

**POLYCOMB REPRESSIVE COMPLEX 2
DEREGULATION IN TISSUE
HOMEOSTASIS AND CANCER**

María Isabel Espejo Díaz

Doctoral Thesis 2021

**Department of experimental and health sciences
Universitat Pompeu Fabra**

Under the supervision of Dr. Sergio Aranda and

Dr. Luciano Di Croce

Department of Gene Regulation, Stem Cells and Cancer

Centre for Genomic Regulation (CRG)



Si alguna vez te sientes que no estas a la altura, que no estas preparada,
entonces estas en el sitio perfecto para aprender.

P.

Abstract

Polycomb group (PcG) proteins are transcriptional repressors that control development and cell fate. Polycomb Repressive Complex 2 (PRC2) dysfunction has been recurrently associated with embryonic lethality and disease, especially cancer. Still, how deregulation of PRC2 composition or its interplay with chromatin contribute to the development of different disorders is far from being understood. Here, we have characterized the function of PHF19, a subunit of PRC2, during haematopoiesis *in vivo*, and we have explored whether PRC2 impaired chromatin interaction might offer therapeutic opportunities for Diffuse Intrinsic Pontine Gliomas (DIPG) treatment. Using a comparative transcriptomic and epigenomic analysis, we have found that PHF19 is an essential regulator of hematopoietic stem cells (HSCs) homeostasis, such that its ablation destabilizes the balance between preservation of cell fate identity and chromatin dynamics, leading to an increase of hematopoietic malignancies. In addition, using proteomic approaches, we have identified chromatin factors that represent potential therapeutic targets for DIPG. Taken together, our findings contribute to unveil the functional complexity of PRC2 deregulation in pathogenesis.

Resumen

Las proteínas del grupo Polycomb son represoras transcripcionales que controlan el desarrollo y la identidad celular. La desregulación de la función del complejo represivo Polycomb 2 (PRC2 por sus siglas en inglés) ha sido asociada a mortalidad embrionaria y al desarrollo de múltiples enfermedades, especialmente cáncer. Sin embargo, aún estamos lejos de comprender con exactitud como la alteración de la composición proteica de PRC2 o de la interacción del complejo con la cromatina contribuyen al desarrollo de diferentes trastornos. En esta tesis hemos caracterizado la función de PHF19, una subunidad de PRC2, durante el proceso de hematopoyesis *in vivo*, y hemos investigado si la alteración de la interacción de PRC2 con la cromatina puede dar lugar a oportunidades terapéuticas para el tratamiento del glioma pontino intrínseco difuso (DIPG por sus siglas en inglés). Utilizando análisis comparativos transcriptómicos y epigenéticos, hemos encontrado que PHF19 es una proteína reguladora esencial de la homeostasis de las células madre hematopoyéticas (HSC por sus siglas en inglés), de modo que su ablación desestabiliza el balance entre el mantenimiento de la identidad celular y la dinámica de la cromatina, lo que conlleva a un incremento de las patologías hematopoyéticas. Así mismo, utilizando un enfoque proteómico, hemos identificado factores de cromatina que podrían suponer potenciales dianas terapéuticas para DIPG. En conjunto, nuestros resultados contribuyen a un mejor conocimiento del complejo papel que desempeña la alteración de PRC2 en el desarrollo de enfermedades.

Preface

As the evidence about the relevance of chromatin function in disease mounts, the efforts to better characterize and understand its regulatory mechanisms have increased tandem. In this search, PcG emerged early as an essential mechanism to preserve proper chromatin function. PRC2 involvement in pathogenesis was revealed with the first PRC2 KO mouse, which highlighted its importance during embryonic development and in adult haematopoiesis. Later, the first genome-wide sequencing studies unveiled recurrent PRC2 mutations in human hematopoietic malignancies, which reinforced the central function of PRC2 in diseases, especially in cancer. More recently, the identification of the first oncohistone, H3K27M, has brought another landmark, highlighting also the role of the chromatin misregulation in promoting PRC2 oncogenic function. This thesis aimed contribute to these major efforts.

To this end, in the first part of my thesis I studied the effect of PHF19 depletion *in vivo*, which illustrated how an alteration in PRC2 composition could has an impact in adult haematopoiesis that ultimately could lead to diseases. Meanwhile, during the second part of my thesis, I aimed to identify chromatin proteins sustaining tumorigenesis in DIPG cells, which could contribute to discover of new therapeutic targets and understanding of how H3K27M promotes PRC2 oncogenic function.

Table of contents

Abstract	vii
Resumen	viii
Preface	ix
Table of contents	xi
INTRODUCTION	3
1. CHROMATIN REGULATION AND FUNCTION.....	3
1.1. <i>The basic structure of chromatin</i>	3
1.2. <i>The changing chromatinome as a mechanism to regulate chromatin function</i>	4
1.1.1. Nucleosomes as a physical barrier to DNA.....	5
1.1.2. Nucleosomes as an additional layer of information.....	6
1.1.3. DNA covalent modifications.....	9
1.1.4. Chromatin 3D structure and architectural proteins.....	9
1.3. The different methodologies for characterization of chromatin structure and chromatinome.....	10
1.3.1. The chromatin structure profiling.....	10
b) Histones PTMs and chromatinome occupancy.....	12
c) DNA modifications.....	14
d) High-order chromatin organization.....	14
1.3.2. The chromatinome characterization.....	15
2. Polycomb Group (PcG) of proteins in chromatin regulation and function.....	18
2.1. Polycomb complex as a model of epigenetic regulation.	18
2.2. Polycomb recruitment at chromatin.....	20
2.3. PRC2 composition in mediating PRC2 function.....	22
2.3.1. Core PRC2.....	22
2.3.2. PRC2.1.....	26
2.3.3. PRC2.2.....	29
2.4. Chromatin components in mediating PRC2 function.....	30
2.4.1. DNA.....	30
2.4.2. RNA.....	31
2.4.3. Histones PTMs.....	33
2.4.4. Other chromatinome proteins.....	34
3. PRC2 FUNCTION IN DEVELOPMENT AND AS A DRIVER OF DISEASE.....	35
3.1. Alterations in PRC2 composition affecting PRC2 function.....	36
3.1.1. Homozygous PRC2 KO mice.....	36
3.1.2. Heterozygous, or conditional PRC2 KO mice.....	37
3.1.3. Loss-of-function and gain-of-function PRC2 mutations in human disease.....	41
3.2. Alterations in the chromatin environment affecting PRC2 function.....	43
3.2.1. H3K27M in mediating PRC2 function in DIPG.....	44
AIM	49
RESULTS	53

