulated (log2(iPS/MEF)>2). These genes are H1f6, H2bu2, H3f4, H2ac1 and H2bc1. Interestingly, these results were also coherent with the total RNAseq we performed on our MEFs and iPSCs, shown in FIG. 3. Overall there was a nice overlap of histone genes upregulated in all 3 datasets (FIG. 19), including the group of 5 that we just mentioned. We named these histone genes "StemHistones" for their specific upregulation in iPS cells.

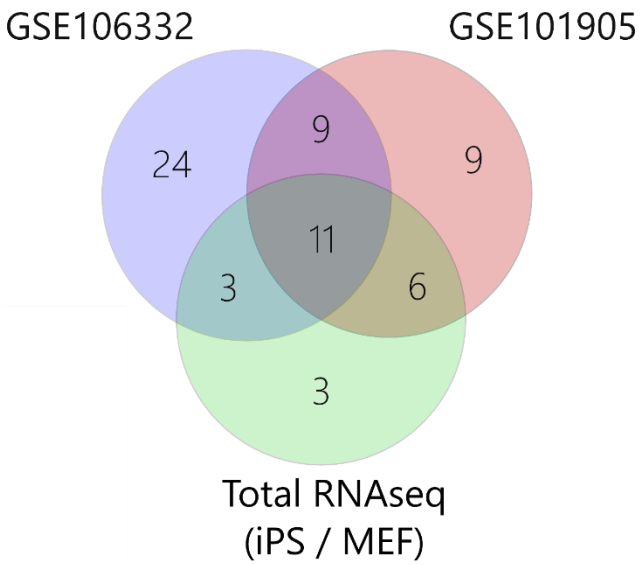

Figure 19 A common subset of histone genes is upregulated in all 3 datasets. Venn diagram showing histone genes that are upregulated in iPSCs compared to MEF according to log2 fold change for all three datasets. 11 genes are consistently upregulated, including StemHistones genes.

2.3 StemHistones depletion affects ES cells differentiation

All five StemHistones genes encode for histone proteins that differ from canonical isoforms of H2A, H2B and H3, while H1f6 is one of many isoforms of histone H1 that has a high degree of sequence divergence. Little is known about the role of these histone isoforms, especially in the context of pluripotent stem cells (see Introduction, chapter 1.3.2), save for tH2A and tH2B, encoded by H2ac1 and H2bc1, that have already been studied in the context of pluripotent cells and reprogramming (Shinagawa *et al*, 2014b). We then decided to focus our attention on H1f6, H2bu2 and H3f4. While H1f6 and H3f4 are reported in the literature, especially regarding their roles in testis and spermatogenesis (Hayakawa *et al*, 2020b; Mahadevan *et al*, 2020; Ueda *et al*, 2017), there is no information about H2bu2.

Interestingly, all these three StemHistones genes are encoded by canonical replication-dependent histone genes and not by replication-independent variant genes. This aspect drove our interest even more, since the functional diversity of these genes is particularly understudied.

2.3.1 Constitutive StemHistones downregulation in ES-E14 mouse embryonic stem cells

We verified that StemHistones genes are upregulated in mESCs, in addition to iPS cells, by performing RT-qPCR (FIG. 20). Then we decided to use mESCs as a model system to study StemHistones in the context of pluripotent cells, since they are easier to handle and manipulate.

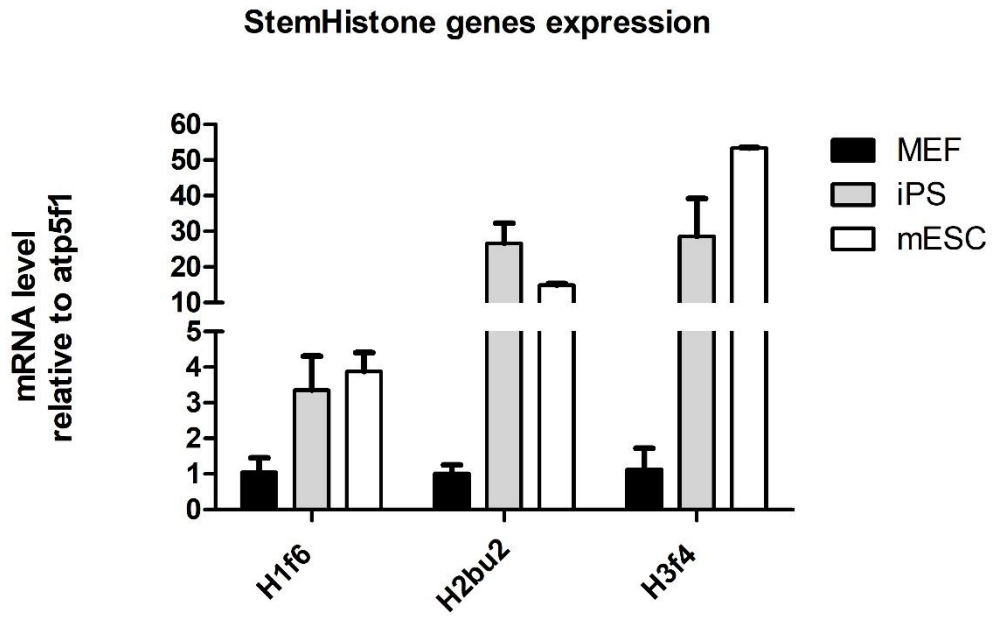

Figure 20 StemHistones genes are overexpressed also in mESCs. RT-qPCR analysis of H1f6, H2bu2 and H3f4 mRNA levels detected in MEF, iPSCs and mESCs. Histogram represents mean ΔCt (Ct gene of interest – Ct housekeeping Atp5f1) and standard deviation of 3 technical replicates.

To start StemHistones characterization, we depleted them via shRNA-mediated RNA interference. We cloned integrating lentiviral constructs bearing a specific shRNA sequence for each StemHistone (H1f6, H2bu2, H3f4) and a scrambled shRNA (control) as a negative control. The shRNA hairpin sequences specific for StemHistones transcripts were designed to target respectively the 3' UTR of H1f6 and the coding sequence (CDS) of H2bu2. For H3f4, we designed two shRNAs targeting respectively 5' UTR and 3' UTR. In all cases, we took advantage of the Genetic Perturbation Platform (GPP, https://portals.broadinstitute.org/gpp/public/) to evaluate sequence specificity and predicted efficiency of our shRNAs design. We also manually verified the specificity of each shRNA against the multiple sequence alignment of all the histone genes of the same family, retrieved through Mouse Genome Informatics (MGI, http://www.informatics.jax.org/). mESCs were transduced with shRNA-expressing lentiviral particles and selected for 1 week with blasticidin. Then we assessed the knockdown efficiency using RT-qPCR (FIG. 21), confirming specific downregulation of

both H1f6 and H2bu2; H3f4 was unaffected using both shRNAs and therefore we excluded it from downstream analysis.

Since H1f6 and H2bu2 are overexpressed in pluripotent cells compared to differentiated cells (FIG. 20), we investigated if their downregulation in mESCs had an effect on pluripotency. We therefore performed an alkaline phosphatase (AP) staining, a common biomarker of stem cells, on mESC knockdown for H1f6 (H1f6 KD) and H2bu2 (H2bu2 KD) (FIG. 22). Colonies in all conditions gave a nice AP signal, confirming their undifferentiated state, but there were differences in the total number of colonies and their area. In particular, H1f6 KD colonies were higher in number and had a bigger area, indicating possible alteration in self-renewal of mESCs upon depletion of H1f6. For example, this could point towards an increased proliferation of these cells.

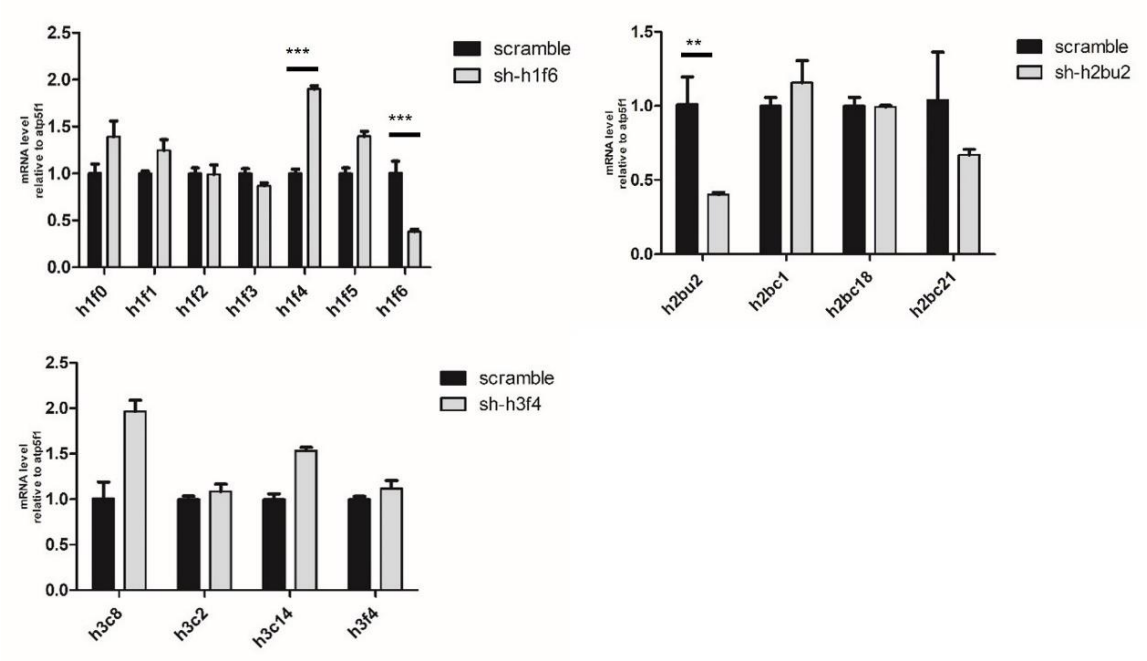

Figure 21 RT-qPCR analysis of the efficiency of StemHistones downregulation. mESCs were infected with control (sh-scramble) or sh-H1f6, sh-H2bu2, sh-H3f4 expressing lentiviruses and assessed for relative H1f6, H2bu, H3f4 mRNA expression. The expression of other histone genes belonging to each histone family was also assessed, to exclude any off-target effects. Histograms represent mean ΔCt (Ct gene of interest – Ct housekeeping ATP5F1) and standard deviation of 3 technical replicates; (***) p-value<0.001, (**) p-value<0.01, unpaired t test.

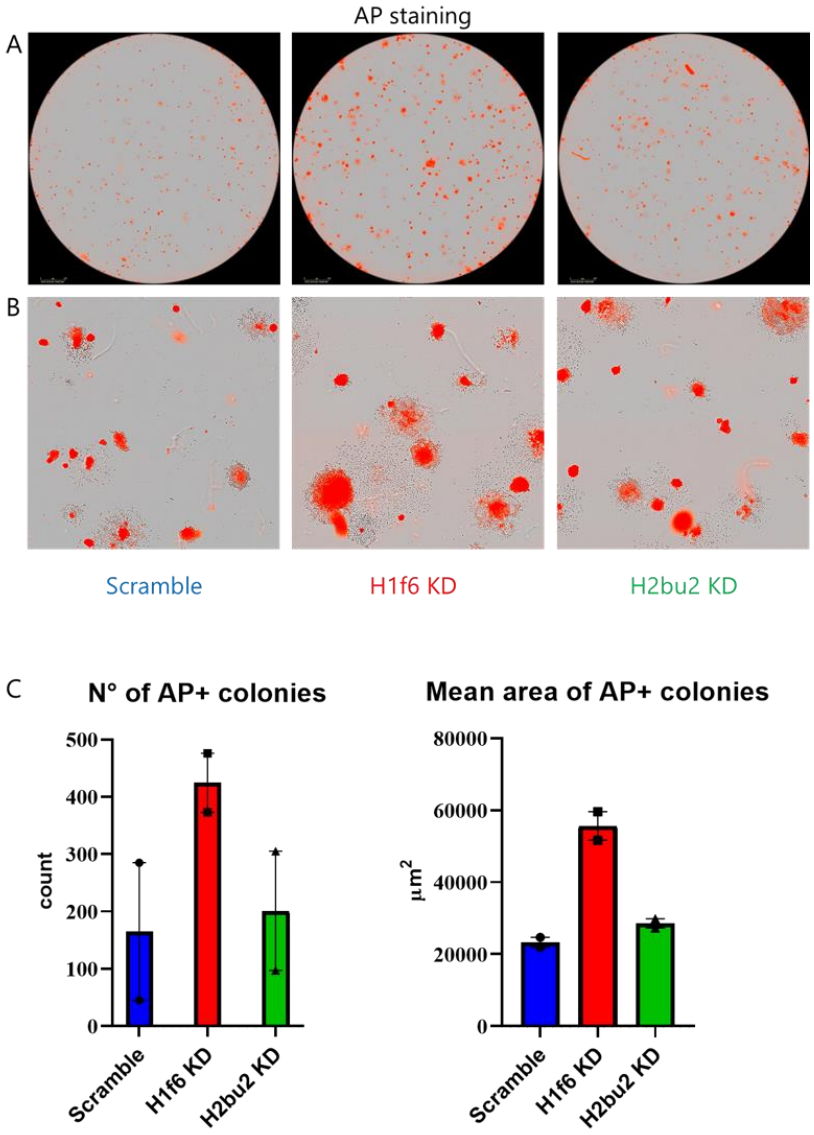

Figure 22 Effects of downregulation of H1f6 and H2bu2 on AP staining. Fluorescent images of alkaline phosphatase staining on mESCs scramble, H1f6 KD and H2bu2 KD after 6 days of culture at low density. (A) Whole well picture and (B) inset. (C) Quantification of number and mean area of AP+ colonies. Error bar represents standard deviation from 2 replicated experiments.

RT-qPCR analysis of core pluripotency gene Nanog (FIG. 23) showed no significant difference in its expression upon StemHistones knockdown. Taken these two results together we concluded that, despite Nanog transcription is not affected, AP staining indicates alteration in mESCs self-renewal upon H1f6 KD. More extensive analysis on pluripotency genes and the epigenetic landscape of their regulatory regions are needed to precisely assess how H1f6 KD affects pluripotency maintenance..

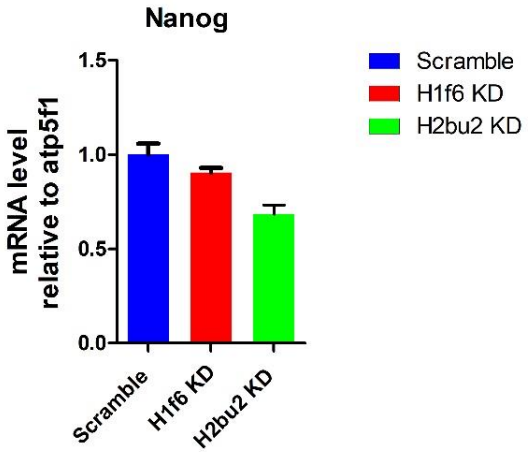

Figure 23 Nanog expression is not affected by StemHistones KD. RT-qPCR analysis of mRNA levels of Nanog in mESCs scramble, H1f6 KD and H2bu2 KD. Histograms represent mean ΔCt (Ct gene of interest – Ct housekeeping ATP5F1) and standard deviation of 3 technical replicates.

2.3.2 H1f6 and H2bu2 KD affect embryoid bodies formation and gene expression

To investigate the effect of H1f6 and H2bu2 KD on pluripotency, we investigated if depleting these StemHistones during differentiation could lead to defects and explain the need for their higher expression levels in pluripotent cells compared to differentiated cells.

We therefore differentiated mESCs into three-dimensional structures called embryoid bodies (EBs), because of their ability to mimic post-implantation embryonic tissues. EBs are mixed populations of differentiating cells that, as differentiation proceeds, gradually lose their pluripotent characteristics and start expressing differentiation markers of all three germ layers.

After 3 days of differentiation we collected and imaged EBs (FIG. 24) and we observed clear morphological differences. In particular H1f6 KD EBs formed less regular spheres compared to scramble, while H2bu2 KD EBs were smaller (FIG. 16 B).