CHAPTER I	55
1. Previous generation and phenotypical hematopoietic characterization of a PHF19 KO mice.....	56
2. Phf19 KO HSCs display an increased HSC identity and a highly quiescent transcriptome.....	62
3. Phf19-KO HSCs present a chromatin accessibility unfavourable to undergo lineage differentiation.....	72
4. Phf19 KO HSC quit quiescence upon stress but display impaired lineage differentiation and a bias towards HSC self-renewal.....	79
5. HPC-7 cell line is not a good alternative to study PHF19 mechanism in LT-HSCs.....	83
CHAPTER II	87
1. The relevance of chromatin and epigenetics in driving Diffuse Intrinsic Pontine Gliomas (DIPG).....	88
2. Identification of epigenetic vulnerabilities in DIPG primary cell lines for selective targeting of tumoral cells.....	91
3. CRISPR screen of 652 chromatin regulator factors to identify potential candidates that compromise DIPG cell proliferation and/or survival.....	93
4. Successful generation of the first proteome and chromatome data from DIPGs cell lines.....	99
5. The chromatin-bound proteins enriched in DIPGs represent potential chromatin interactors driving oncogenic outcomes.....	108
6. Validation of our list of potential therapeutic candidates proteins for DIPG treatment.....	112
Discussion.....	119
Conclusions	143
Materials and methods	147
INDEX.....	178
References	182
Annex.....	215

INTRODUCTION

1. Chromatin regulation and function

Humans harbour about trillions of cells that, despite containing the same genome, exert multiple functions due to differences in the regulation of their genomic information [1]. The regulatory mechanisms that govern these genomic processes, as transcription, DNA repair or replication, as well as the faithful transmission of genetic and epigenetic information, rely on chromatin [2].

1.1. The basic structure of chromatin

Chromatin is the macromolecular complex formed by DNA together with physically associated proteins and non-coding RNAs [3, 4]. The functional unit of the eukaryotic chromatin is the nucleosomes, a stable octameric core of two each of the histones H2A, H2B, H3 and H4, wrapped by up ~ 147 bp of DNA [5]. Histones are the most abundant eukaryotic nuclear proteins, with ~ 8500 - 17000 copies per Mb of the genome, which allows accomplishing the impressive feat of packaging about 2 meters of DNA into a 10-micrometre human cell nucleus [6, 7]. More significantly, DNA compaction into the nucleus is not stochastic; rather, nucleosome distribution promotes the functional chromatin organisation into the densely arranged constitutive and facultative heterochromatin or the low nucleosome-occupied euchromatin (Figure I1) [8, 9]. The stable constitutive heterochromatin generally associates with the preservation of genome integrity and accurate chromosome segregation. In contrast, the dynamic interchange of nucleosomes endorses large fluctuations between facultative heterochromatin and euchromatin. This

nucleosome dynamic regulates the accessibility of chromatin-bound proteins, also known as ‘chromatome’, to the DNA replication start sites, the double-strand breaks, or the protein-coding and non-coding regulatory regions [10-12].

1.2. The changing chromatome as a mechanism to regulate chromatin function.

In eukaryotic cells, 98% of the genome does not encode for proteins, but it contains non-coding regions with functions in chromatin regulation [13]. Functional annotations identified different categories of non-coding regulatory regions, which largely concern promoters, enhancers, and insulators. Promoters regions associate with transcription initiation and are close to transcription start sites (TSS), whereas enhancers and silencers regions associate with amplification or reduction of gene expression and can be located up to 1 Mb distance from TSS [14].

Regulatory regions are particularly functionally relevant as they harbour 90% of the total DNA sequences recognised by **Transcription Factor (TF) proteins** [12, 15]. In humans, over 1600 TFs have been estimated [16]. TFs can bind to specific DNA sequences and guide the recruitment of co-factors proteins to the chromatin. The term TF refers to their main function in regulating gene expression via the interaction with ‘coactivators’ and ‘corepressors’ proteins. For instance, close to TSS, promoters harbour specific DNA sequences that guide the RNA polymerase II (RNA pol II) recruitment based on the cooperation between TFs and

the coactivator Mediator complex [17, 18]. In addition to their main function over transcription [19], TF-cofactor interactions can also contribute to the three-dimensional chromatin structure organisation, to guide the proteasome degradation of other chromatin-bound protein [20], or to decrease the head-to-head collisions between RNA and DNA polymerases upon DNA replication [21]. TF binding to their DNA-binding sequences is mainly modulated by nucleosomes and DNA covalent modifications. TF binds almost exclusively at euchromatin, with nucleosomes representing a physical barrier to DNA. However, rarely pioneer TF may directly bind the nucleosome surface and induce sequence-specific chromatin unfolding by recruiting ATP-dependent chromatin remodeling complexes (CRCs) and histone-modifying proteins [8, 22, 23].

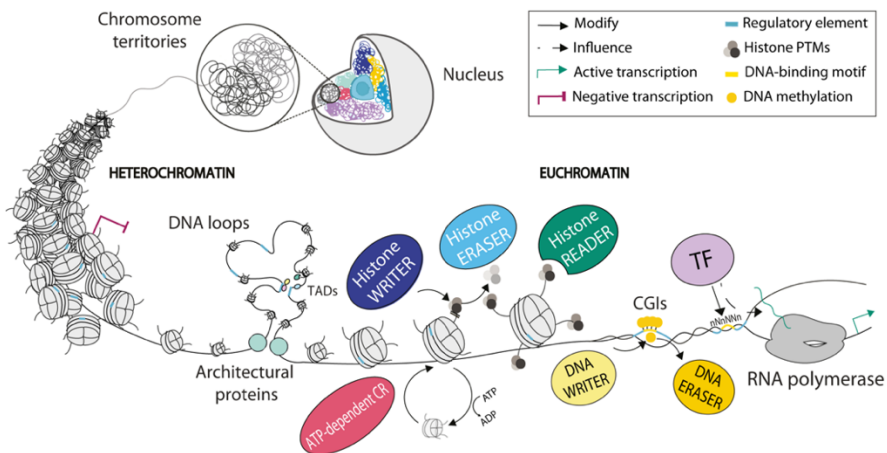


Figure 11. Schematic representation (from the largest scale, chromosome territories, to the linear DNA) of chromatin structure and function regulation through chromatin-bound proteins interactions. CR = chromatin remodeling; TF = transcription factor; TADs = topological associated domains.