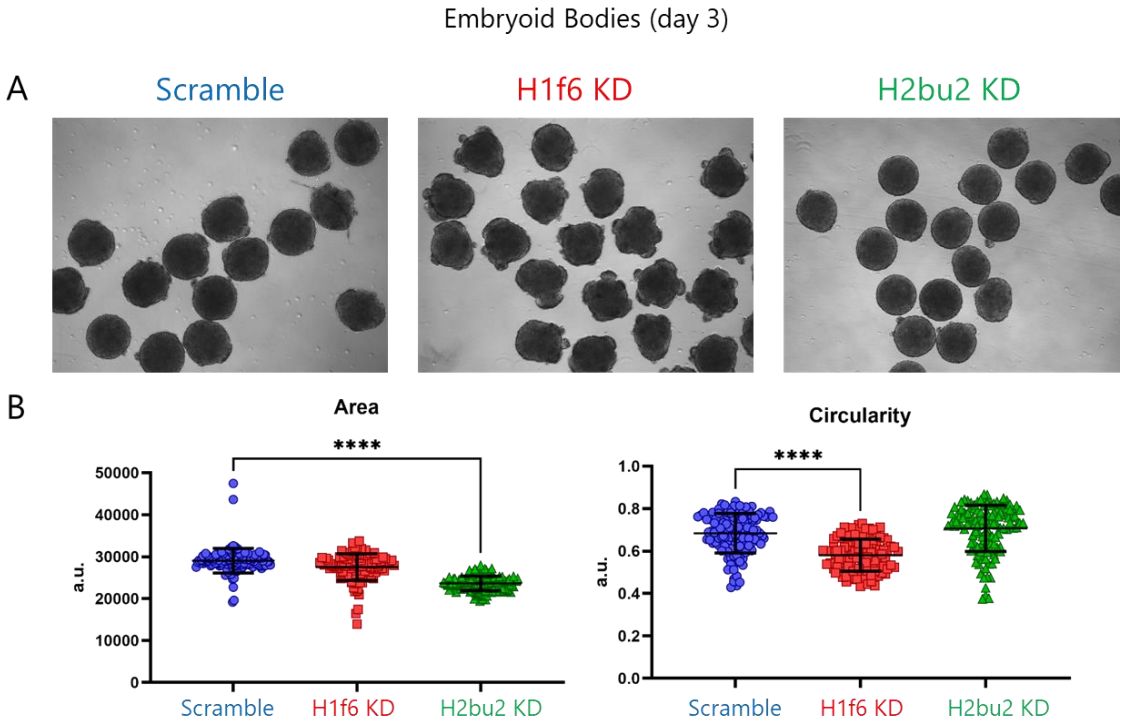

Figure 24 EBs morphology of scramble, H1f6 KD and H2bu2 KD. (A) Phase-contrast representative images of EBs. (B) Scatter plot of Area ad Circularity measures of EBs, calculated using FIJI software. One-Way Analysis of Variance (ANOVA) was used to determine statistical significance. (****) p-value<0.001.

We performed total RNAseq on both mESCs and EBs in all three conditions (scramble, H1f6 KD, H2bu2 KD), to uncover possible differences in gene expression induced by H1f6 and H2bu2 KD, both in mESCs and upon differentiation into EBs.

We first compared scramble mESCs and scramble EBs, to confirm that our EBs are indeed undergoing differentiation. Gene Ontology (GO) enrichment analysis identified the top 20 enriched terms in both upregulated and downregulated genes in EBs (FIG. 25). This analysis confirmed an ongoing differentiation process, as most enriched terms for upregulated genes relate to neuroectodermal formation (“axon development”, “axonogonesis”) or organogenesis (“epithelial tube morphogenesis”) or general differentiation-associated processes (“mesenchyme development”, “regionalization”),

while downregulated genes are clearly enriched for “response to leukemia inhibitory factor” used to keep mESCs pluripotent.

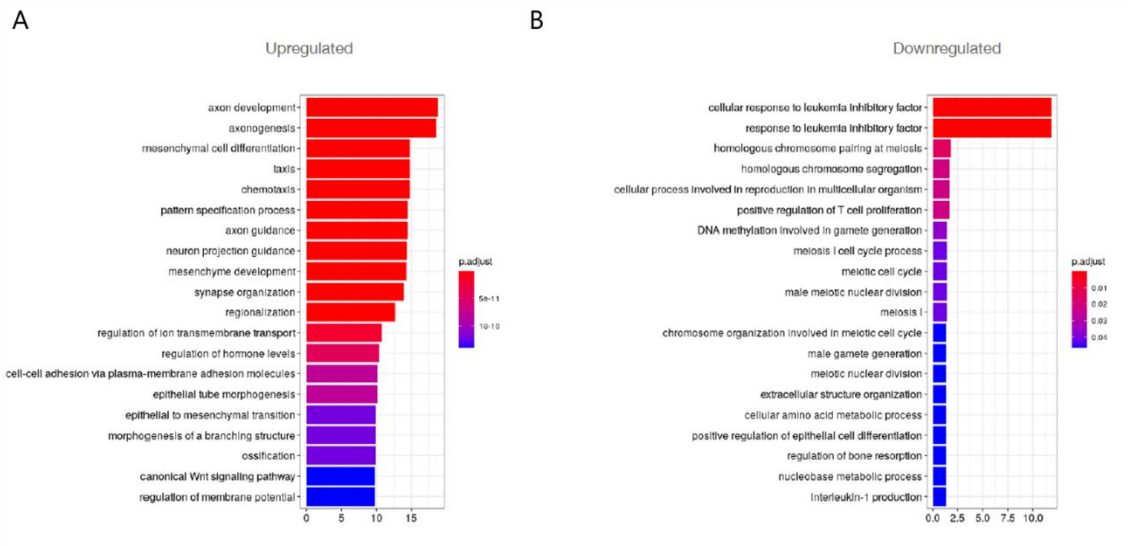

Figure 25 Control EBs transcription reflects exit from pluripotency and correct differentiation onset. Top 20 Gene Ontology terms enriched in either upregulated (A) or downregulated (B) genes in EBs scramble compared to mESCs scramble. GO terms are ranked according to the adjusted p value (FDR≤0.05), X axis is -log10 of adjusted p value.

We then focused on the specific effects caused by StemHistone depletion.

We investigated gene expression profiles of H1f6 KD mESCs and EBs by performing an unsupervised hierarchical cluster analysis of differentially expressed genes. The resulting dendrogram tree was cut to obtain 4 different clusters, corresponding to genes upregulated in each condition (scramble mESCs and EBs, H1f6 KD mESCs and EBs) (FIG. 26A). Only genes contained in cluster 1 were enriched for GO terms related to development and differentiation: these genes display higher expression in H1f6 KD mESCs and even higher expression in H1f6 KD EBs compared to control mESCs and EBs. GO analysis (FIG. 26B) revealed that these genes are mainly related to endodermal development (“endoderm development”, “endoderm formation”, “digestive tract morphogenesis”) among general differentiation terms. H1f6 KD seems therefore to increase expression of endoderm-related genes during differentiation; most notably, this effect is already visible in undifferendiated mESCs. We also carefully looked at genes in cluster 3, as these seem to be downregulated upon H1f6 KD in mESCs and further decreased in EBs. We

speculated they could be genes related to pluripotency, but to our surprise none of them was related to pluripotency and GO analysis returned no enriched term.

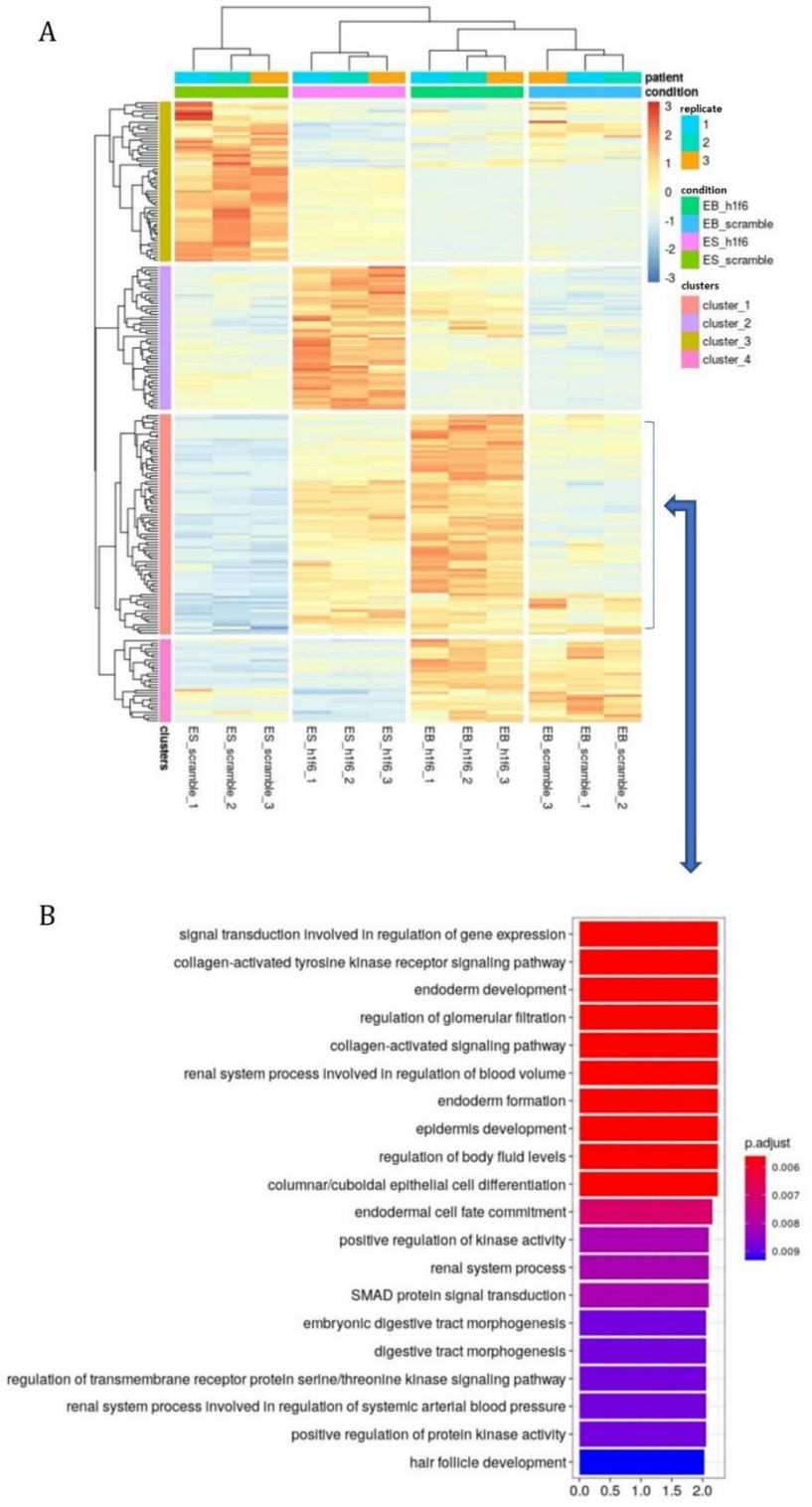

Figure 26 Endodermal genes are upregulated upon H1f6 KD. (A) Heat map showing gene expression of scramble control and H1f6 KD mESCs and EBs according to the 4 clusters found using unsupervised hierarchical clustering. (B) Cluster 1 magnification showing Top 20 Gene Ontology enriched terms GO terms are ranked according to the adjusted p value (FDR≤0.05), X axis is -log10 of adjusted p value.

On the other hand, H2bu2 KD did not show any effect in mESCs, as there was no GO term enriched. Instead, GO enrichment analysis on H2bu2 KD EBs showed that upregulated genes where enriched for mesoderm-related terms, especially regarding cardiac development (“cardiac septum development”, “heart morphogenesis”) and angiogenesis (“regulation of angiogenesis”, “regulation of vasculature development”), while downregulated genes were mostly related to ectoderm and neural development (“synapse organization”, “synapse assembly”) (FIG. 27). Taken together, these results suggest a role for H2bu2 in both mesodermal and ectodermal development.

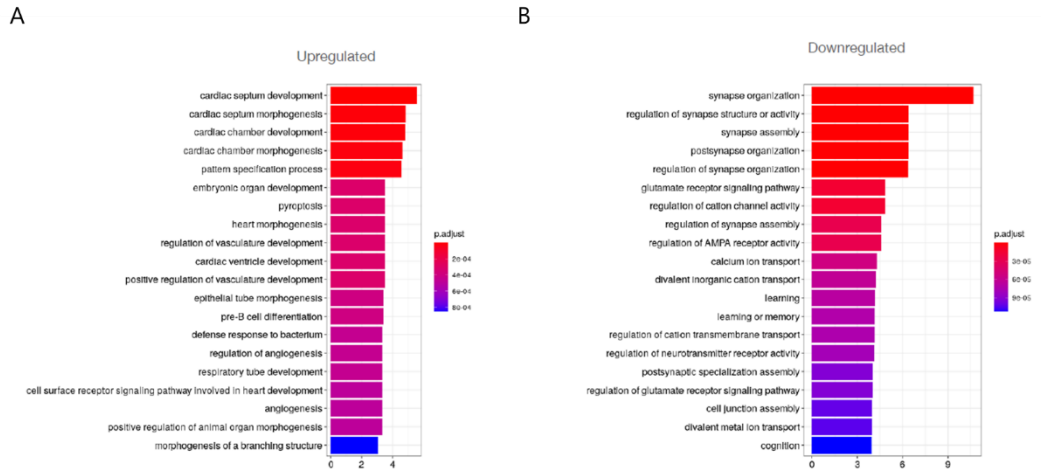

Figure 27 Transcriptional profilig of H2bu2 KD EBs. Top 20 Gene Ontology terms enriched in either upregulated (A) or downregulated (B) genes in H2bu2 KD EBs compared to scramble EBs. GO terms are ranked according to the adjusted p value (FDR≤0.05), X axis is -log10 of adjusted p value.

For confirmation, we analysed our transcriptomes at single gene level. Using a published list of germ layers markers (Yelagandula *et al*, 2021), we performed a cluster analysis on differentially expressed genes in our RNAseq experiment (FIG. 28). We overall confirmed the GO enrichment analysis, in particular the striking upregulation of endodermal genes in H1f6 KD mESCs and EBs compared to their scramble controls, while H2bu2 KD mESCs do not show any differences compared to their scramble control. On the other hand, H2bu2 KD EBs upregulated a number of mesodermal markers and decrease the expression of some ectodermal ones. This is in line with the GO analysis (FIG. 27), even though from the single-gene perspective the picture is less clear-cut.

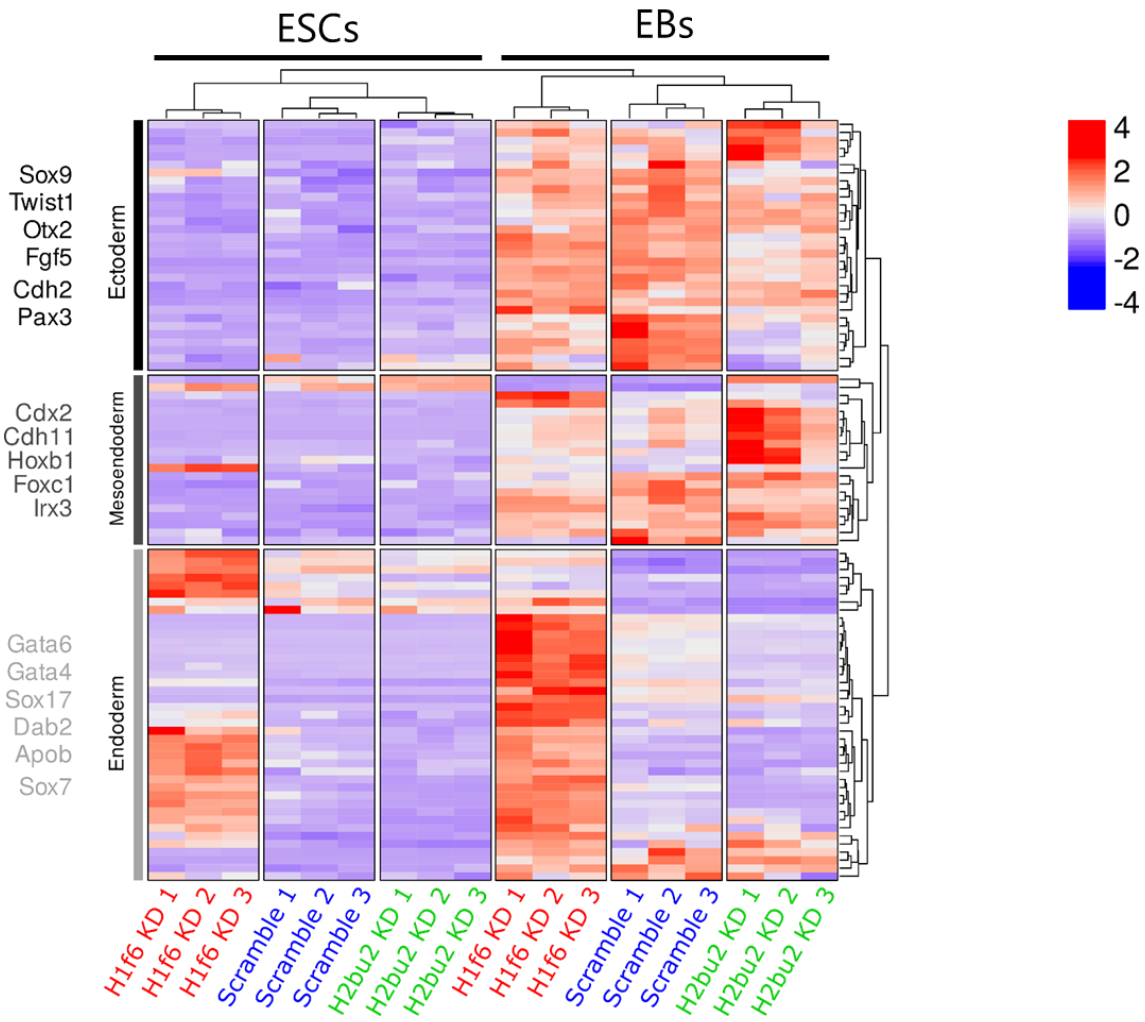

Figure 28 Differentiation into the three germ layers is affected upon H2bu2 and H1f6 KD. Heat map showing differential gene expression of selected genes of the 3 germ layers representative of ectoderm (top), mesoderm (middle), endoderm (bottom), with an absolute fold change value >1, FDR < 0.05. The most representative genes for each germ layer are indicated.