1.1.1. Nucleosomes as a physical barrier to DNA.

Introduction

Four major ATP-dependent CRCs regulate nucleosome density and spacing by breaking the histone-DNA interactions at the expense of ATP [23, 24]. The chromodomain helicase DNA-binding (CHD) and the switch/sucrose non-fermentable (SWI/SNF) subfamilies can move and eject histones to render the chromatin more accessible to TF. On the other hand, the CHD together with the imitation switch (ISWI) subfamily and histone chaperones, can also assemble, mature, deliver and space canonical nucleosomes at nascent DNA to maintain chromatin state upon replication [23, 24]. The last major complex, the inositol-requiring 80 (INO80) subfamily, can alter the nucleosomes composition through the replacement of canonical histones by non-canonical or replication-independent variants, which differ in their functionality presenting different cell-type specificity, deposition patterns, and post-translational modifications (PTM) [25, 26]. Thus, for instance, in the presence of DNA damage the INO80 complex replaces the H2A.Z variant with H2A.X, which specific phosphorylation at Ser139 (γ -H2A.X) serves to recruit the DNA repair machinery [27, 28].

1.1.2. Nucleosomes as an additional layer of information.

More than 100 histone 'writers' and 50 'eraser' human proteins regulate the chromatome interactions by introducing more than 400 different histone post-translational modifications (PTMs) [29-31]. The best functionally described modifications are the methylation and acetylation of the H3 lysine residues, albeit phosphorylation and ubiquitination of other histones are also well studied (Figure I2A) [32-34]. These chemical modifications can impact chromatin

accessibility, as histone acetylation, which destabilise DNA-histone interactions promoting euchromatin [33]. However, even more importantly, histones PTMs serve as a docking-site for chromatin-bound proteins, or histone 'readers', thus representing a regulatory layer for chromatin function [35]. The abundance and combination of different histone PTMs regulate gene transcription [36]. For instance, H3K27ac together with H3K4me3 or H3K4me1 associate with accessible or active promoters or enhancers, while at the gene bodies H3K36me3 and H3K79me3 correlates with actively transcribing genes (Figure I2B). H3K27me3 associates both with inaccessible or silenced promoters and enhancers, and H3K9me3 is essential to establish constitutive silenced heterochromatin [37, 38]. These histone PTMs may represent both, driving and maintaining mechanisms of gene expression, as exemplified by H3K4me3, which recognition drive the recruitment of the polymerase transcriptional initiation complex (PIC) at some genes and, simultaneously, maintain active gene expression by inhibition of the de novo DNA methyltransferase DNMT3A activity [38, 39].

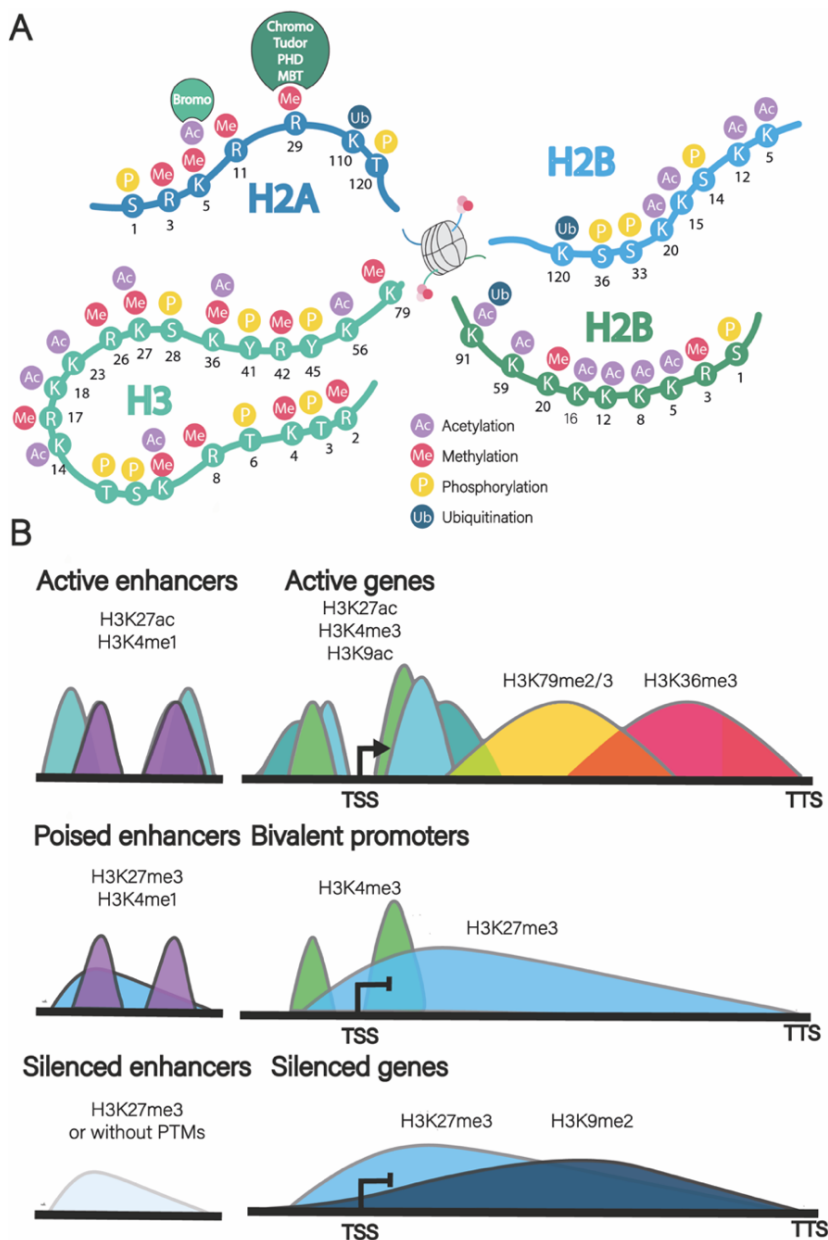


Figure 12. (A) Core histones residues subject to acetylation, methylation, phosphorylation and ubiquitination. These histone PTMs can be differentially recognized by histone readers harboring specific protein domains, such as bromo for acetylation, or chromo domains for methylation. (B) Simplified illustration of the different H3 methylation and acetylation PTMs at promoters and enhancers according to their transcriptional state. Histone PTMs distribution is depicted as defined by chromatin immunoprecipitation and sequencing (ChIP-seq) assays (adapted from [38]).

1.1.3. DNA covalent modifications.

At the linear level, 3 DNA methyltransferases regulate the DNA methylation at the 5-carbon of the cytosine pyrimidine ring (**5mC**) (Figure I1). The 5mC disrupts >60% of the TFs interactions and mainly leads to gene silencing [40]. Thus, differences in the DNA methylation profile correlate with the transcriptional outcome and contribute to set cellular identity. Consistently, while the majority (>80%) of CpGs in the genome harbour 5mC, promoters CpGs contains less than 10% of 5mC and its specific distribution profile can discriminate between different tissues [41]. 3 ten-eleven translocation (TET) methylcytosine dioxygenase mediated DNA demethylation through oxidation of 5mC to 5hmC, 5caC, and 5fmC, followed by replication-dependent dilution of oxidized 5mC or thymine DNA glycosylase (TDG)-mediated excision of 5fC and 5caC [42]. Importantly, 5mC and its oxidative derivatives, also serve as an anchor to DNA-methylation's 'readers', and cooperate with histone PTMs to regulate chromatin function [41, 43]. For instance, the methyl-CpG binding protein 2 (MeCP2) recognise 5mC and either H3K36me3 or H3K4me3 to discriminate between gene bodies and promoters. Thus, at gene bodies, MeCP2 suppress intragenic promoters and regulate alternative splicing by RNA pol II eviction [44], while at promoters, it induces gene repression by histone deacetylases HDAC1-2 recruitment [45-47].

1.1.4. Chromatin 3D structure and architectural proteins.

Finally, architectural proteins contribute to the hierarchical 3D organization of the genome within the nucleus [48]. Chromosomes

localize into discrete nuclear regions, termed chromosome territories, which are partitioned at the large megabase scale into A/B compartments [49]. A compartment clusters euchromatin regions towards the interior nuclear space, whereas B compartment tends to segregate tightly compacted heterochromatin regions towards the nuclear periphery, where it can attach to the nuclear lamina to form lamin-associated domains (LADs) [50, 51]. Inside A/B compartments, chromatin folds into Topological Associated Domains (TADs), comprising ten to hundred kilobases regions within which the chromatin interactions are more frequent than with regions outside the domain [52, 53]. Finally, intra-TADs DNA loops arise from cohesin-mediated interactions between paired CCCTC-binding factor (CTCF) sites, in order to facilitate contacts between regulatory regions located at 10-100 kilobase of linear DNA distance [48, 54].

Overall, the regulation of the genomic information involves a precise chromatin structure and a notable heterogenic, networked, and dynamic chromatome composition, which characterisation is essential to understand cell identity and disease.

1.3. The different methodologies for characterization of chromatin structure and chromatome.

1.3.1. The chromatin structure profiling

Once the human genome sequencing was achieved, the widespread and fast emergence of high-throughput sequencing-based technologies has enabled the characterization of chromatin structure

at the genome-wide level, including nucleosomes positioning, DNA or histones modifications, and high-order chromatin structure [55]. Co-integration of these and other data, as transcriptome profiling by RNA sequencing technique, have set an unprecedented scenario for the understanding of the molecular mechanisms governing chromatin function [56].

a) **Chromatin accessibility**

Chromatin accessibility assays are used to unravel DNA regions occupied or depleted of nucleosomes [8, 57]. Nuclease assays, such as the micrococcal nuclease sequencing (MNase-seq) or DNase I hypersensitive sites (DNase-seq) [58], relies on the sequencing of the remaining DNA fragments after extensive nuclease-mediated digestion, which correlates with nucleosome occupied regions (Figure I3A). In contrast, direct assays relay on the sequencing of the accessible chromatin regions isolated either by chromatin sonication, as Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE-seq) assay [59] or the Assay for Transposase-Accessible Chromatin (ATAC-Seq) (Figure I3A). The last, ATAC-seq uses the hyperactive transposase Tn5, which cleavages and ligates sequencing adapters in a single step known as ‘tagmentation’ [60]. ATAC-seq benefits over previous methods include the lack of bias for particular sequence and the tagmentation step, which simplifies PCR library amplification and increases preparation efficiency. Optimized protocols, such as Omni-ATAC-seq, allows to remove mitochondrial DNA and provides one of the most sensitive and specific chromatin accessibility data [61].

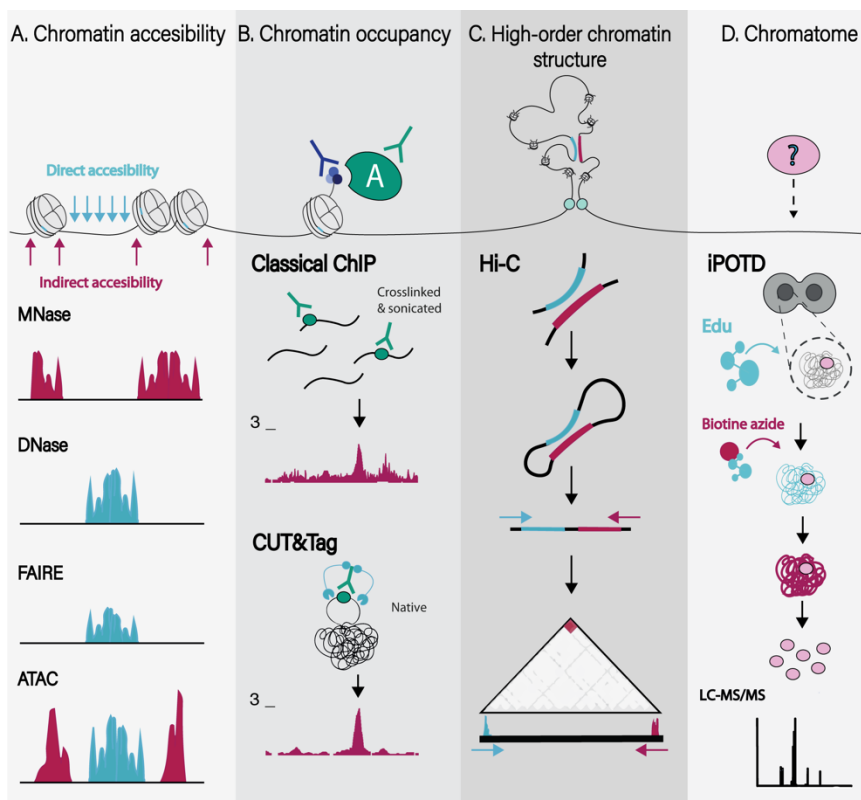


Figure I3. Principal techniques to study chromatin structure and composition. (A) Principal techniques to directly (red) or indirectly (blue) assess chromatin accessible regions. (B) Principal techniques to assess genome-wide distribution of known chromatin features. (C) Hi-C gold-standard technique to assess chromatin long-range interactions and high-order organization. (D) iPOTD protocol to assess total chromatome composition.

b) Histones PTMs and chromatome occupancy

Chromatin immunoprecipitation assay followed by sequencing (ChIP-seq) is used to profile the genomic occupancy of a protein of interest (Figure I3B) [62]. In the standard ChIP protocol, the chromatin-bound proteins are first formaldehyde (FA)-crosslinked to DNA [63]. Next, the cross-linking chromatin is sonicated, and the fragments are incubated with an antibody against the protein of interest, which allows to immunoprecipitate the DNA fragments specifically bound by this protein. Thus, after chromatin de-

crosslinking, these DNA fragments are isolated, used for library preparation, and sustained to massive DNA sequencing [64]. The resulting reads are mapped against a reference genome, which can result in pileups of reads (peaks) that correspond to the genomic locations of the protein of interest [65].