2.3.3 Generation of StemHistones KO cell lines

In the previous section we described how decreasing StemHistone expression via shRNA-mediated knockdown causes alteration in EBs formation. While we were able to obtain a 60-70% depletion for both H1f6 and H2bu2, both our shRNAs constructs for H3f4 did not induce any knockdown. To generate cell lines that completely lack StemHistones expression and extend the analysis to H3f4, we decided to knock-out each StemHistone via CRISPR/Cas9 technology.

We designed single-guide RNAs (sgRNAs) on H1f6, H2bu2 and H3f4 to target both 3' (sgRNA2, sgRNA3) and 5' regions (sgRNA1) of their sequence . The guides were highly scoring in VBC-score (https://www.vbc-score.org/) and good-scoring in CRISPick (https://portals.broadinstitute.org/gppx/crispick/public), two public web tools for scoring efficiency and off-target effects of Cas9 sgRNAs (Michlits *et al*, 2020; Doench *et al*, 2016). We cloned the sgRNAs in the lentiCRISPRv2 plasmid available through Addgene and transfected mESCs using lipofection. We transfected each population with two different plasmids together, targeting both 5' and the 3' end of each StemHistone CDS. We used two different combination of guides per histone (sgRNA1 + sgRNA2; sgRNA1 + sgRNA3) (FIG. 29A). After transfection, cells were selected with puromycin for 3 days. To check which of the two combinations of sgRNAs worked, we performed a PCR analysis in the pool of transfected mESCs as preliminary screening. We used forward (F)

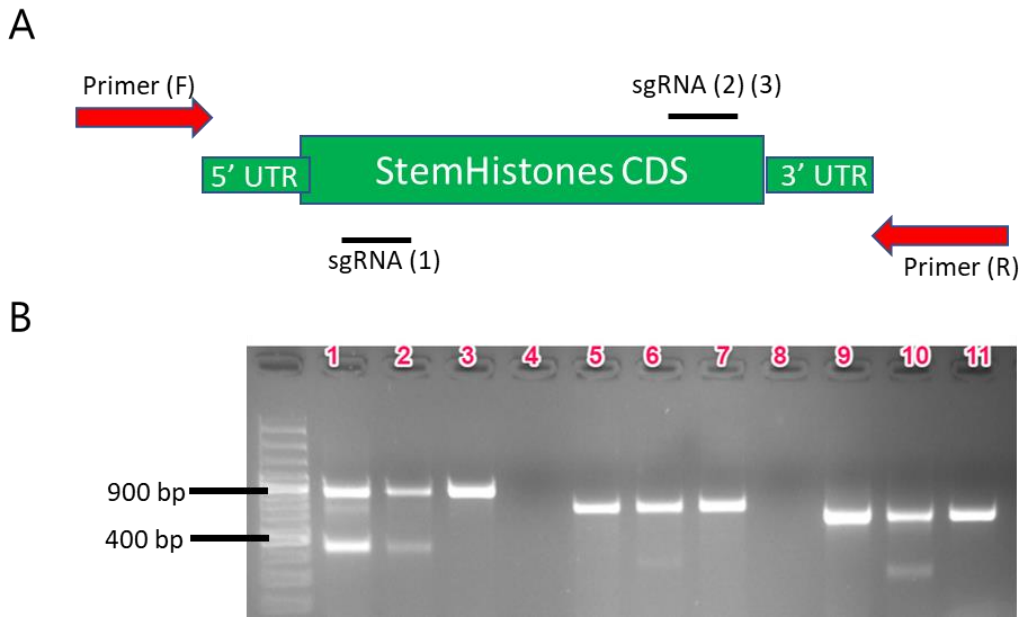

Figure 29 StemHistones KO CRISPR/Cas9 design and targeting. (A) Design strategy for sgRNAs targeting StemHistones genes CDS: two guides at the same time were used to generate a deletion containing the majority of the CDS from 5' to 3'. (B) Agarose gel for PCR reactions run on pools of mESCs co-transfected with lentiCRISPRv2 encoding for 2 sgRNAs in combination, either sgRNA1+sgRNA2 or sgRNA1+sgRNA3. Lanes 1 and 2 containing PCR products from genomic DNA of mESCs transfected with sgRNAs targeting H1f6, 5 and 6 targeting H2bu2 and 9 and 10 targeting H3f4 (3, 7, 11 are WT samples used as negative controls, 4 and 8 are "no sample" controls). The expected size for H1f6 WT is 955 bp while the expected size of H1f6 KO is 463 bp. Both combinations of guides for H1f6 worked, as faster bands are present on the gel in both lanes 1 and 2. For H2bu2, the expected size for WT is 758 bp, the expected size for H2bu2 KO is 362 bp: just one combination of guides for H2bu2 worked (lane 6). The expected size for H3f4 WT is 798 bp, for H3f4 KO is 397 bp. Also for H3f4 just one combination of guide worked (lane 10).

and reverse (R) primers outside the pair of double strand breaks generated by sgRNAs. (FIG. 29B) and confirmed that at least 1 combination per StemHistone was able to induce deletion of the region spanned by our sgRNAs.

Transfected populations of mESCs were then single-cell cloned to obtain mESCs bearing the same CRISPR/Cas9 editing. We confirmed StemHistone KO of both alleles in 1 or 2 clones per histone by PCR and Sanger sequencing, as well as RT-qPCR to confirm complete depletion at the RNA level (FIG. 30).

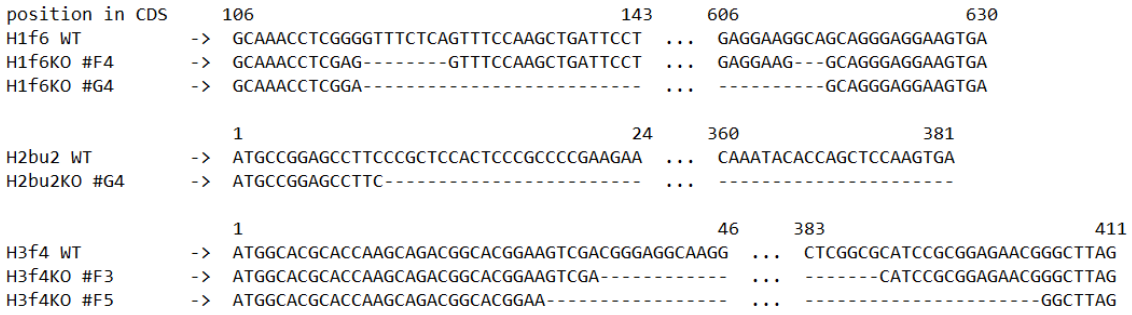

Figure 30 Sanger sequencing of selected clones. PCR-amplified genomic DNA from each clone was subjected to Sanger sequencing and then aligned with the reference sequence from RefSeq. Shown are regions were KO clones had gaps causing either frame-shift mutations with early stop codon (H1f6 and H3f4) or nearly-complete deletion of the CDS (H2bu2).

2.3.4 StemHistones KO confirm the role of unusual histone isoforms in differentiation onset

All StemHistones KO cell lines were viable and we therefore proceeded to their characterization, as we did for StemHistones KD.

First, we verified their undifferentiated status by performing AP staining and measuring Nanog expression via RT-qPCR (FIG. 31). As for StemHistones KD, there was no clear difference in Nanog transcription. For the AP staining, H1f6 KO clones were both close resembling WT control, differently from H1f6 KD. H2bu2 KO had a higher number of colonies, which were slightly smaller in size. H3f4 KO clones were even more difficult to interpretate as clone #H5 had both more and bigger AP positive colonies compared to clone #F3. Despite these observations, overall the staining was quite homogenous and the

differences were rather small. As previously noted, a more thorough analysis is needed to fully assess the self-renewal of StemHistones KO mESCs but the two marker we took into analysis did not show prominent alterations. The only clear difference is the increased number of colonies in H2bu2 and H3f4 KO.

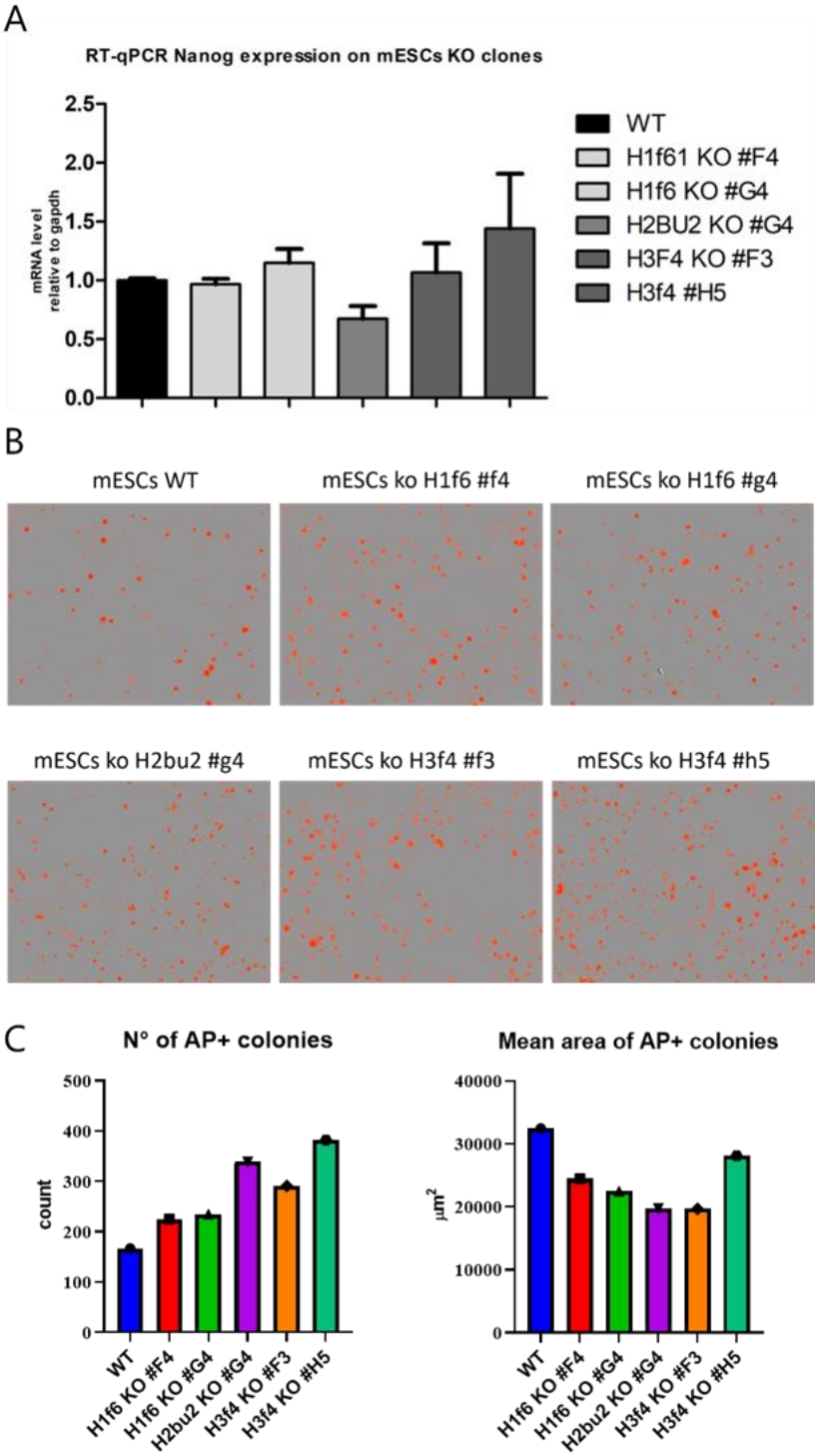

Figure 31 Effects of knockout of H1f6, H2bu2 and H3f4 on Nanog transcription and AP staining. (A) RT-qPCR analysis of mRNA levels of Nanog in mESCs WT, H1f6 KO (clones F4 and G4), H2bu2 KO clone G4 and H3f4 KO clones F3 and H5. Histograms represent mean ΔCt (Ct gene of interest – Ct housekeeping GAPDH) and standard deviation of 3 technical replicates. (B) Fluorescent microscopy images of Alkaline Phosphatase staining on mESCs WT, H1f6 KO (clones F4 and G4), H2bu2 KO clone G4 and H3f4 KO (clones F3 and H5) after 6 days of culture in presence of LIF at low density. (C) Quantification of number and mean area of AP+ colonies from (B).

Then we differentiated StemHistone KO mESCs into EBs and assessed their morphology (FIG. 32). We did not detect differences in the circularity of EBs, while the size was very variable. WT mESCs were very uniform in their size, while all StemHistones KO had somehow altered shape: both H1f6 KO clones formed more heterogenous and bigger EBs, while no difference was observed for the single H2bu2 KO clone. H3f4 KO clones had a

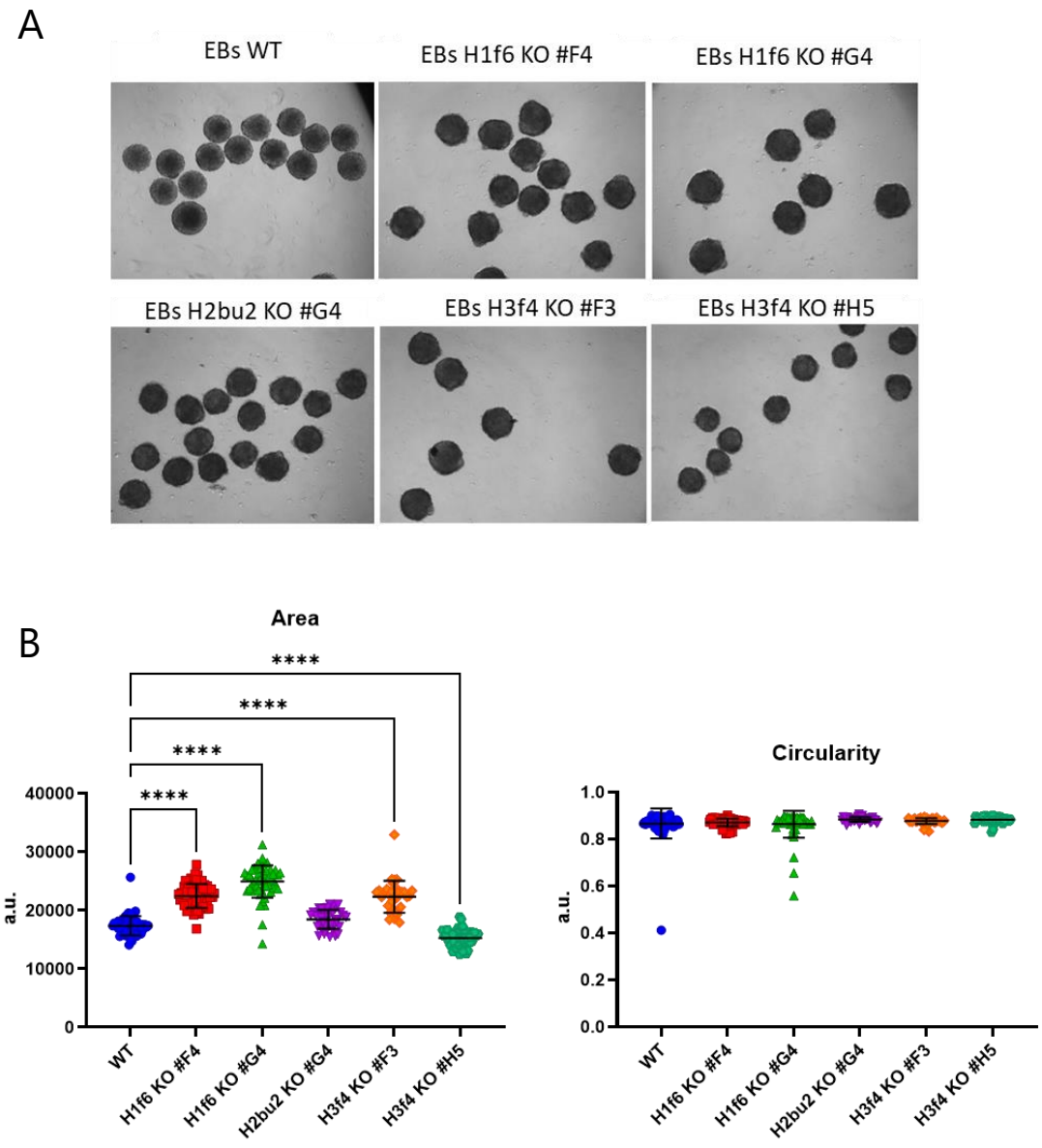

Figure 32 Differences in EBs morphology within WT, H1f6 KO (clones #F4 #G4), H2bu2 KO (clone #G4), H3f4 KO (clones #F3 #H5). (A) Pictures of EBs taken at inverted microscopy, 5x objective. (B) (B) Scatter plot of Area ad Circularity measures of EBs, calculated using FIJI software. One-Way Analysis of Variance (ANOVA) was used to determine statistical significance. (****) p-value<0.001.

variable phenotype, with clone #F3 forming bigger EBs and clone #H5 forming smaller ones.

We set out to investigate deeper on the impact of StemHistone depletion by analysing via RT-qPCR the transcriptional profile of a panel of differentiation markers in EBs. We chose markers for each germ layers that were upregulated in scramble control EBs in our RNAseq experiment: Sox17 and Gata6 for endoderm, Otx2 and Cd24a for ectoderm, Eomes for mesoderm (FIG. 33).