Multiple upgrades of the standard ChIP-seq protocol have been described, the most relevant incorporating i) tethering enzymes to increase sensibility, and ii) native chromatin conditions to avoid chromatin shearing and FA-fixation while preserving natural electrostatic contacts. The Cleavage Under Targets and Release Using Nuclease (CUT&RUN) assay [66], and its improved adaptation the Cleavage Under Targets & Tagmentation (CUT&TAG) [67, 68], take advantage of the standard ChIP antibody-targeting to guide the recruitment of a Protein A, fused either with a MNase (pA-MN) or with a hyperactive Tn5 transposase (pA-Tn5) respectively, to the genomic regions where a protein of interest is located (Figure I3B). Following intact cells permeabilization and antibody-pA-MN/Tn5 incubation, the activation of the tethering enzymes induces the cleavage or tagmentation of DNA in the vicinity of the antibody-bound genomic regions. The DNA fragments are then released from the nucleus to the supernatant, while the rest of the undigested genome is retained within the intact nucleus. By doing so, CUT&RUN and CUT&TAG simplify DNA collection to generate sequencing libraries, are suitable for native or lightly fixed chromatin conditions and, importantly, increase sensibility by reducing the standard ChIP signal-to-noise ratio produced by complete genomic

fragmentation. Altogether, CUT's techniques decrease significantly cell input and deep sequencing requirements, opening up ChIP single-cell analysis for histone marks or highly expressed TF with robust epitopes, such as CTCF [67, 69].

c) DNA modifications

The characterisation of 5mC DNA methylation does not require specific antibodies or chromatin fragments purification. The accurate whole-genome bisulfite sequencing (WGBS) detects 5mC locations by treating genomic DNA with sodium bisulfite, resulting in deamination of unmethylated cytosines to uracil [70]. Methylated cytosines remain intact, and post DNA fragmentation, the PCR-amplification with an uracil-tolerant polymerase converts uracil into thymine. The resultant sequencing comparison with a bisulfite untreated DNA provides a base-pair resolution readout of the genome-wide location of 5mC. However, the standard bisulfite conversion cannot distinguish between 5mC and 5hmC, so specialised protocols and independent library preparations are required to detect each mark [71]. The emerging long-read sequencing techniques as Oxford Nanopore may provide a reliable solution and a wider knowledge about other DNA modifications, as they allow to sequence native DNA and to discriminate between different modifications, as 6mA and 5hmC [72].

d) High-order chromatin organization

The techniques based on chromosome conformation capture (3C) assay are the top methods to obtain information about the three-dimensional chromatin structures, such as chromosome territories,

TAD domains, and DNA loops [73]. All the 3C-based techniques rely on the principle of proximity ligation, in which FA-fixation cause interacting loci to be bound with each other via covalent DNA-protein cross-links. Following DNA fragmentation by combining restriction enzymes, the treatment with DNA ligases favours the formation of chimeric products combining the interacting genomic sequences [74]. Thus, the short and long-range chromatin interactions can be inferred at a specific region using 4C-technology [75], among several sites around a specific region using 5C [76], or genome-wide using high-throughput sequencing Hi-C [77] (Figure I3C). The interactions are measured within fixed-sized bins, and the number of interactions between bins is stored in contact matrixes. The standard Hi-C protocol cover bins ranking from >10Mb of linear distance for long-range interaction to 1kb resolution for short-range interaction, although a recently improved Hi-C protocol termed Micro-C, which use a MNase digestion of DNA, display a resolution of 1-100 nucleosome scale and identify 3 to 5 times more looping interactions within 1kb resolution than Hi-C [78].

Overall, the continuing efforts to increase techniques sensibility and the improvement of the NGS approaches are providing unprecedented opportunities as the emergence of single-cell techniques (scATAC-seq, scChIP-seq, scDNA-methylation, and scHi-C [79-82]), which is opening the door to characterize the cell-to-cell variation within tissues or rare cellular populations.

1.3.2. The chromatome characterization

Introduction

In addition to high-throughput genomics techniques, much of our current knowledge regarding chromatin function is based on the mass spectrometry (MS) characterisation of the chromatome, which provided a picture of the chromatin scenario in which these functions are taken place and the main actors involved [2].

In particular, MS approaches can directly provide information about the chromatome composition of a specific locus or the global chromatin. Locus specific chromatome isolation strategies represent a powerful tool, although they still require major improvements to rich the tough task of isolating, enriching and analysing the chromatome of a specific locus among thousands of loci (for a nice review in locus-specific isolation see [83]). For instance, the recent dCas9–APEX2 biotinylation at genomic elements by restricted spatial tagging (C-BERST) assay combine the targeting of a specific locus using sgRNAs with the biotinylation of any proteins around the target locus using the ascorbate peroxidase APEX system, which allows to posterior identification of biotinylated proteins using MS [84]. However, although dCas9:APEX2 generated a great reception, it soon turns out to significantly biotinylated unspecific proteins, so multiple controls and replicates are required to mitigate these issues.

On the other side, bulk chromatin proteomic methods have proved to significantly provide a picture of the total chromatin composition, as well as of large chromatome subtypes such as heterochromatin or DNA replication forks [2, 85]. The first methods rely on the physicochemical properties of the chromatin, and combine different ionic buffers and centrifugation steps to finally purify an insoluble

material highly enriched in chromatin proteins [86]. Subsequent techniques introduced increasing gradients of digestion enzymes to exploit the differences in DNA-protein strength interactions, such as the differential chromatin-associated proteins (D-CAP) [87]. Similarly, the chromatin enriching salt preparations (CHESS) introduced increasing gradients of salt content buffers to exploit the differences in DNA solubility to discriminate between euchromatin and heterochromatin bound proteins [88]. However, the trade-off necessary to maintain solubility by using chelating compounds, salts, or detergents, inevitably leads to the dissociation of important interactors in all these techniques. Consequently, the use of crosslinked nucleus was incorporated by the Chromatin Enrichment for Proteomics (ChEP) protocol, which also allow to apply extremely stringent washing conditions to remove proteins contaminants while preserving specific interactions [89]. Recently, two protocols, the isolation of proteins on nascent DNA (iPOND) [90], and the DNA-mediated chromatin pulldown (Dm-ChP) [91] adapted the ‘click chemistry’ to combine the incorporation of a thymidine analogue 5-ethynyl-2'-deoxyuridine (EdU) molecule during DNA replication with biotin [91]. These strategies allow to efficiently purify EdU-labelled nascent chromatin using streptavidin–biotin affinity and to obtain with high specificity *bona fide* chromatin-bound proteins. Long incubations with EdU can result in genome-wide incorporation, thus enabling the identification of proteins on Total DNA (iPOTD) [91].

2. Polycomb Group (PcG) of proteins in chromatin regulation and function.

2.1. Polycomb complex as a model of epigenetic regulation.

The mechanisms regulating chromatin are extremely dynamic. Thus, transcriptional programs would be rapidly lost in the absence of self-reinforcing loops that maintain cell fate and transcriptional memory through cellular generations [92, 93]. After a long debate regarding the term ‘epigenetics’, a recent definition has been proposed as ‘the study of molecules and mechanisms that can perpetuate alternative gene activity states in the context of the same DNA sequence’ [93]. This operational definition requires to consider a molecule or mechanism as ‘epigenetics’ i) to regulate gene transcription, ii) to persist for long periods in post-mitotic adult cells and, importantly, iii) to be mitotic and transgenerational inherited [4]. In this regard, the most-well characterized carriers of epigenetic modification are the histone PTMs H3K9me3 and H3K27me3, the DNA modification 5mC, and some non-coding RNAs.

The PcG complex is a histone-modifying multiprotein complex whose chromatin recruitment and enzymatic activity is an essential epigenetic mechanism to safeguard facultative heterochromatin [94]. Proteomic and functional studies classified PcG into two main complexes, the Polycomb Repressive Complexes 1 (PRC1) and 2 (PRC2). PRC1 and PRC2 harbour the catalytic capacity to induce H2AK119ub and **H3K27me1/2/3**, respectively, which strongly

associated with transcriptional repression. Early studies identified PcG as a group of proteins necessary to maintain correct spatiotemporal expression of homeotic (HOX) genes during *Drosophila* development [95]. In mammals, PcG is also required to define the specific developmental and cell fate transcriptional programs [94] (Figure I4). Therefore, it is involved in a plethora of particular cellular processes, including regulation of X chromosome inactivation, genomic imprinting, cell cycle, trans-generational epigenetic inheritance of facultative heterochromatin, and also development of disease [93]. However, how the interplay between PcG complexes and chromatin orchestrates chromatin function is still not fully understood.

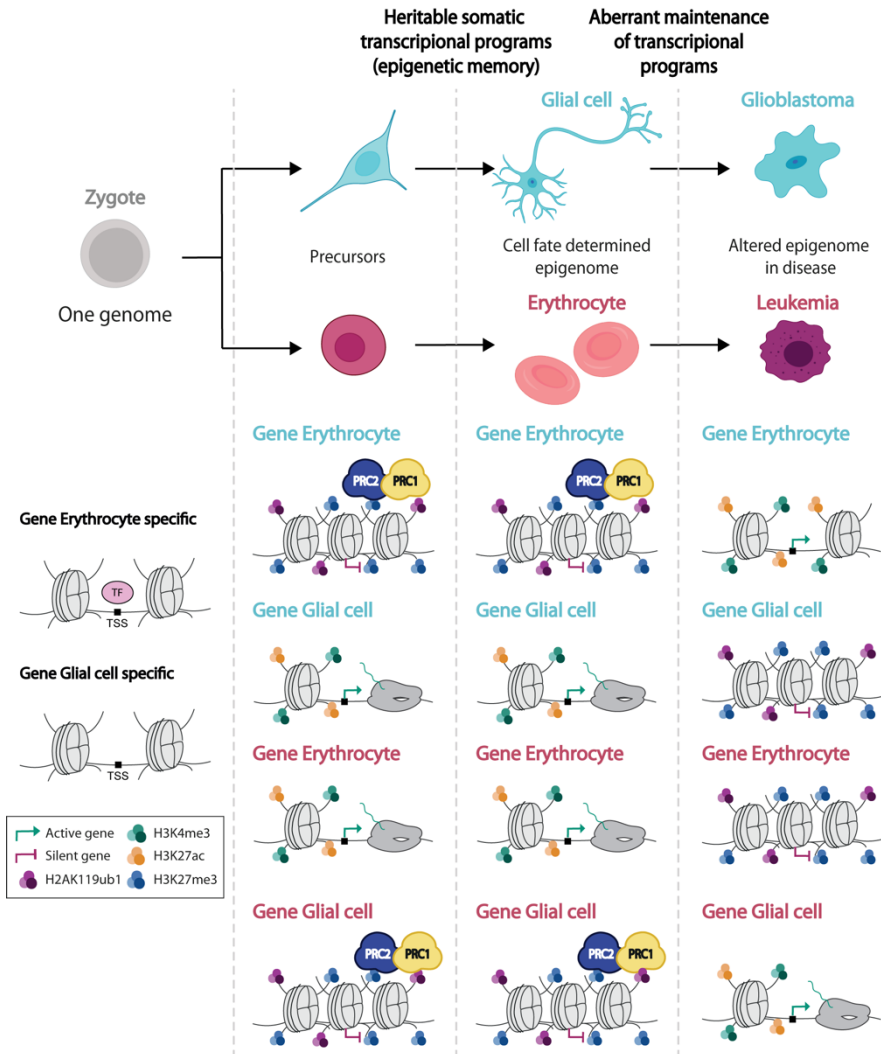


Figure 14. PcG mediated epigenetic maintenance of cell identities during development and throughout life. Starting from the zygotic genome, stage- and cell-type-specific TF initiate regulatory cascades that induce cell differentiation. Epigenetic components as PcG maintain the ‘off’ states of certain genes in a cell-type- and time-specific manner (erythrocyte genes ‘off’ in glial cells (blue) and glial cell genes ‘off’ in erythrocytes (red)). Failures in this mechanism promote altered epigenomes that can lead to changes in specific cell fate or, when accidental, to disease (adapted from [68]).