Transcriptional analysis in H1f6 KO clones showed a strong upregulation for both endodermal markers Sox17 and Gata6, already apparent in H1f6 KO mESCs and for clone #F4 also present in H1f6 KO EBs. Ectodermal markers were instead less coherent, Otx2 was upregulated in clone #F4 EBs compared to WT, but not in clone #G4 EBs, while Cd24a was downregulated for both clones compared to WT, but clone #F4 retained higher expression compared to clone #G4. Finally, mesoderm marker Eomes was downregulated in EBs from both KO clones. Overall, these data confirm what we observed for H1f6 KD in RNAseq, with a clear derepression of endodermal markers in H1f6 KO mESCs, also resulting in overexpression in EBs for clone #F4. Mesodermal and ectodermal markers are less clear-cut but overall they are similar or lower than WT.

H2bu2 KO instead showed a downregulation in EBs for all markers compared to WT, save for Otx2 which does not change. In KO mESCs, Gata6 is upregulated and, to a lesser extent, also Sox17. Importantly, mesodermal marker Eomes did not show any upregulation in KO EBs. This was not among the upregulated mesodermal markers in H2bu2 KD samples, therefore analysis of other markers could give a more coherent picture. Overall, we could not obtain a confirmation of the H2bu2 KD EBs phenotype, even though we detected a downregulation of Cd24a as ectodermal marker. A more comprehensive analysis is probably needed to clarify the picture.

For H3f4 KO we analysed two clones and we do not have a transcriptomic analysis for comparison. What we detected was similar to H2bu2 KO: endodermal markers are slightly upregulated in KO mESCs but downregulated in KO EBs; ectodermal markers are not coherent between clones, with Otx2 being upregulated in clone #F3 but unchanged in clone #H5, while Cd24a is downregulated for both; Eomes is downregulated for both clones EBs. Therefore, again we do not gain a clear picture from these analyses, but we

can speculate that lacking H3f4 interferes with differentiation as well, since all markers for all germ layers are affected to some extent in H3f4 KO EBs.

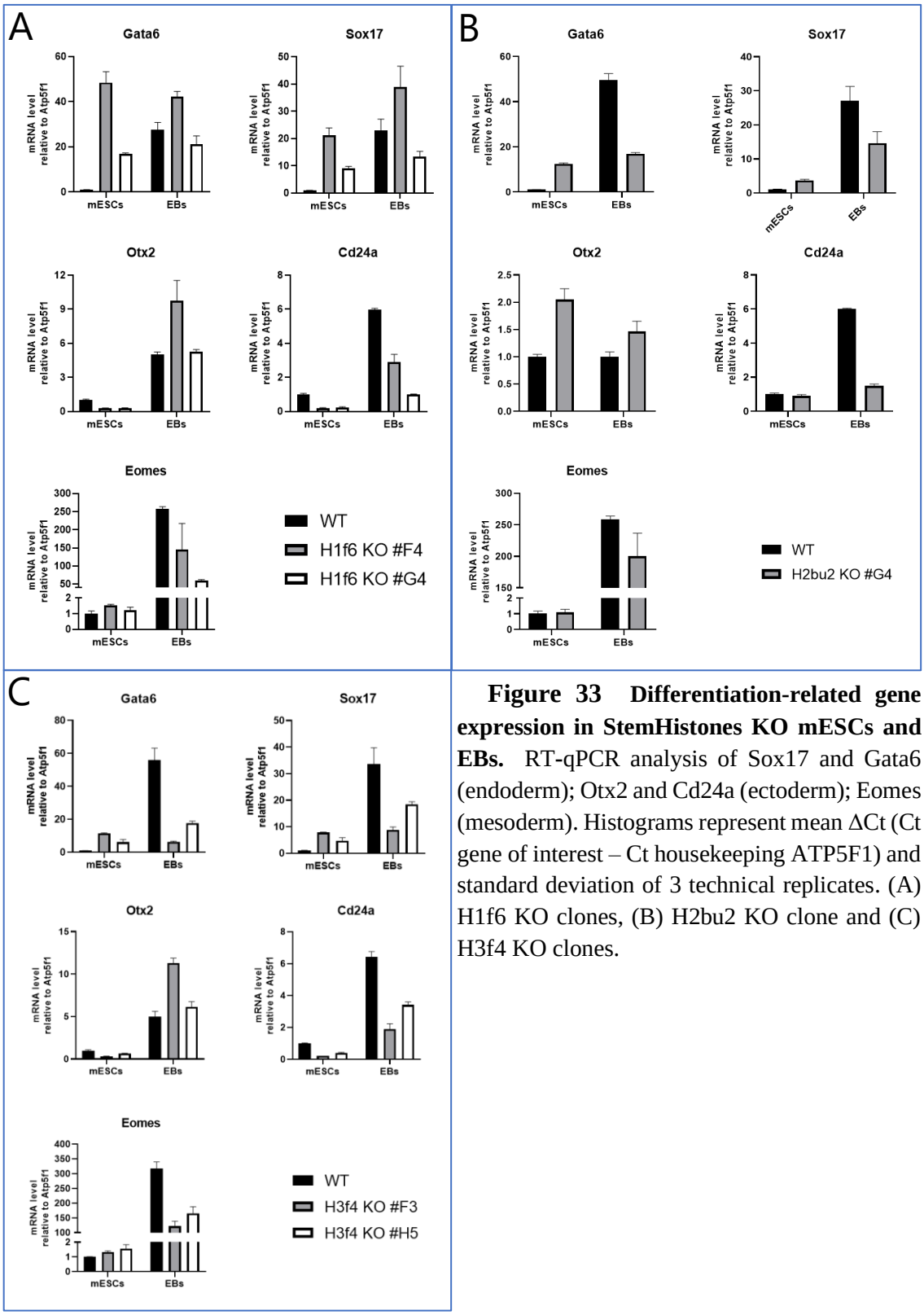

Figure 33 Differentiation-related gene expression in StemHistones KO mESCs and EBs. RT-qPCR analysis of Sox17 and Gata6 (endoderm); Otx2 and Cd24a (ectoderm); Eomes (mesoderm). Histograms represent mean ΔCt (Ct gene of interest – Ct housekeeping ATP5F1) and standard deviation of 3 technical replicates. (A) H1f6 KO clones, (B) H2bu2 KO clone and (C) H3f4 KO clones.

In conclusion, we confirmed that all 3 StemHistones appear to be involved in the correct onset of differentiation, as differentiating mESCs lacking their expression have altered transcription of germ layers markers. Specifically, H1f6 seems to be directly regulating endodermal genes expression as they are upregulated in H1f6 KO even in the pluripotent state, where these genes are normally repressed. H2bu2 and H3f4 KO instead do not have strong effects in mESCs, save for a slight upregulation of Sox17 and Gata6. In KO EBs, they cause a general downregulation of germ layers markers, suggesting that their absence is somehow impeding the normal differentiation onset.

2.4 StemHistones are in chromatin

The lack of commercially available antibodies does not allow us to image and detect StemHistones at protein level. We therefore decided to rescue their expression in StemHistones KO cell lines using lentiviral-driven expression of their CDS fused with a C-terminal 3XFLAG tag. This strategy allows us to have cell lines where the expression of StemHistones is only coming from our lentiviral construct and therefore provides a way both to analyse the effect of StemHistones overexpression and interrogate their behaviour at protein level.

2.4.1 Establishing StemHistones-FLAG cell lines

We generated constructs in a lentiviral backbone for each StemHistone CDS with a C-terminal 3xFLAG, under a strong constitutive promoter (EF-1α). KO mESCs were transduced with lentiviral particles encoding for the StemHistone they were lacking and selected with blasticidin for 7 days. We first verified the expression of StemHistones performing a western blot on whole-cell lysates of the pool of transduced and selected cells and we detected FLAG signal as single sharp bands at the expected molecular weights (FIG. 34).

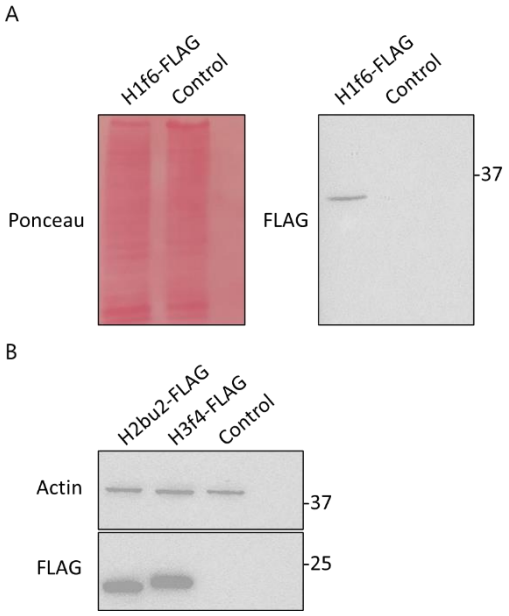

Figure 34 mESCs transduced with StemHistones-FLAG construct express the protein. (A) H1f6-FLAG protein is expressed at expected molecular weight, Ponceau used as loading control. (B) H2bu2-FLAG and H3f4-FLAG are expressed at expected molecular weight, Actin used as loading control.

We then decided to single-cell clone our pools to obtain cell lines with homogenous expression of the transgene. We isolated clones that we imaged using immunofluorescence staining, to verify that our StemHistones are localized in the nucleus. Indeed, all the clones showed co-localization with DAPI signal, confirming that they are expressed in the nucleus (FIG. 35).

These results confirmed that we StemHistones overexpressing cell lines are viable and histones localize in the nucleus as expected. They will be useful tools for further characterization of StemHistones in mESCs.

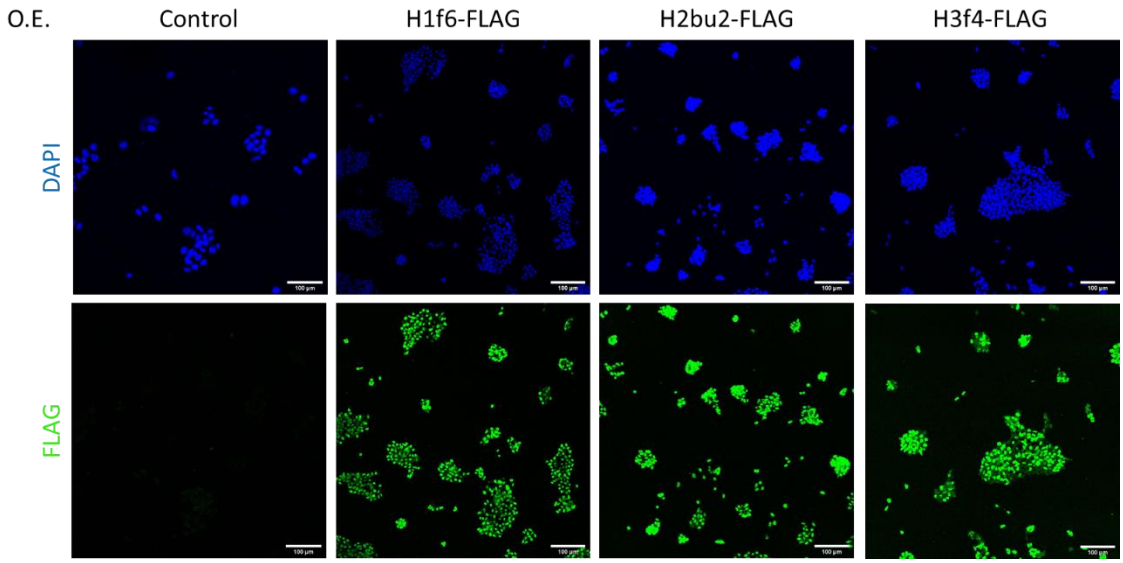

Figure 35 StemHistones show nuclear localization in mESCs. Immunofluorescent staining of mESCS StemHistones-FLAG, expressing the indicated construct. DAPI staining of nuclei in blue, FLAG staining in green. Scale size is 100 mM

DISCUSSION

3.1 Hmgb1 alters chromatin in iPS cells but does not influence reprogramming

Hmgb1 is a chromatin protein with a number of different functions, both inside the cell and in the extracellular milieu. Importantly, its function in chromatin regulation is not well understood: it is known that Hmgb1 is one of the most abundant proteins in the nucleus and that it binds DNA, acting as a DNA-chaperone to favour nucleosome assembly, but it is very difficult to map its binding sides on DNA (Goodwin *et al*, 1973; Celona *et al*, 2011b). ChIPseq approaches have long failed to capture Hmgb1 on chromatin, because of its insensitivity to formaldehyde fixation methods (Pallier *et al*, 2003; Teves *et al*, 2016) and only recently a double-crosslinking approach managed to map Hmgb1 on chromatin in the context of senescence in human cells (Sofiadis *et al*, 2021).

We therefore decided to follow a different route, starting off from the observations made in our lab that MEFs lacking Hmgb1 contain less histones (Celona *et al*, 2011a) than WT MEFs and that mESCs contain less histones than differentiated cells (Karnavas *et al*, 2014). Our idea was to try and assess whether total histone content could impact on the ability of MEFs to reprogram to iPS cells, potentially increasing the efficiency of the reprogramming process by mimicking the open, "histone-low", chromatin of iPS cells, using the lack of Hmgb1 as a tool to induce "histone-low" MEFs.

We established a reprogramming system in our lab, using a commercially available inducible OKSM mouse model. We were able to reprogram MEFs to iPS cells by culturing them in doxycycline for approximately two weeks and assess the activation of the endogenous pluripotency network by activation of Oct4GFP transgene (FIG. 9).

Then, we needed to confirm that reprogramming MEFs to iPS cells also induces a decrease in histone content, as previously observed in our lab for mESCs. Indeed we confirmed this reduction by western blot (FIG. 10). We asked whether this reduction was due to down-modulation of transcription or post-transcriptional processes. Both qPCR

analysis and RNAseq on MEFs and iPSCs clearly showed that histone transcription is not downregulated in iPSCs compared to MEFs (FIG. 11). On the contrary most genes are slightly upregulated, with some specific low-expressed genes that have a strong increase (FIG. 11B). On the same line, genes that are involved in post-transcriptional regulation of replication-dependent histone mRNAs do not decrease their expression in iPS cells and some are instead slightly upregulated (SLBP, Cpsf3) (FIG. 11C). Therefore, there has to be some mechanism at the mRNA or at the protein level regulating histone decrease. Although we decided not to investigate this mechanism, we can imagine three possible ways for the cell to decrease histone content. First, the histone mRNAs might be present, but not translated; this is unlikely because the protein responsible for their maturation and translation, SLBP, is highly expressed. At the protein level, the cell could increase histone turn-over via proteasomal degradation and one could test this hypothesis by inhibiting the proteasome itself. Another interesting possibility is that histones get actively secreted by the cells, as we previously showed to happen in LPS-stimulated macrophages (Nair *et al*, 2018). Both hypotheses are nevertheless very challenging to test in a process like reprogramming, which happens over a long period of time (approximately 2 weeks) with constant medium replacement. The length of the process makes pretty much impossible to test proteasomal inhibition, which after long time periods precludes cell viability; the medium replacement needed for constant induction of the OKSM transgene and maintenance of emerging iPSCs colonies does not allow to test the release of histones in the medium. We consequently decided to focus our attention on the histone decrease itself and not on the mechanism that causes it.

We therefore successfully reprogrammed MEFs lacking Hmgb1 (both Hmgb1 HET and Hmgb1 KO) to iPS cells and confirmed they are pluripotent by analysis of markers such as Nanog and Oct4 at mRNA and protein level (FIG. 12 – 13). We demonstrated that pluripotent cells could sustain the absence of Hmgb1 by establishing iPS cell lines from Hmgb1 KO MEFs. In order to check whether Hmgb1 KO iPS cells contained a decreased amount of histones, we decided to take advantage of the SUPER-Silac technique. This proteomic approach can help in precisely quantifying the histone amount in our iPSCs and also provide an overview on histone post-translational modifications. Indeed, SUPER-Silac confirmed the reduction in core histones also in Hmgb1 KO iPS cells (FIG 14). H3 and H4 levels were significantly reduced by about 20% compared to

Hmgb1 WT iPS cells and we observed a non-significant trend also for H2A and H2B. When we looked at all histones taken together, we could confirm a general decrease of about 20% for histone content in Hmgb1 KO iPSCs. This is in line with experiments that were previously done in MEFs and cancer cells in our lab (Celona *et al*, 2011a; Brambilla *et al*, 2020).

To our surprise, we also found a few changes in the abundance of histone PTMs, as measured using SUPER-Silac (FIG 15). In particular, we observed a decrease in markers of active chromatin such as H3K4me3 and H4 acetylation (FIG. 15A-B). This is interesting because we previously observed (Celona *et al*, 2011a) increased transcription in MEFs lacking HMGB1 with decreased histones. One could imagine that decreasing active chromatin marks could allow iPS cells to avoid unwanted transcriptional activation. On the other hand, also H3K9me3 is strikingly reduced (FIG. 15C). This resembles the reduction in H3K9me3 associated with depletion of the histone chaperone CAF1 (Cheloufi *et al*, 2015; Wang *et al*, 2018). CAF1 acts as a chaperone for nucleosome formation on the histone side and Hmgb1 plays the same role on the DNA side. At the same time, we know that HMGB1 depletion decreases histone content and it is reasonable to imagine CAF1 depletion to have the same effect. We therefore speculated that, conversely, HMGB1 could mirror CAF1 function in being a stabilizer of cell identity (Cheloufi & Hochedlinger, 2017) and lead to increased reprogramming. Instead, we found that reprogramming efficiency of MEFs lacking Hmgb1 was not different from WT MEFs (FIG. 16). Of course, lack of Hmgb1 could have an impact on other aspects of cell fitness and proliferation and therefore mask the effect of histone depletion. It would be interested to directly test the hypothesis of histone depletion as a mechanism for increasing reprogramming efficiency, for example by knocking-out a highly expressed histone gene and observe how cells adapt and eventually behave during reprogramming.