2.2. Polycomb recruitment at chromatin.

In mammals, the classical PcG recruitment model suggested a hierarchical binding mechanism in which PRC2 might interact with

chromatin via specific DNA-binding factors, deposit H3K27me₃, and promote PRC1 recruitment to reinforce PcG gene silencing [94]. In this line, genome-wide mapping studies in mouse ESCs (mESC) and human ESCs (hESC) identified hypomethylated CpG islands (CGIs) in more than 90% of the promoters and the enhancers of PRC2 repressed genes, which leads to the hypothesis of CGIs DNA-sequences as head effectors of PRC2 target specificity (Figure I5) [94, 96]. However, several data quickly challenged this model, as i) PRC1 have been found to not necessarily correlate with PRC2 on chromatin, ii) H3K27me₃ and H2AK119ub can be localized outside CGIs, iii) a large fraction of unmethylated CGIs remains PcG free, and iv) PRC1 can deposit H2AK119ub and trigger PRC2 recruitment at H3K27me₃ depleted chromatin regions [97]. Hence, although the PRC2 classical model could occasionally explain PRC2 function, additional mechanisms should regulate its chromatin activity.

Interdependency

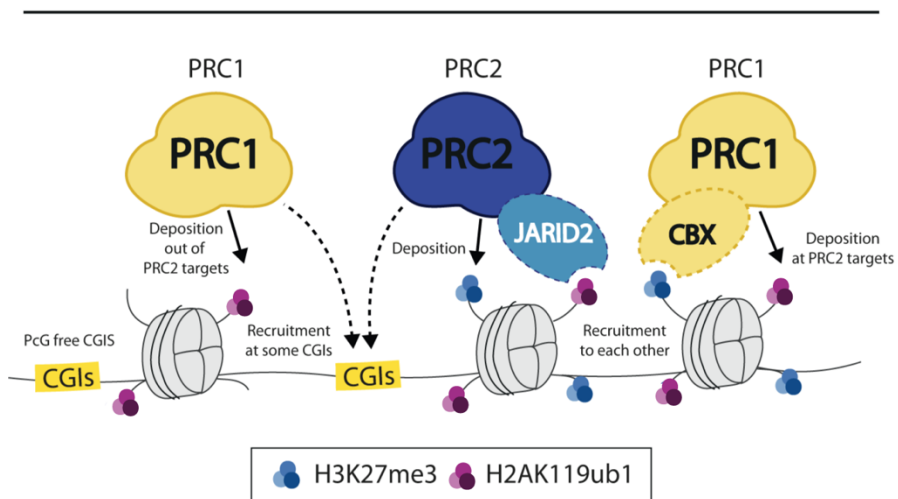


Figure I5. Schematic representation of PcG recruitment model. PRC2 and PRC1 complexes can be recruited independently to chromatin at some unmethylated CGIs and deposits H3K27me₃ and H2AK119Ub respectively. In turn, these marks

serve to recruit each other, as PRC2 complex (via JARID2 subunit) is recruited at H2AK119Ub sites and PRC1 complex (via CBX subunit) is recruited at H3K27me3 sites (adapted from [98]).

Indeed, besides the slight preference for CGIs, PRC2 has currently demonstrated relatively poor DNA sequence specificity and interacts with most of the genomic chromatin through weak and transient interactions [99]. In this line, PRC2 only stably co-localise at H3K27me3 regions, which suggest further mechanisms mediating recruitment, stabilisation, and catalytic activation of PRC2 at those regions. As a result of these shreds of evidence, the more likely model for PRC2 regulation relies on a complex but tightly regulated interaction between many different factors, including incorporation of its different subunits, allosteric activators, interactions with various histone modifications or chromatin structures, and PRC2-interacting partners, including DNA and RNA [94, 100, 101].

2.3. PRC2 composition in mediating PRC2 function

2.3.1. Core PRC2

The PRC2 minimal core is composed of three proteins: i) one of the SET domain-containing histone methyltransferases enhancer of zeste (**EZH2 or EZH1**); ii) the embryonic ectoderm development (**EED**) protein; and iii) the suppressor of zeste (**SUZ12**). All three proteins associate with equimolar stoichiometries and are indispensable for the enzymatic activity of the complex. Additionally, the PRC2 core can contain one of the retinoblastoma-binding protein (**RBBP4-7**), which present a slightly lower stoichiometry and is dispensable for PRC2 activity [94, 102].

EZH1 and EZH2 harbour the unique **catalytic activity** to mono-, di-, and tri-methylate lysine 27 of histone H3 [94]. Importantly, the catalytic activity and expression of the two mutually exclusive subunits differ (Figure I6A). PRC2-EZH2 activity is higher than PRC2-EZH1 [103, 104], which may be explained by the capacity of EED to bind H3K27me3 to promote a more catalytic active conformation for EZH2, but not for EZH1, also referred as EZH2 allosteric activation [105-108]. The EED-mediated positive loop between EZH2 and H3K27me3 allows for more effective methylation of PRC2 targets, a better spread of H3K27me3 towards neighbouring regions, and the establishment of a H3K27me3-positive loop that ensure **epigenetic inheritance** of parental H3K27me3 domains during cell division [109-112]. These different enzymatic capacities may account for the higher expression of EZH2 during development [113-115], while in adult differentiated cells EZH1 is the predominant subunit (Figure I6A).

Importantly, both EZH1 and EZH2 can catalyse H3K27me_{1/2} in mESCs, but only EZH2 efficiently perform H3K27me₃ [103]. This could also refer to the differences in EED-mediated allosteric activation, as EZH1/2 alone only methylate efficiently H3K27me_{0/1} [116], underscoring the necessity of additional mechanisms to promote H3K27me₃ deposition. The existence of EZH2 stabilization mechanisms to promote H3K27me₃ can also explain the high co-localization of PRC2 with H3K27me₃ but not H3K27me_{1/2}, which suggests a more transient PRC2 chromatin interaction at H3K27me_{1/2} regions (Figure I6B). These differences in PRC2 interaction and the distinct genomic distribution of H3K27me_{1/2/3}

Introduction

highly potential specific biological functions, with H3K27me1 found on 5%–10% of total H3 and enriched within gene bodies of active genes; H3K27me2 covering up to 70% of H3 at non-coding regions, which suggest a role in the prevention of inappropriate promoter or enhancer activation; and, finally, H3K27me3 deposited on 5%–10% of total H3, and correlate with steady silenced promoters and enhancers (Figure I6B).