In conclusion, we were able to derive Hmgb1 KO iPSCs and confirm their pluripotency according to commonly used molecular markers. Lack of Hmgb1 does not impact on the efficiency of reprogramming, but alters chromatin in iPS cells. In fact, we observed that they contain less core histones using the precise proteomic approach of SUPER-Silac, confirming a common phenotype of cell lines lacking Hmgb1. Since iPS cells contain less histones than MEFs, we plan to expand the analysis and compare MEFs and iPSCs Hmgb1 WT and KO to have a comprehensive landscape, that could confirm

how cells can survive with a strongly reduced amount of histones. It is interesting to observe that Hmgb1 KO iPSCs contain approximately half the histones of WT MEFs, based on our indirect comparison. This points towards the idea that cells can survive with a strongly reduced amount of histones, even more than what we hypothesised previously. Therefore, the amount of histones a cell needs to properly package its chromatin and survive it is likely low and more importantly cell type dependent.
We also observed for the first time that Hmgb1 KO iPSCs modify their chromatin by altering the level of various histone PTMs, in particular decreasing H3K9me3. Analysing the genomic distribution of H3K9me3 in the absence of Hmgb1 could give some functional insight about this reduction, especially given how important heterochromatin maintenance is in the context of pluripotency (Chen *et al*, 2012).

3.2 StemHistones are unusual histone variants involved in differentiation onset

When we performed total RNAseq in MEFs and iPSCs to look at histone gene transcription, we were struck by the overexpression of 5 lowly expressed genes (Fig 11: H2ac1, H2bc1, H1f6, H2bu2, H3f4). We named these genes “StemHistones” because of their specific upregulation in iPS cells and we confirmed their overexpression in pluripotent cells using publicly available RNAseq datasets (Amlani *et al*, 2018; Knaupp *et al*, 2017) (Fig. 18). We also confirmed their upregulation in ES-e14 mESCs line via qPCR (Fig 19).

Interestingly, all 5 genes encode for isoforms of canonical histone genes. H2ac1 and H2bc1 encode for tH2A and tH2B variants, which are oocyte- and testis-specific. Previous work (Shinagawa *et al*, 2014b) showed that overexpressing these 2 variants during reprogramming leads to increased efficiency of iPS cells generation via creation of an open chromatin environment. We speculated that also the other 3 genes (H1f6, H2bu2 and H3f4) could have a specific function in mESCs.

H1f6 and H3f4 encode for H1t and H3t histones, deemed to be testis-specific variants as well (Drabent *et al*, 1993b; Maehara *et al*, 2015), while H2bu2 encodes the H2b

isoform H2B.U that does not have any characterization in the literature and is only cited in a study about mouse sperm (Golas *et al*, 2003). Therefore, all have a link with the male reproductive system but no characterization in the context of embryonic stem cells. Interestingly, all these variants are conserved in humans: H1T is a testis-specific H1 isoform which also forms less compacted chromatin (MacHida *et al*, 2016); H3T is also a conserved testis-specific H3 that forms unstable nucleosomes, despite through different residues compared to mouse H3t (Tachiwana *et al*, 2010); H2B.U is also conserved and very little studied in humans, the same way it is not characterized in mice.

We started our characterization by depleting our chosen three StemHistones via shRNA mediated knock-down in ES-e14 mESCs. We designed shRNA so that they would be as specific as possible to our genes of interest, to avoid off target effects on other histone genes. We obtained a good decrease for both H1f6 and H2bu2, but both our shRNA designs were not effective for H3f4 depletion (FIG. 21). Therefore, we developed tools for CRISPR/Cas9-mediated deletion of StemHistones, which turned out to be efficient. We evaluated CRISPR/Cas9 specificity by assessing expression levels of histone genes very similar to StemHistones, using RT-qPCR. A more thorough analysis of off-targets needs to be carried out to rule out differences between clones due to gene editing events happening at other genomic sites.

The undifferentiated state of mESCs was measured in either StemHistones down-regulation or knock-out by both AP staining and transcriptional analysis of Nanog (FIG. 22-23-31). AP staining showed a difference in size and number of colonies, which were clearly increased upon H1f6 KD, while H1f6 KO were much more similar to the control. Both KD and KO for H2bu2 showed only mild differences, while H3f4 KO clones had an increased number of AP+ colonies but also notable differences among them. Overall, AP staining suggests that depleting StemHistones does affect mESCs self-renewal but the effects are confounding, given differences between KD and KO and even among KO clones for the same gene. We would need to assess changes in histone marks profile and global gene expression for all conditions in order to have a complete understanding. Moreover, assessing off-target effects could explain the differences in H3f4 KO clones.

We then hypothesised that StemHistones could affect the correct differentiation onset and indeed we detected some morphological differences upon mESCs differentiation in

EBs: H1f6 KD EBs were less regular in shape and H2bu2 KD EBs were smaller compared to Scramble control EBs (FIG. 24); StemHistones KO EBs all had altered sizes compared to WT controls (FIG. 32). We speculated that the morphological defect could result from an unbalance in lineage commitment, since EBs are usually formed by cells of all three germ layers (ectoderm, mesoderm, endoderm). Therefore, we first profiled StemHistones KD mESCs and EBs via RNAseq to try and get a general picture of transcriptional misregulation. Then, we backed up these results by analysing a number of germ layer markers in StemHistones KO mESCs and EBs via RT-qPCR. Indeed, we found that lack of StemHistones had an impact on differentiation onset, in distinct ways.

H1f6 depletion clearly increased the expression of endodermal genes, as highlighted by top 20 GO terms enriched in H1f6 KD mESCs and EBs compared to controls (FIG. 26). Interestingly, H1f6 KD mESCs already show an increase in endodermal gene expression, as highlighted in FIG. 20. These genes get further overexpressed when H1f6 KD mESCs are differentiated to H1f6 KD EBs. Altogether, it seems that depleting H1f6 has a direct effect on de-repressing endodermal markers, as they get activated already in mESCs, where they are normally repressed. Knocking-out H1f6 confirms the effect of endodermal genes upregulation that we noted in H1f6 KD, which is striking in mESCs while in EBs varies in the two KO clones we analysed (FIG. 33A). It is very interesting to note that so far knock-out of a single histone H1 gene was reported to have no impact on mES cells pluripotency and differentiation, nor on mice viability. The only effects were visible when three somatic H1 genes were deleted at the same time, affecting neural differentiation by incomplete repression of pluripotency genes and impaired activation of neuron-specific genes (Fan *et al*, 2001b; Zhang *et al*, 2012). Our results highlight that H1f6 could have a specific and distinct role in mESCs compared to somatic H1s.

On the other hand, H2bu2 knock-down did not affect transcription in mESCs. In fact, there were no GO terms enriched in either down- or up-regulated genes in H2bu2 KD mESCs and if we look at lineage-specific markers we can appreciate that there is no difference in expression between H2bu2 KD and Scramble control mESCs (FIG. 27). On the contrary, upon differentiation to EBs there is a clear difference: H2bu2 KD EBs upregulate genes related to mesodermal and cardiac differentiation, while top GO terms for downregulated genes relate to synaptic organization. Looking at the markers in FIG. 20 we confirmed that some mesodermal markers are upregulated in H2bu2 KD EBs

compared to Scramble control EBs. At the same time, some ectodermal markers appear to be downregulated, even though not in an uniform way, as some others are unchanged or even upregulated. Therefore, H2bu2 KD seems to affect differentiation onset as well, although its effect is milder than H1f6 KD and less clear-cut. We did not confirm in H2bu2 KO EBs the upregulation of mesodermal markers that we had seen in our RNAseq on H2bu2 KD EBs (FIG. 33B). We did see downregulation of ectodermal marker Cd24a while Otx2 was unchanged. Overall, H2bu2 KO effect on differentiation appears different from H2bu2 KD, but a more extended analysis is needed to comprehensively evaluate their effects.

For H3f4 KO we were struck by the different phenotype of EBs (FIG. 32). The clones we analysed were in fact different in size, clone F3 bigger and clone H5 smaller than WT control, and also had some difference in transcription (FIG. 33C). For example, Otx2 was clearly upregulated in clone F3 while it was unchanged in clone H5 compared to control. Based on this limited transcriptional analysis, we can only say that H3f4 KO seems to impact on differentiation as well, although its effects need deeper investigation to be clarified.

To extend the analysis, we will profile the transcriptome of StemHistone KO mESCs and EBs through RNAseq in order to further confirm our observations and validate data observed in StemHistones KD RNAseq.Since we only analysed the effect of StemHistones depletion at early time-point in differentiation, having looked only in EBs after three days, we will also perform *in vitro* differentiation experiments of StemHistone KO mESCs towards endoderm, mesoderm and ectoderm using published protocols. Analysis of cell morphology and differentiation markers through immuno-fluorescence and transcriptional profiling will tell us whether lacking StemHistones is indeed impairing proper differentiation of pluripotent stem cells to specific lineages.

We also wanted to provide evidence of the cellular localization of StemHistones and their incorporation into chromatin. We then established rescue cell lines, overexpressing FLAG-tagged H1f6, H2bu2 and H3f4 in KO mESCs. We verified their expression at protein level and their expected nuclear localization (FIG. 34 and 35). With these tools, we can now further investigate on StemHistones from two different viewpoints. On one side we will differentiate these mES cell lines in a rescue experiment, to verify whether

overexpression leads to interference with specific differentiation pathways. For example, it would be interesting to see if H1f6 overexpression leads to strong silencing of endodermal genes. On the other hand, the 3xFLAG tag allows us to analyse the genomic distribution of our variants using CUT&Tag and map their positions on the genome compared to canonical isoforms (Kaya-Okur *et al*, 2019). On this note, the use of a strong promoter to drive StemHistones-FLAG expression, such as the Ef1α we are employing, could lead to confounding results, as these StemHistones are lowly expressed compared to canonical histones. In particular, if StemHistones are normally localized in specific genomic sites, the overexpression could lead to saturation of the molecular machinery that controls their deposition, leading to artifacts in their genome-binding profile. Therefore, we will carefully screen a number of clones to obtain a range of expression closer to the endogenous one.

Our work underlines the importance of small sequence variation in histone genes and in particular canonical histone genes. Through a simple comparison of transcriptomic datasets, we could identify some differentially expressed histone genes and dissect their function. We also highlight that even though StemHistones contribute for a very small percentage to the total pool of histones, they can exert specific function and contribute to cell lineage specification and development. How they achieve their function remains to be elucidated, but there are growing examples in the literature of tissue-specific functions for histone variants, as for H3mm variants (Hirai *et al*, 2022; Harada *et al*, 2018). One can imagine that the low expression level of these histone isoforms could correlate with a specific function in precise genomic locations rather than the general chromatin-packaging function of highly-expressed isoforms. For example, we could speculate that H1t impacts on endodermal gene expression because it is recruited to regulatory regions exerting a repressive function in mESCs. Our future CUT&Tag experiments aim at understanding where StemHistones are specifically recruited to possibly regulate particular gene sets.

In conclusion, we started to characterize three unusual histones in the context of pluripotent stem cells (hence the name, StemHistones). We provided initial evidence of their role in the onset of differentiation *in vitro* and showed how they have different individual roles: H1f6 appears to specifically regulate endodermal genes, while H2bu2

and H3f4 both impair differentiation when depleted. H2bu2 might have some specific effect on mesoderm and ectoderm that require further confirmation. We established tagged versions of all three StemHistones, confirmed their nuclear localization and we will now use these tools to further investigate the effect on chromatin and differentiation.

We therefore propose these three unusual histone variants as new players in the regulation of germ layers determination upon pluripotency exit.

MATERIALS AND METHODS

4.1 Mice

All mouse lines were bred and maintained according to approved IACUC protocols. C57/BL6 mice, heterozygous for Hmgb1, were crossed to each other and with B6;CBA-Tg (Pou5f1 – EGFP)2Mn/J (OG2) mice, purchased from Jackson Laboratory (mouse strain 004654), for isolation of mouse embryonic fibroblasts (MEFs) for reprogramming upon lentiviral infection. Double transgenic Hmgb1het/Pou5f1-EGFP mice were crossed with Gt(ROSA_26Sor$^{tm1(rtMA*M2)Jae}$Col1a1$^{tm3(tet0-Pouf1-Sox2-Klf4-Myc)Jae}$ mice, purchased by Jackson Laboratory (mouse strain 011004), for isolation of mouse embryonic fibroblasts (MEFs) for reprogramming upon doxycycline addition.

4.2 Derivation, culture and inactivation of primary mouse embryonic fibroblasts (MEFs)

Pregnant C57/Bl6 female mice were sacrificed at 14.5 dpc. Uteri were first placed in a tube with 50 ml isolation medium containing DMEM Glutamax (Life Technologies, #61965-059), 10% FCS (Life Technologies, #15140-122), 100 U/ml penicillin, 100 μg/ml streptomycin (PEN/STREP, Life Technologies, #15140-122). The uteri were placed in a 100 mm petri dish containing 20 ml sterile PBS (Life Technologies, #14190-169). Uterine deciduas were removed and each embryo in its yolk sac was transferred to a fresh dish with 10-15 ml PBS. The yolk sac was removed. The embryo head, the tail, the spleen and liver were removed and the tail was saved for genotyping. Embryos were minced in very small pieces with a scalpel in PBS and transferred in a 15 ml tube. Minced embryos in PBS were centrifuged at 1200 rpm for 5 minutes at RT. PBS was discarded and 5 ml of Trypsin/EDTA 0.05% (Life Technologies, #25300-054) was added. Tubes were wrapped with aluminium foil and incubated at 37ºC on a shaker at 50 rpm for 15 min. The tissue fragments were allowed to settle and supernatant was collected and transferred to 50 ml tubes containing 10 ml isolation medium. Addition of 5 ml Trypsin/EDTA and incubation at 37ºC was repeated once more and the supernatants collected from both Trypsin/EDTA incubations were combined and centrifuged at 1200 rpm for 5 min at RT. The pellets of each single embryo were resuspended in 12 ml of fresh MEFs medium containing DMEM

Glutamax (Life Technologies, #61965-059), 10% FCS (Life Technologies, #15140-122), 100 U/ml penicillin, 100 μg/ml streptomycin (Life Technologies, #15140-122), 2 mM L-Glutamine (Life Technologies, #25030-081), 1 mM Na-Pyruvate (Life Technologies, #11360-039), 1X Non Essential Amino Acids -NEAA (Life Technologies, #11140 – 035), 50 μM 2-mercaptoethanol (Life Technologies, #31350-010) and plated in a 10 cm TC plate (P1). The following day cells were 75% confluent and they were trypsinised, counted and replated at a density of $3x10^6$ cells/10 cm TC plate (P2). Two days later, P2 cells were trypsinized and frozen at -80ºC in cryovials at a concentration of $5x10^6$/vial in 1 ml cryopreserving medium (90% MEFs medium, 10% DMSO). After 24-36 hours cryovials were transferred to liquid N_2 for long term storage. MEFs were inactivated with 10 μg/ml mitomycin C (Sigma, #M0503) in normal culture medium for 3.5 hrs in 37ºC incubator before being used as a feeder layer for iPSCs.

4.3 Cell Culture and EBs differentiation

MEFs were cultured as described above. iPS cells were grown on 0.1% gelatin-coated plates on a feeder layer of mitomycin inactivated MEFs in ES medium, containing DMEM Glutamax (Life Technologies, #61965-059), 15%mFCS (Life Technologies, #10439024), 100 U/ml penicillin, 100 μg/ml streptomycin (Life Technologies, #15140-122), 2 mM L-Glutamine (Life Technologies, #25030-081), 1 mM Na-Pyruvate (Life Technologies, #11360-039), 1X Non Essential Amino Acids - NEAA (Life Technologies, #11140 – 035), 50 μM 2-mercaptoethanol (Life Technologies, #31350-010), 10^3 u/ml Leukemia Inhibitory Factor (LIF) (Millipore, #ESG1107). Medium was changed daily and cells were passaged every 2-3 days. mESCs (ES-E14) were cultured in ES medium on 0.1% gelatin-coated plates. Medium was changed daily and cells were passaged every 2-3 days.

For reprogramming experiments, MEFs of the indicated genotype at P3 were plated at a density of $2x10^4$ cells/cm^2 on 0.1% gelatin-coated plates and cultured in ES medium with 1 μg/ml doxycycline, changed daily to ensure constant activation of the tetO-OKSM transgene. After 14-18 days, the GFP fluorescence signal was acquired on the whole plate using Incucyte® Live-Cell Analysis Systems. GFP positive colonies were manually

picked and expanded on feeder layer of inactivated MEFs, to establish independent iPSCs lines.

For EBs differentiation, mESCs were dissociated with pre warmed 1x trypsin solution and resuspended to a concentration of $3x10^4$ cells/ml in ES medium without LIF. We plated 30 μL drops containing 900 mES cells on the lids of 10 mm dishes filled with 10 mL of PBS to prevent the drops from drying out and we let the EBs grow for 3 days. Then we harvested EBs in 60 mm plates in ES medium without LIF, imaged them and lysed them for RNA extraction. Images were analysed using a custom macro in FIJI software, to score area and circularity of EBs.

4.4 Plasmid and lentiviral particles production

pLKO.1 plasmid for the lentiviral delivery of shRNA sequences in mammalian cells was cloned to insert the specific shRNA sequence and blasticidin was used for selection. shRNA hairpin sequences specific for StemHistones transcripts were designed to target respectively the 3' UTR of H1f6; the coding sequence of H2bu2; and both 5' and 3' UTR of H3f4. A control shRNA sequence that does not target any mammalian transcript was used as control (Scramble). The designed template oligos were phosphorylated and annealed. The oligos were ligated in recipient plasmid pLKO.1 digested with AgeI and EcoRI to obtain pLKO-shStemHistones (shH1f6, sh H2bu2, shH3f4_1, shH3f4_2).

Lentiviral plasmids for the expression of FLAG-tagged StemHistones were obtained using an available lentiviral backbone, modified using restriction digestion cloning. The E1fα promoter was inserted using EcoRV-MluI, StemHistone-3xFLAG (H1f6-3xFLAG, H2bu2-3xFLAG, H3f4-3xFLAG) were added downstream using BamHI-XbaI and an IRES-BlastR cassette was cloned further downstream using BstXI-SalI.

For lentiviral particle production, HEK 293T cells were cultured in DMEM (Gibco) plus 10% fetal bovine serum (Gibco), penicillin 100 U/mL, streptomycin 100 U/mL, glutamine and sodium pyruvate (Gibco). Two hours before transfection, medium was changed to IMDM (sigma) 10% FBS, penicillin 100 U/mL, streptomycin 100 U/mL, glutamine. $9x10^6$ low passage HEK 293T cells were transfected with the packaging plasmids pDM2-VSVG and pCVM-ΔR8.91 and the target vector using

calcium/phosphate method. A plasmid solution (7 µg pDM2-VSVG, 28 µg pCMV-ΔR8.91, 32 µg pLKO-shStemHistones) was mixed with 0.125 M $CaCl_2$ in a final volume of 1250 µL and incubated at room temperature for 5 minutes; the CaPi-DNA precipitate was formed by dropwise addition of 1250 µL of 2xHBS (to a final concentration of NaCl 140.5 mM; HEPES 50 mM; Na_2HPO_4 0.75 mM pH 7.12); the precipitates were immediately added dropwise to the cell media. Virus-containing media was collected 36 hours after transfection, centrifuged 5 minutes at 200 g and filtered through a 0.22 µm membrane. mESCs were infected with lentivirus containing medium in presence of 8 µg/ml polybrene. Cells were selected with blasticidin at final concentration of 10 µg/mL for 7 days.

4.5 CRISPR/Cas9-mediated knockout

We used CRISPR/Cas9 technology to generate RNA-guided double strand breaks in DNA to produce mESCs StemHistones knockouts (mESC-KO). We designed 20 bp oligos based on H1f6, H2bu2 and H3f4 sequences to target 3' (sgRNA2, sgRNA3) and 5' regions (sgRNA1) in the transcript. The guides had high score in VBC-score (https://www.vbc-score.org/) and a good score in CRISPick (https://portals.broadinstitute.org/gppx/crispick/public). We used a publicly available plasmid, lentiCRISPRv2. This plasmid contains two expression cassettes, hSpCas9 and the chimeric guide RNA. Annealed oligos designed for the targeting of StemHistones genes were cloned upstream of the single guide RNA scaffold using BsmBI restriction enzyme. The lentiCRISPRv2 plasmids encoding sgRNAs targeting StemHistone genes were transfected in mESCs, using Lipofectamin 2000. We plated $5x10^5$ mESCs just before transfection, using 2 µg of each plasmid. 24 h after transfection the medium was changed with ES medium containing puromycin at final concentration of 1 µg/mL and selection was continued for 3 days. Afterwards, cells were sparsely seeded on a 15 cm plate for clone-picking. To improve the efficiency of editing, we used two different combination of guides per histone (sgRNA1 + sgRNA2; sgRNA1 + sgRNA3).

Genomic DNA used for screening colonies positive to CRISPR/Cas9-medaited KO was extracted from 96 well plates containing single cell clones. When mESCs cell were confluent, we lysed them with Bradley Lysis Buffer containing Proteinase K (10 mM Tris-HCl pH 7.5; 10 mM EDTA; 0.5% SDS; 10 mM NaCl; 1 mg/mL Proteinase K)

following the manufacturer's protocol. For PCR we used GoTaq G2 DNA polymerase from Promega and we followed the manufacturer's instruction.

4.6 Alkaline phosphatase assay

In order to estimate the reprogramming efficiency, we perfomed alkaline phosphatase staining using Vector® Red Substrate Kit. Briefly, reprogramming plates at day 15 were fixed with 4% paraformaldehyde (PFA) for 15 minutes. Plates were washed twice with PBS and left overnight at 4C in PBS. Then, AP staining was performed following the manufacturer's instructions. We prepared 150 mM Tris-HCl, pH 8 and we added 3 drops (80 μL) of Vector reagent 1, reagent 2 and reagent 3. We incubated cells for 20-30 minutes in the dark and then we washed cells with PBS twice. For the acquisition, we used the Incucyte® Live-Cell Analysis Systems, which takes a picture of the whole well with an acquisition time of 400 ms for red fluorescence.

We used the same kit to evaluate the pluripotent state of mESCs in colony forming assay. We plated 1000 cells on a 6 well plate, and we let them grow as clearly distinguishable single colonies. After 6-7 days cells were fixed with 4% paraformaldehyde (PFA) and processed as described above.

4.7 Protein extraction and western blot

For histone content quantification, cells were lysed directly in 5% Laemmli reducing buffer. Lysates, before incubation in 95^{o}C were subjected to DNA quantification using Picogreen KIT (Invitrogen). Equal amounts of DNA were loaded in each lane and each sample from each experiment run in technical replicates. For iPSCs a standard curve was designed using 100 ng, 200 ng, 300 ng DNA equivalent. For StemHistones-FLAG detection, cells were lysed in RIPA buffer (25 mM Tris HCl pH 7.6; 150 mM NaCl; 1% NP-40 sodium deoxycholate; 0,1% SDS; protease inhibitors 1x; DTT 1 mM). Samples were sonicated, denatured 5 minutes at 95°C and separated using SDS-PAGE. Resolving gel consisted of polymerised acrylamide-bisacrylamide (ratio 29:1) in 375 mM Tris pH 8.8, 0.1% SDS, at concentrations 15%. The stacking gel consisted of 4% acrylamide-bisacrylamide, 125 mM Tris pH 6.8, 0.1% SDS. The gels were run at 5-15 V/cm in

running buffer (144 g/l glycine, 30 g/l Tris, 0.5% SDS). Proteins separated by SDS-PAGE were transferred onto 0.2 μm nitrocellulose membranes using Trans-Blot® Turbo™ Transfer System (Bio-Rad). Membranes were blocked for 1 hour at room temperature in 5% skim milk/TBST (20 mM Tris, pH 7.5, 137 mM NaCl, 0.1% Tween 20), washed in TBST and incubated overnight with primary antibodies, rabbit anti-H3 (Abcam), mouse anti-FLAG (Cell Signaling) and mouse anti-Actin (Sigma). An ECL chromogenic substrate (ImmunoCruz™ Western Blotting Luminol) was used to visualize the bands. Ponceau staining was used for assessing equivalent protein loading.

4.8 Histone isolation and digestion

Briefly, an equal number of cells per sample were resuspended in lysis buffer (10% sucrose; 0.5 mM EGTA, pH 8.0; 15 mM NaCl; 60 mM KCl; 15 mM HEPES; 0.5% Triton; 0.5 mM PMSF; 1 mM DTT; 5 mM NaF; 5 mM Na_3VO_4; 5 mM Na-butyrate; protease inhibitors) and nuclei were separated from the cytoplasm by centrifugation on a sucrose cushion. Histones were extracted by a 4 hours incubation in 0.4 M HCl at 4°C and dialyzed overnight against 100 mM CH_3COOH, using dialysis membranes with a 6–8 kDa cutoff (Spectrum Laboratories, Inc, Rancho Dominguez, CA). About 5–10 μg of histones per run per sample were mixed with an approximately equal amount of super-SILAC mix. For in-gel digestions, bands corresponding to histone H3 were excised, chemically alkylated with D6-acetic anhydride (Merck KGaA, Darmstadt, Germany), and in-gel digested with trypsin (the combination of chemical alkylation and trypsin digestion generates an “Arg-C-like” digestion). Digested peptides were desalted and concentrated using a combination of reversed-phase C18/C and strong cation exchange (SCX) chromatography on handmade nanocolumns (StageTips). Digested peptides were then eluted with 80% ACN/0.5% acetic acid and 5% NH_4OH/30% methanol from C18/C and SCX StageTips, respectively. Eluted peptides were lyophilized, resuspended in 1% TFA, pooled and subjected to LC-MS/MS analysis.

4.9 LC-MS/MS

Peptide mixtures were separated by reversed-phase chromatography on an in-house-made 25 cm column (inner diameter 75 μm, outer diameter 350 μm outer diameter, 1.9

μm ReproSil, Pur C18AQ medium), using a ultra nanoflow high-performance liquid chromatography (HPLC) system (EASY-nLC™ 1000, Thermo Fisher Scientic) connected online to a Q Exactive instrument (Thermo Fisher Scientific) through a nanoelectrospray ion source. Solvent A was 0.1% formic acid (FA) in ddH2O and solvent B was 80% ACN plus 0.1% FA. Peptides were injected in an aqueous 1% TFA solution at a flow rate of 500 nl/min and were separated with a 100 min linear gradient of 0–40% solvent B, followed by a 5 min gradient of 40–60% and a 5 min gradient of 60–95% at a flow rate of 250 nl/min. The Q Exactive instrument was operated in the data-dependent acquisition (DDA) mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS spectra (m/z 300–1650) were analyzed in the Orbitrap detector with resolution of 35,000 at m/z 400. The five most intense peptide ions with charge states ≥ 2 were sequentially isolated to a target value for MS1 of 3×10^6 and fragmented by HCD with a normalized collision energy setting of 25%. The maximum allowed ion accumulation times were 20 msec for full scans and 50 msec for MS/MS and the target value for MS/MS was set to 1×10^6. The dynamic exclusion time was set to 20 s and the standard mass spectrometric conditions for all experiments were as follows: spray voltage of 2.4 kV, no sheath and auxiliary gas flow.

4.10 Histone peptides and PTM Data Analysis

Acquired raw data were analyzed using the integrated MaxQuant software v.1.5.2.8 (Max Planck Institute of Biochemistry; https://maxquant.org/), which performed peak list generation and protein identification using the Andromeda search engine. The Uniprot MOUSE 1301 (33202 entries) database was used for histone peptide identification. Enzyme specificity was set to Arg-C. The estimated false discovery rate (FDR) of all peptide identifications was set at a maximum of 1%. The mass tolerance was set to 6 ppm for precursor and fragment ions. One missed cleavage was allowed and the minimum peptide length was set to 6 amino acids. Variable modifications for in-gel digestions included lysine D6-acetylation (+45.0294 Da), lysine monomethylation (+59.0454, corresponding to the sum of D6-acetylation (+45.0294) and monomethylation (+14.016 Da), dimethylation (+28.031 Da), trimethylation (+42.046 Da), and lysine acetylation (+42.010 Da). To reduce the search time and the rate of false positives, which increase

with increasing the number of variable modifications included in the database search, the raw data were analyzed through multiple parallel search jobs, setting different combinations of variable modifications, as previously described [15]. Variable modifications for in-solution Arg-C digestions were lysine monomethylation (+14.016 Da), dimethylation (+28.031 Da), trimethylation (+42.046 Da), and acetylation (+42.010 Da). Peptides with Andromeda scores less than 60 and localization probability scores less than 0.75 were removed. Identifications, retention times, and elution patterns of isobaric peptides were used to guide the manual quantification of each modified peptide using QualBrowser version 2.0.7 (Thermo Fisher Scientific, Waltham, MA, USA). Site assignment was evaluated using QualBrowser (Thermo Fisher Scientific, Waltham, MA, USA) and MaxQuant Viewer (Max Planck Institute of Biochemistry https://maxquant.org/) from MS2 spectra. Extracted ion chromatograms (XICs) were constructed for each doubly/triply charged precursor, based on its m/z value, using a mass tolerance of 10 ppm and a mass precision up to four decimals. For each histone modified peptide, the % relative abundance (%RA) was estimated by dividing the area under the curve (AUC) of each modified peptide for the sum of the areas corresponding to all the observed forms of that peptide and multiplying by 100. For SILAC experiments, Arg10 was selected as a heavy label (multiplicity = 2) in MaxQuant. The heavy form of each modified peptide was quantified from its XIC and the relative abundance was quantified. Heavy peptides without a light counterpart were not considered for quantification. L/H ratios of relative abundances were calculated for each modified peptide.

4.11 RNA extraction and Real Time PCR analysis

Total RNA from iPSCs, mESCs and EBs was isolated using the NucleoSpin RNA II kit (Macherey-Nagel). For iPSCs, cells were feeder-depleted before RNA extraction. Briefly, cells were trypsinized, resuspended with medium and then placed back in the incubator for 30’ to allow feeders to attach back to the plate. Floating iPSCs were collected and processed as indicated. Reverse transcription was performed using SuperScript™ VILO™ cDNA Synthesis Kit (ThermoScientific) following the manufacturer’s instructions. Quantitative real-time PCR was performed in a LightCycler480 (Roche) apparatus using SYBR Green I master mix (Roche). Ct is calculated on technical triplicates using LightCycler480 Software

(Roche) and analysis is performed using the ΔCt method, Atp5F1 and GAPDH were used as housekeeping genes.

4.12 Total RNAseq

mESCs and EBs were lysed using TRIzol reagent, and total RNA extracted using RNA Clean and Concentrator kit from Zymo Research, following the manufacturer's instructions. Libraries were prepared by the Center of Omic Sciences of San Raffaele Hospital (COSR), using TruSeq Stranded Total RNA kit with Ribo-Zero. Libraries were run on NovaSeq 6000 platform (paired-end, 100 nucleotides per read). After trimming of the adapter sequences (cutadapt, https://cutadapt.readthedocs.io) reads were mapped to mouse genome (mm10) using hisat2 (http://daehwankimlab.github.io/hisat2/) using parameters "-p 20 -5 5". Read counting was performed using featureCounts from the Subread Package and features displaying less than 10 reads were filtered out. Differential expression analysis was performed using DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html). Gene ontology was performed using the "clusterProfiler" R package (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html). Heatmap and Hierarchical clustering was performed using the "Pheatmap" R package.

4.13 Immunofluorescent staining

iPSCs and mESCs were grown for two-four days on feeder layer-coated or 0.1% gelatine glass coverslips, respectively. Then, cells were fixated with 4% PFA for 10 minutes at room temperature, washed three times with PBS, and permeabilized using PBS-T (0.1% Triton in PBS) for 10 min. Fixed and permeabilized cells were then blocked with blocking solution (2% normal goat serum, 0.1% BSA, 0.05% Tween in PBS) for 1 hour at room temperature, followed by incubation with primary antibody diluted in blocking solution overnight at 4°C. The antibodies used were mouse anti-FLAG (Cell Signaling), rabbit anti-Oct4 (Abcam), rabbit anti-Nanog (abcam). Cells were washed three times with washing solution (0.05% Tween in PBS). Appropriate Alexa Fluor (Alexa 488 or Alexa 546)-conjugate antibodies (1:500; Invitrogen) were used as second-step reagent. Nuclei were counterstained with Hoechst (1:2000; Sigma Aldrich), and

coverslips were mounted with Fluorescence Mounting Medium (Dako). Fluorescent and phase contrast images were taken by using Nikon Eclipse E600 or Zeiss Imager M2 microscopes. Image acquisition was done by using the Nikon digital camera DXM1200 or AxioCam MRc5 camera and the acquisition software ACT-1 (Plan Fluor Ienses: X4/0.13, X10/0.33, X20/0.50, X40/0.75) or AxioVision. Images showing double fluorescence were first acquired separately by using appropriate filters, and then the different layers were merged by using Fiji ImageJ.

REFERENCES

Ahmad K & Henikoff S (2002) The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Molecular Cell* 9: 1191–1200

Almeida R, Fernández-Justel JM, Santa-María C, Cadoret JC, Cano-Aroca L, Lombraña R, Herranz G, Agresti A & Gómez M (2018) Chromatin conformation regulates the coordination between DNA replication and transcription. *Nature Communications 2018 9:1* 9: 1–14

Amlani B, Liu Y, Chen T, Ee LS, Lopez P, Heguy A, Apostolou E, Kim SY & Stadtfeld M (2018) Nascent Induced Pluripotent Stem Cells Efficiently Generate Entirely iPSC-Derived Mice while Expressing Differentiation-Associated Genes. *Cell Reports* 22: 876–884

Babiarz JE, Halley JE & Rine J (2006) Telomeric heterochromatin boundaries require NuA4-dependent acetylation of histone variant H2A.Z in Saccharomyces cerevisiae. *Genes Dev* 20: 700–710

Banaszynski LA, Wen D, Dewell S, Whitcomb SJ, Lin M, Diaz N, Elsässer SJ, Chapgier A, Goldberg AD, Canaani E, *et al* (2013) Hira-Dependent Histone H3.3 Deposition Facilitates PRC2 Recruitment at Developmental Loci in ES Cells. *Cell* 155: 107–120

Bar-Nur O, Brumbaugh J, Verheul C, Apostolou E, Pruteanu-Malinici I, Walsh RM, Ramaswamy S & Hochedlinger K (2014) Small molecules facilitate rapid and synchronous iPSC generation. *Nature Methods 2014 11:11* 11: 1170–1176

Battle DJ & Doudna JA (2001) The stem-loop binding protein forms a highly stable and specific complex with the 3′ stem-loop of histone mRNAs. *RNA* 7: 123–132

ten Berge D, Koole W, Fuerer C, Fish M, Eroglu E & Nusse R (2008) Wnt Signaling Mediates Self-Organization and Axis Formation in Embryoid Bodies. *Cell Stem Cell* 3: 508–518

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, *et al* (2006a) A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125: 315–326

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, *et al* (2006b) A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125: 315–326

Bonaldi T, Längst G, Strohner R, Becker PB & Bianchi ME (2002) The DNA chaperone HMGB1 facilitates ACF/CHRAC-dependent nucleosome sliding. *The EMBO Journal* 21: 6865–6873

Bönisch C & Hake SB (2012) Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic Acids Research* 40: 10719–10741

Bradley A, Evans M, Kaufman MH & Robertson E (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature 1984 309:5965* 309: 255–256

Brambilla F, Garcia-Manteiga JM, Monteleone E, Hoelzen L, Zocchi A, Agresti A & Bianchi ME (2020) Nucleosomes effectively shield DNA from radiation damage in living cells. *Nucleic Acids Research* 48: 8993–9006

Brickman JM & Serup P (2017) Properties of embryoid bodies. *Wiley Interdisciplinary Reviews: Developmental Biology* 6: e259

van den Brink SC, Baillie-Johnson P, Balayo T, Hadjantonakis AK, Nowotschin S, Turner DA & Arias AM (2014) Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development (Cambridge)* 141: 4231–4242

Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, Simeone A, Tan M, Swigut T & Wysocka J (2014) Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency. *Cell Stem Cell* 14: 838–853

Canzio D, Chang EY, Shankar S, Kuchenbecker KM, Simon MD, Madhani HD, Narlikar GJ & Al-Sady B (2011) Chromodomain-Mediated Oligomerization of HP1 Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly. *Molecular Cell* 41: 67–81

Cao K, Lailler N, Zhang Y, Kumar A, Uppal K, Liu Z, Lee EK, Wu H, Medrzycki M, Pan C, *et al* (2013) High-Resolution Mapping of H1 Linker Histone Variants in Embryonic Stem Cells. *PLOS Genetics* 9: e1003417

Carey BW, Markoulaki S, Beard C, Hanna J & Jaenisch R (2009) Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nature Methods 2009 7:1* 7: 56–59

Celeste A, Petersen S, Romanienko PJ, Fernandez-Capetillo O, Chen HT, Sedelnikova OA, Reina-San-Martin B, Coppola V, Meffre E, Difilippantonio MJ, *et al* (2002) Genomic instability in mice lacking histone H2AX. *Science (1979)* 296: 922–927

Celona B, Weiner A, Felice D di, Mancuso F & Cesarini FM (2011a) Substantial Histone Reduction Modulates Genomewide Nucleosomal Occupancy and Global Transcriptional Output. *PLoS Biol* 9: 1001086

Celona B, Weiner A, di Felice F, Mancuso FM, Cesarini E, Rossi RL, Gregory L, Baban D, Rossetti G, Grianti P, *et al* (2011b) Substantial Histone reduction modulates Genomewide nucleosomal occupancy and global transcriptional output. *PLoS Biology* 9

Chakravarthy S, Patel A & Bowman GD (2012) The basic linker of macroH2A stabilizes DNA at the entry/exit site of the nucleosome. *Nucleic Acids Research* 40: 8285–8295

Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S & Smith A (2003) Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell* 113: 643–655

Cheloufi S, Elling U, Hopfgartner B, Jung YL, Murn J, Ninova M, Hubmann M, Badeaux AI, Euong Ang C, Tenen D, *et al* (2015) The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* 528: 218–224

Cheloufi S & Hochedlinger K (2017) Emerging roles of the histone chaperone CAF-1 in cellular plasticity. *Current Opinion in Genetics and Development* 46: 83–94 doi:10.1016/j.gde.2017.06.004 [PREPRINT]

Chen J, Liu H, Liu J, Qi J, Wei B, Yang J, Liang H, Chen Y, Chen J, Wu Y, *et al* (2012) H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature Genetics 2012 45:1* 45: 34–42

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, *et al* (2008) Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* 133: 1106–1117

Clapier CR, Iwasa J, Cairns BR & Peterson CL (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Publishing Group* 18

Cossec JC, Theurillat I, Chica C, Búa Aguín S, Gaume X, Andrieux A, Iturbide A, Jouvion G, Li H, Bossis G, *et al* (2018) SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* 23: 742-757.e8

Dion MF, Altschuler SJ, Wu LF & Rando OJ (2005) Genomic characterization reveals a simple histone H4 acetylation code. *Proc Natl Acad Sci U S A* 102: 5501–5506

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, *et al* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology 2015 34:2* 34: 184–191

Dorigo B, Schalch T, Bystricky K & Richmond TJ (2003) Chromatin Fiber Folding: Requirement for the Histone H4 N-terminal Tail. *Journal of Molecular Biology* 327: 85–96

Drabent B, Bode C & Doenecke D (1993a) Structure and expression of the mouse testicular H1 histone gene (H1t). *Biochim Biophys Acta* 1216: 311–313

Drabent B, Bode C & Doenecke D (1993b) Structure and expression of the mouse testicular H1 histone gene (H1t). *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1216: 311–313

Durrin LK, Mann RK, Kayne PS & Grunstein M (1991) Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. *Cell* 65: 1023–1031

Elsässer SJ, Huang H, Lewis PW, Chin JW, Allis CD & Patel DJ (2012) DAXX envelops a histone H3.3–H4 dimer for H3.3-specific recognition. *Nature 2012 491:7425* 491: 560–565

Fan Y, Sirotkin A, Russell RG, Ayala J & Skoultchi AI (2001a) Individual Somatic H1 Subtypes Are Dispensable for Mouse Development Even in Mice Lacking the H1 0 Replacement Subtype. *Molecular and Cellular Biology* 21: 7933–7943

Fan Y, Sirotkin A, Russell RG, Ayala J & Skoultchi AI (2001b) Individual Somatic H1 Subtypes Are Dispensable for Mouse Development Even in Mice Lacking the H1 0 Replacement Subtype . *Molecular and Cellular Biology* 21: 7933–7943

Fujii-Yamamoto H, Jung MK, Arai KI & Masai H (2005) Cell cycle and developmental regulations of replication factors in mouse embryonic stem cells. *Journal of Biological Chemistry* 280: 12976–12987

Gajović S, St-Onge L, Yokota Y & Gruss P (1998) Retinoic acid mediates Pax6 expression during in vitro differentiation of embryonic stem cells. *Differentiation* 62: 187–192

Gaspar-Maia A, Qadeer ZA, Hasson D, Ratnakumar K, Adrian Leu N, Leroy G, Liu S, Costanzi C, Valle-Garcia D, Schaniel C, *et al* (2013) MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. *Nature Communications 2013 4:1* 4: 1–13

Ge E, Nora P, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, *et al* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre.

Geiger T, Cox J, Ostasiewicz P, Wisniewski JR & Mann M (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods 2010 7:5* 7: 383–385

Golas A, Dzieza A, Kuzniarz K & Styrna J (2003) Gene mapping of sperm quality parameters in recombinant inbred strains of mice. *International Journal of Developmental Biology* 52: 287–293

Goodwin GH, Sanders C & Johns EW (1973) A New Group of Chromatin-Associated Proteins with a High Content of Acidic and Basic Amino Acids. *European Journal of Biochemistry* 38: 14–19

Grewal SIS & Elgin SCR (2002) Heterochromatin: new possibilities for the inheritance of structure. *Current Opinion in Genetics & Development* 12: 178–187

Gurdon JB (1962) The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, *et al* (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature 2011 477:7364* 477: 295–300

Hamazaki T, Oka M, Yamanaka S & Terada N (2004) Aggregation of embryonic stem cells induces Nanog repression and primitive endoderm differentiation. *Journal of Cell Science* 117: 5681–5686

Hao F, Kale S, Dimitrov S & Hayes JJ (2021) Unraveling linker histone interactions in nucleosomes. *Current Opinion in Structural Biology* 71: 87–93

Harada A, Maehara K, Ono Y, Taguchi H, Yoshioka K, Kitajima Y, Xie Y, Sato Y, Iwasaki T, Nogami J, *et al* (2018) Histone H3.3 sub-variant H3mm7 is required for normal skeletal muscle regeneration. *Nature Communications 2018 9:1* 9: 1–13

Hayakawa K, Tani R, Nishitani K & Tanaka S (2020a) Linker histone variant H1T functions as a chromatin de-condenser on genic regions. *Biochemical and Biophysical Research Communications* 528: 685–690

Hayakawa K, Tani R, Nishitani K & Tanaka S (2020b) Linker histone variant H1T functions as a chromatin de-condenser on genic regions. *Biochemical and Biophysical Research Communications* 528: 685–690

Hendzel MJ, Lever MA, Crawford E & Th'Ng JPH (2004) The C-terminal Domain Is the Primary Determinant of Histone H1 Binding to Chromatin in Vivo. *Journal of Biological Chemistry* 279: 20028–20034

Hirai S, Tomimatsu K, Miyawaki-Kuwakado A, Takizawa Y, Komatsu T, Tachibana T, Fukushima Y, Takeda Y, Negishi L, Kujirai T, *et al* (2022) Unusual nucleosome formation and transcriptome influence by the histone H3mm18 variant. *Nucleic Acids Research* 50: 72–91

Hirota T, Lipp JJ, Toh BH & Peters JM (2005) Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature 2005 438:7071* 438: 1176–1180

Hu G, Cui K, Northrup D, Liu C, Wang C, Tang Q, Ge K, Levens D, Crane-Robinson C & Zhao K (2013) H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* 12: 180–192

Ishiuchi T, Enriquez-Gasca R, Mizutani E, Bošković A, Ziegler-Birling C, Rodriguez-Terrones D, Wakayama T, Vaquerizas JM & Torres-Padilla ME (2015) Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nature Structural & Molecular Biology 2015 22:9* 22: 662–671

Jenuwein T & Allis CD (2001) Translating the histone code. *Science (1979)* 293: 1074–1080

Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K & Felsenfeld G (2009) H3.3/H2A.Z double variant–containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. *Nature Genetics 2009 41:8* 41: 941–945

Karnavas T, Pintonello L, Agresti A & Bianchi ME (2014) Histone content increases in differentiating embryonic stem cells. *Frontiers in Physiology* 5 AUG

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K & Henikoff S (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications 2019 10:1* 10: 1–10

Khadake JR & Rao MRS (1995) DNA- and Chromatin-Condensing Properties of Rat Testes Hla and Hit Compared to Those of Rat Liver Hlbdec: Hit Is a Poor Condenser of Chromatin. *Biochemistry* 34: 15792–15801

Knaupp AS, Buckberry S, Pflueger J, Lim SM, Ford E, Larcombe MR, Rossello FJ, de Mendoza A, Alaei S, Firas J, *et al* (2017) Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell* 21: 834-845.e6

Kolundzic E, Ofenbauer A, Bulut SI, Uyar B, Baytek G, Sommermeier A, Seelk S, He M, Hirsekorn A, Vucicevic D, *et al* (2018) FACT Sets a Barrier for Cell Fate Reprogramming in Caenorhabditis elegans and Human Cells. *Developmental Cell* 46: 611-626.e12

Kornberg RD (1977) Structure of chromatin. *Annu Rev Biochem* 46: 931–954

Kubo A, Shinozaki K, Shannon JM, Kouskoff V, Kennedy M, Woo S, Fehling HJ & Keller G (2004) Development of definitive endoderm from embryonic stem cells in culture. *Development* 131: 1651–1662

Lachner M, O'Carroll D, Rea S, Mechtler K & Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410: 116–120

Lichter P, Cremer T, Borden J, Manuelidis L & Ward DC (1988) Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, *et al* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (1979)* 326: 289–293

Lin Q, Sirotkin A & Skoultchi AI (2000) Normal Spermatogenesis in Mice Lacking the Testis-Specific Linker Histone H1t. *Molecular and Cellular Biology* 20: 2122–2128

Ling X, Harkness TAA, Schultz MC, Fisher-Adams G & Grunstein M (1996) Yeast histone H3 and H4 amino termini are important for nucleosome assembly in vivo and in vitro: redundant and position-independent functions in assembly but not in gene regulation. *Genes & Development* 10: 686–699

Long M, Sun X, Shi W, Yanru A, Leung STC, Ding D, Cheema MS, MacPherson N, Nelson CJ, Ausio J, *et al* (2019) A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Research* 47: 8399–8409

Luger K, Mäder AW, Richmond RK, Sargent DF & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature 1997 389:6648* 389: 251–260

Luger K & Richmond TJ (1998) The histone tails of the nucleosome. *Current Opinion in Genetics and Development* 8: 140–146

MacHida S, Hayashida R, Takaku M, Fukuto A, Sun J, Kinomura A, Tashiro S & Kurumizaka H (2016) Relaxed Chromatin Formation and Weak Suppression of Homologous Pairing by the Testis-Specific Linker Histone H1T. *Biochemistry* 55: 637–646

Maehara K, Harada A, Sato Y, Matsumoto M, Nakayama KI, Kimura H & Ohkawa Y (2015) Tissue-specific expression of histone H3 variants diversified after species separation. *Epigenetics and Chromatin* 8: 1–17

Mahadevan IA, Kumar S & Rao MRS (2020) Linker histone variant H1t is closely associated with repressed repeat-element chromatin domains in pachytene spermatocytes. *Epigenetics and Chromatin* 13

Martello G, Bertone P & Smith A (2013) Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *The EMBO Journal* 32: 2561–2574

Martello G & Smith A (2014) The Nature of Embryonic Stem Cells. *http://dx.doi.org/101146/annurev-cellbio-100913-013116* 30: 647–675

Martin AM, Pouchnik DJ, Walker JL & Wyrick JJ (2004) Redundant Roles for Histone H3 N-Terminal Lysine Residues in Subtelomeric Gene Repression in Saccharomyces cerevisiae. *Genetics* 167: 1123–1132

Martire S & Banaszynski LA (2020) The roles of histone variants in fine-tuning chromatin organization and function. *Nature Reviews Molecular Cell Biology 2020 21:9* 21: 522–541

Martire S, Gogate AA, Whitmill A, Tafessu A, Nguyen J, Teng YC, Tastemel M & Banaszynski LA (2019) Phosphorylation of histone H3.3 at serine 31 promotes p300 activity and enhancer acetylation. *Nature Genetics 2019 51:6* 51: 941–946

Marzluff WF (2005) Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Current Opinion in Cell Biology* 17: 274–280

Marzluff WF, Gongidi P, Woods KR, Jin J & Maltais LJ (2002) The Human and Mouse Replication-Dependent Histone Genes. *Genomics* 80: 487–498

Marzluff WF, Wagner EJ & Duronio RJ (2008) Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail. *Nature Reviews Genetics* 9: 843–854 doi:10.1038/nrg2438 [PREPRINT]

Matsuda T, Nakamura T, Nakao K, Arai T, Katsuki M, Heike T & Yokota T (1999) STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *The EMBO Journal* 18: 4261–4269

Michlits G, Jude J, Hinterndorfer M, de Almeida M, Vainorius G, Hubmann M, Neumann T, Schleiffer A, Burkard TR, Fellner M, *et al* (2020) Multilayered VBC score predicts sgRNAs that efficiently generate loss-of-function alleles. *Nature Methods 2020 17:7* 17: 708–716

Molaro A, Young JM & Malik HS (2018) Evolutionary origins and diversification of testis-specific short histone H2A variants in mammals. *Genome Research* 28: 460–473

Nair RR, Mazza D, Brambilla F, Gorzanelli A, Agresti A & Bianchi ME (2018) LPS-challenged macrophages release microvesicles coated with histones. *Frontiers in Immunology* 9: 1463

Niwa H, Miyazaki JI & Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics 2000 24:4* 24: 372–376

Noberini R, Osti D, Miccolo C, Richichi C, Lupia M, Corleone G, Hong SP, Colombo P, Pollo B, Fornasari L, *et al* (2018) Extensive and systematic rewiring of histone post-

translational modifications in cancer model systems. *Nucleic Acids Research* 46: 3817–3832

Okamoto K, Okazawa H, Okuda A, Sakai M, Muramatsu M & Hamada H (1990) A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* 60: 461–472

Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, Gupta PB, *et al* (2012) Chromatin-modifying enzymes as modulators of reprogramming. *Nature 2012 483:7391* 483: 598–602

Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH & O'Shea CC (2017) ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science (1979)* 357

Pallier C, Scaffidi P, Chopineau-Proust S, Agresti A, Nordmann P, Bianchi ME & Marechal V (2003) Association of Chromatin Proteins High Mobility Group Box (HMGB) 1 and HMGB2 with Mitotic Chromosomes. *Molecular Biology of the Cell* 14: 3414–3426

Park YJ, Dyer PN, Tremethick DJ & Luger K (2004) A New Fluorescence Resonance Energy Transfer Approach Demonstrates That the Histone Variant H2AZ Stabilizes the Histone Octamer within the Nucleosome. *Journal of Biological Chemistry* 279: 24274–24282

Paull TT, Rogakou EP, Yamazaki V, Kirchgessner CU, Gellert M & Bonner WM (2000) A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Current Biology* 10: 886–895

Pengelly AR, Copur Ö, Jäckle H, Herzig A & Müller J (2013) A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb. *Science (1979)* 339: 698–699

Quivy JP, Gérard A, Cook AJL, Roche D & Almouzni G (2008) The HP1–p150/CAF-1 interaction is required for pericentric heterochromatin replication and S-phase progression in mouse cells. *Nature Structural & Molecular Biology 2008 15:9* 15: 972–979

Ramesh S, Bharath MMS, Chandra NR & Rao MRS (2006) A K52Q substitution in the globular domain of histone H1t modulates its nucleosome binding properties. *FEBS Letters* 580: 5999–6006

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, *et al* (2014a) A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159: 1665–1680

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, *et al* (2014b) A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159: 1665–1680

Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M & Cosma MP (2015) Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell* 160: 1145–1158

Rodriguez-Madoz JR, Jose-Eneriz ES, Rabal O, Zapata-Linares N, Miranda E, Rodriguez S, Porciuncula A, Vilas-Zornoza A, Garate L, Segura V, *et al* (2017) Reversible dual inhibitor against G9a and DNMT1 improves human iPSC derivation enhancing MET and facilitating transcription factor engagement to the genome. *PLOS ONE* 12: e0190275

Santoro SW & Dulac C (2012) The activity-dependent histone variant H2BE modulates the life span of olfactory neurons. *Elife* 2012

Savić N, Bär D, Leone S, Frommel SC, Weber FA, Vollenweider E, Ferrari E, Ziegler U, Kaech A, Shakhova O, *et al* (2014) lncRNA Maturation to Initiate Heterochromatin Formation in the Nucleolus Is Required for Exit from Pluripotency in ESCs. *Cell Stem Cell* 15: 720–734

Schalch T, Duda S, Sargent DF & Richmond TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre.

Sheban D, Shani T, Maor R, Aguilera-Castrejon A, Mor N, Oldak B, Shmueli MD, Eisenberg-Lerner A, Bayerl J, Hebert J, *et al* (2022) SUMOylation of linker histone

H1 drives chromatin condensation and restriction of embryonic cell fate identity. *Molecular Cell* 82: 106-122.e9

Shinagawa T, Takagi T, Tsukamoto D, Tomaru C, Huynh LM, Sivaraman P, Kumarevel T, Inoue K, Nakato R, Katou Y, *et al* (2014a) Histone Variants Enriched in Oocytes Enhance Reprogramming to Induced Pluripotent Stem Cells. *Cell Stem Cell* 14: 217–227

Shinagawa T, Takagi T, Tsukamoto D, Tomaru C, Huynh LM, Sivaraman P, Kumarevel T, Inoue K, Nakato R, Katou Y, *et al* (2014b) Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells. *Cell Stem Cell* 14: 217–227

Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR & Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science (1979)* 311: 844–847

Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW & Smith A (2008) Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. *PLOS Biology* 6: e253

Singh AM, Hamazaki T, Hankowski KE & Terada N (2007) A Heterogeneous Expression Pattern for Nanog in Embryonic Stem Cells. *Stem Cells* 25: 2534–2542

Sofiadis K, Josipovic N, Nikolic M, Kargapolova Y, Ubelmesser N€, Varamogianni-Mamatsi V, Zirkel A, Papadionysiou I, Loughran G, Keane J, *et al* (2021) HMGB1 coordinates SASP-related chromatin folding and RNA homeostasis on the path to senescence. *Molecular Systems Biology* 17: e9760

Soufi A, Donahue G & Zaret KS (2012) Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell* 151: 994–1004

Sridharan R, Gonzales-Cope M, Chronis C, Bonora G, McKee R, Huang C, Patel S, Lopez D, Mishra N, Pellegrini M, *et al* (2013) Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. *Nature Cell Biology 2013 15:7* 15: 872–882

di Stefano B, Collombet S, Jakobsen JS, Wierer M, Sardina JL, Lackner A, Stadhouders R, Segura-Morales C, Francesconi M, Limone F, *et al* (2016) C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nature Cell Biology* 18: 371–381

Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature 2000 403:6765* 403: 41–45

Szabo Q, Bantignies F & Cavalli G (2019) Principles of genome folding into topologically associating domains. *Science Advances* 5

Tachiwana H, Kagawa W, Osakabe A, Kawaguchi K, Shiga T, Hayashi-Takanaka Y, Kimura H & Kurumizaka H (2010) Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc Natl Acad Sci U S A* 107: 10454–10459

Tagami H, Ray-Gallet D, Almouzni G & Nakatani Y (2004) Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. *Cell* 116: 51–61

Takahashi K & Yamanaka S (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* 126: 663–676

Talbert PB, Ahmad K, Almouzni G, Ausiá J, Berger F, Bhalla PL, Bonner WM, Cande WZ, Chadwick BP, Chan SWL, *et al* (2012) A unified phylogeny-based nomenclature for histone variants. *Epigenetics and Chromatin* 5: 1–19

Talbert PB & Henikoff S (2021) Histone variants at a glance. *Journal of Cell Science* 134

Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, Buchou T, Cheng Z, Rousseaux S, Rajagopal N, *et al* (2011) Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* 146: 1016–1028

Tani R, Hayakawa K, Tanaka S & Shiota K (2016) Linker histone variant H1T targets rDNA repeats. *Epigenetics* 11: 288–302

Teif VB, Vainshtein Y, Caudron-Herger M, Mallm JP, Marth C, Höfer T & Rippe K (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nature Structural and Molecular Biology* 19: 1185–1192

Terme JM, Sesé B, Millán-Ariño L, Mayor R, Belmonte JCI, Barrero MJ & Jordan A (2011) Histone H1 Variants Are Differentially Expressed and Incorporated into Chromatin during Differentiation and Reprogramming to Pluripotency. *Journal of Biological Chemistry* 286: 35347–35357

Teves SS, An L, Hansen AS, Xie L, Darzacq X & Tjian R (2016) A dynamic mode of mitotic bookmarking by transcription factors. *Elife* 5

Tonge PD, Corso AJ, Monetti C, Hussein SMI, Puri MC, Michael IP, Li M, Lee DS, Mar JC, Cloonan N, *et al* (2014) Divergent reprogramming routes lead to alternative stem-cell states. *Nature 2014 516:7530* 516: 192–197

Torres CM, Biran A, Burney MJ, Patel H, Henser-Brownhill T, Cohen AHS, Li Y, Ben-Hamo R, Nye E, Spencer-Dene B, *et al* (2016) The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. *Science (1979)* 353

Tropberger P & Schneider R (2013) Scratching the (lateral) surface of chromatin regulation by histone modifications. *Nature Structural & Molecular Biology 2013 20:6* 20: 657–661

Trott J & Arias AM (2013) Single cell lineage analysis of mouse embryonic stem cells at the exit from pluripotency. *Biology Open* 2: 1049–1056

Turinetto V & Giachino C (2015) Histone variants as emerging regulators of embryonic stem cell identity. *Epigenetics* 10: 563–573 doi:10.1080/15592294.2015.1053682 [PREPRINT]

Ueda J, Harada A, Urahama T, Machida S, Maehara K, Hada M, Makino Y, Nogami J, Horikoshi N, Osakabe A, *et al* (2017) Testis-Specific Histone Variant H3t Gene Is Essential for Entry into Spermatogenesis. *Cell Reports* 18: 593–600

Vidal SE, Amlani B, Chen T, Tsirigos A & Stadtfeld M (2014) Combinatorial modulation of signaling pathways reveals cell-type-specific requirements for highly efficient and synchronous iPSC reprogramming. *Stem Cell Reports* 3: 574–584

Wang C, Liu X, Gao Y, Yang L, Li C, Liu W, Chen C, Kou X, Zhao Y, Chen J, *et al* (2018) Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nature Cell Biology 2018 20:5* 20: 620–631

Watson M, Stott K, Fischl H, Cato L & Thomas JO (2014) Characterization of the interaction between HMGB1 and H3—a possible means of positioning HMGB1 in chromatin. *Nucleic Acids Research* 42: 848–859

West JA, Cook A, Alver BH, Stadtfeld M, Deaton AM, Hochedlinger K, Park PJ, Tolstorukov MY & Kingston RE (2014) Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nature Communications* 5

Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA & Gough NM (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature 1988 336:6200* 336: 684–687

Yang J, Zhang X, Feng J, Leng H, Li S, Xiao J, Liu S, Xu Z, Xu J, Li D, *et al* (2016) The Histone Chaperone FACT Contributes to DNA Replication-Coupled Nucleosome Assembly. *Cell Reports* 14: 1128–1141

Yelagandula R, Stecher K, Novatchkova M, Michetti L, Michlits G, Wang J, Hofbauer P, Pribitzer C, Vainorius G, Isbel L, *et al* (2021) ZFP462 targets heterochromatin to transposon-derived enhancers restricting transcription factor binding and expression of lineage-specifying genes. *bioRxiv*: 2021.06.28.449463

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P & Smith A (2008) The ground state of embryonic stem cell self-renewal. *Nature 2008 453:7194* 453: 519–523

Young RA (2011) Control of the Embryonic Stem Cell State. *Cell* 144: 940–954

Yun M, Wu J, Workman JL & Li B (2011) Readers of histone modifications. *Cell Research 2011 21:4* 21: 564–578

Zentner GE & Henikoff S (2013) Regulation of nucleosome dynamics by histone modifications. *Nature Structural & Molecular Biology 2013 20:3* 20: 259–266

Zhang Y, Cooke M, Panjwani S, Cao K, Krauth B, Ho PY, Medrzycki M, Berhe DT, Pan C, McDevitt TC, *et al* (2012) Histone H1 depletion impairs embryonic stem cell differentiation. *PLoS Genetics* 8

Zlatanova J & Doenecke D (1994) Historie H1°: a major player in cell differentiation? *The FASEB Journal* 8: 1260–1268

APPENDICES

Appendix 1 – sequences of primers used in this study

panH2B F	CCATGGGCATCATGAACTC
panH2B R	GGTCGAGCGCTTGTTGTAAT
panH3 F	AGGACTTCAAGACCGACCTG
panH3 R	GTGTCCTCAAACAGACCCAC
Nanog.YP_F	CAAGCGGTGGCAGAAAAACC
Nanog.YP_R	ATGCGTTCACCAGATAGCCC
H1f6 F	GCTGATTCCTGAGGCCCTTT
H1f6 R	TTCTCCACGTCGTAACCAGC
H2bu2 F	AGCCAGTGCAGCAGGATG
H2bu2 R	CCGTTTGCGTTTCTTGCCAT
H3f4 F	GACACGAACCTGTGCGCTAT
H3f4 R	GAGCCTAAGCCCGTTCTCC
H1f0 F	GATGAGCCCAAAAGGTCGGT
H1f0 R	TTCTTGCTTGGGGCTTTGGA
H1f1 F	TCAGCAACCATGTCGGAGAC
H1f1 R	CTTTTGCCGGCTTCTTGGTC
H1f2 F	CGCGTCTAAAGCCGTAAAGC
H1f2 R	GGGGAGGCAGCCTACTTTTT
H1f3 F	CTGCCACACCCAAAAAGACG
H1f3 R	CCTTGGCTGGACTCTTTGCT
H1f4 F	AAGAGCCCGAAGAAGGCAAA
H1f4 R	TTGGCTGCGGTTTTCTTTGG
H1f5 F	AAGGCGGTGAAGTCTAAGGC
H1f5 R	GAGCCTTTGGGGCTTTGTTG
H1f7 F	ACAACGCTCCCAGCATACAA
H1f7 R	AAAGCAACCGTACCCAGAGG
H1f8 F	GACAAAGAAAGCCTGTGCCC
H1f8 R	CCTTTCTCCATCGTGGCCTT
H1f9 F	TCAAGCGTGTGCTCCAGAAT
H1f9 R	ATCTCTTGGCCCTGTGGTTG
H2bc1 F	GTGGCGGTAAAGGGTGCTAC
H2bc1 R	CTCTCCTTGCGGCATCTCTT
H2bc21 F	AAGGCCGTCACCAAGTACAC

H2bc21 R	TCTGATTTGAGCGGGTGACTC
H2bc18 F	GTTGCCCGGTCTACCTTACG
H2bc18 R	CTGCACTTTCGTGACAGCTT
H3c14 F	TTTCAAGTCGCTGTCTCCGC
H3c14 R	ACGGAAGACACCCAAACGAG
H3c8 F	CAAGCGTGTCACCATCATGC
H3c8 R	AGCCTTTAAAGTGGTTGTGTGG
H3c2 F	TTTGAGGACACCAACCTGTGC
H3c2 R	CCAGTTTGCACTTTGTTTCGTT
EcoRV-Ef1a F	ATCTGGATATCGGCTCCGGTGCCCGTCA
MluI-BamHI-Ef1a R	AGATACGCGTGGATCCTCACGACACCTGAAATGGA
Ef1a_junction F	TTATCGTTTCAGACCCACCTCC
Ef1a_junction R	TTATCGAAAGCAGCGAGACAGG
BstXI_blastR F	ATCTGGCCACAACCATGGCCAAGCCTTTGTCTCA
SalI_blastR R	TGCATGTCGACTTAGCCCTCCCACACATAACC
Blast_junction F	GAGAACAGGGGCATCTTGAG
Blast_junction R	GCGTCAGCAAACACAGTGCA
sg3-H1f6 F	CACCGGAAACTGAGAAACCCCG
sg3-H1f6 R	AAACCGGGGTTTCTCAGTTTCC
sg4-H1f6 F	CACCGCTGCGGCTGGTTACGACG
sg4-H1f6 R	AAACCGTCGTAACCAGCCGCAGC
H1f6KO F	CGGCCTCAAGTACCCTTGT
H1f6KO R	CTTGGTCTGCACCAGGACTC
sg3-H2bu2 F	CACCGGGCGGGAGTGGAGCGGGA
sg3-H2bu2 R	AAACTCCCGCTCCACTCCCGCCC
H2bu2KO F	AGGAAAGAAGAAGCGGGGTA
sg3-H3f4 F	CACCGCAAGCAGCTAGCCACGAAGG
sg3-H3f4 R	AAACCCTTCGTGGCTAGCTGCTTGC
sg4-H3f4 F	CACCGACGGCACGGAAGTCGACGGG
sg4-H3f4 R	AAACCCCGTCGACTTCCGTGCCGTC
H3f4KO F	GGGCAGCCAATAGGAATGTA
H3f4KO R	GTAGCGGTGCGGCTTCTT
Otx2 F	ACTTGCCAGAATCCAGGGTG
Otx2 R	CTTCTTCTTGGCAGGCCTCA

Cd24a F	TTCGCATGGTCACACACTGA
Cd24a R	ACACACACAGTAGCTTCGGG
3xFLAG_H1f6 R	CAATATCATGATCTTTATAATCACC GTCATGGTCTTTGTAGTCCTTCCTC CCTGCTGCCT
XbaI_3xFLAG_COMMON R	CTTATCTAGATCACTTGTCATCGTC ATCCTTGTAATCAATATCATGATCT TTATAATCAC
3xFLAG_H1f6 F	ATGTCGGAAACGGCTCCT
3xFLAG_H2bu2 F	GCCACCATGCCGGAGC
3xFLAG_H2bu2 R	TATCATGATCTTTATAATCACCGTC ATGGTCTTTGTAGTCCTTGGAGCTG GTGTATTTGG
3xFLAG_H3f4 F	ATGGCACGCACCAAGCAG
3xFLAG_H3f4 R	TCAATATCATGATCTTTATAATCAC CGTCATGGTCTTTGTAGTCAGCCCG TTCTCCGCGG
Sox17 F	CTTGGAAGGCGTTGACCTTG
Sox17 R	ACTTGTAGTTGGGGTGGTCC
Eomes F	TCGTGGAAGTGACAGAGGAC
Eomes R	AGCTGGGTGATATCCGTGTT
Gata6 F	TTCTACACAAGCGACCACCT
Gata6 R	TTGAGGTCACTGTTCTCGGG
H2bu2 80 F	CCGGAGCCATCACCAAGGCTCAGA ACTCGAGTTCTGAGCCTTGGTGAT GGCTTTTTTG
H2bu2.80 R	AATTCAAAAAAGCCATCACCAAGG CTCAGAACTCGAGTTCTGAGCCTT GGTGATGGCT
H3f4.17 F	CCGGGAATGTAGCTGGGTCTATAA ACTCGAGTTTATAGACCCAGCTAC ATTCTTTTTG
H3f4.17 R	AATTCAAAAAGAATGTAGCTGGGT CTATAAACTCGAGTTTATAGACCC AGCTACATTC
H3f4.530 F	CCGGTCATGAACCCAACGGCTCTT TCTCGAGAAAGAGCCGTTGGGTTC ATGATTTTTG
H3f4.530 R	AATTCAAAAATCATGAACCCAACG GCTCTTTCTCGAGAAAGAGCCGTT GGGTTCATGA

H1f6.674 F	CCGGGGGAGGAAGTGAGTTTCAAAGCTCGAGCTTTGAAACTCACTTCCTCCCTTTTTG
H1f6.674 R	AATTCAAAAAGGGAGGAAGTGAGTTTCAAAGCTCGAGCTTTGAAACTCACTTCCTCCC