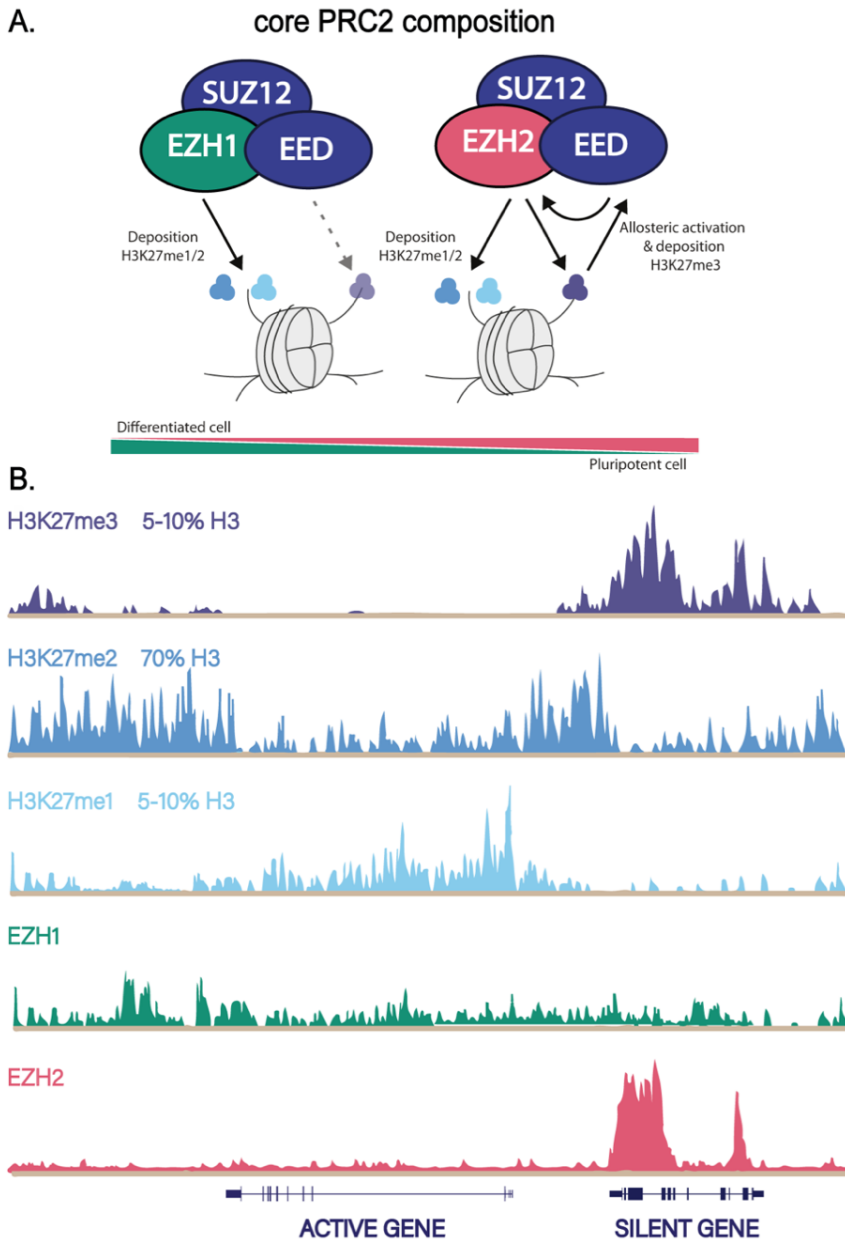


Figure 16. Different core PRC2 composition and function. (A) Schematic representation of core PRC2 subunits and the differences between PRC2-EZH1 and PRC2-EZH2 enzymatic activity and expression. (B) Overview of the genomic distribution of H3K27 methylation and its colocalization with EZH1-2. Representation of ChIP-seq tracks for H3K27 methylation and EZH1/2 illustrate the overlap between H3K27me3 and EZH2 but not EZH1 in an inactive gene, enrichment of H3K27me1 over an active gene and intergenic H3K27me2 (adapted from [102]).

However, despite EZH2 harbour DNA-binding domains that seem to stabilise the PRC2 catalytic activity, EZH2 and EED alone have proved insufficient to mediate PRC2 target specificity and strong chromatin binding [117]. This point towards **non-core PRC2** substoichiometrical **accessory proteins** also referred to as cofactors, and other **chromatin components** as the main effectors of PRC2 recruitment, stability, residence time on chromatin, and overall modulation of PRC2 enzymatic activity [103, 117-123]. SUZ12 mediates the assemble of core PRC2 with these accessory proteins [99]. Importantly, some accessory proteins assemble to PRC2 through the same SUZ12 domains, which results in two different PRC2 subcomplexes comprising mutually exclusive combinations of cofactors, **PRC2.1** and **PRC2.2** [118, 124].

Finally, **EZHIP** cofactor has been identified outside the PRC2 core and subcomplexe, and has been demonstrated to inhibits H3K27 methylation by binding EZH2 and preventing its activity [125-128].

2.3.2. PRC2.1

PRC2.1 contain one of the Polycomb-like proteins (**PCL**) cofactors (PCL1/PHF1, PCL2/MTF2, PCL3/Phf19), and either Elongin BC and Polycomb Repressive Complex 2 Associated (**EPOP**), or PRC2-associated LCOR isoform 1 (**PALI1/2**) proteins.

PCLs proteins can bind to unmethylated CGIs and H3K36me_{2/3} to promote PRC2 recruitment (Figure I7A). Interestingly, ablation of the individual PCL proteins only modestly affected the recruitment of PRC2.1 in mESCs, which suggest a specific but mostly

compensatory role for these subunits [97]. This redundancy may be sufficient to maintain settled transcriptional programs but deficient to respond to different stimulus, as suggested changes in PCL expression during mESCs differentiation or the defects in PRC2.1 knockout (KO) mice (discussed below) [129]. Recently, the multiple PCL1-3 depletion in mESCs shown that loss of all PCL1–3 caused PRC2.1 eviction [97]. However, PRC2.1 eviction only induce partial reduction of H3K27me₃, and do not affect PRC2.2 binding (Figure I7C). This propose that PRC2.1 and PRC2.2 may synergize in mESC to ensure the faithful deposition of H3K27me₃.

The PCL recognition of H3K36me_{2/3} has been associated with a dual function in guidance and regulation of PRC2 activity. Indeed, PHF1 and PHF19 targeting of PRC2 at H3K36me₃ regions have been associated with both, inhibition of PRC2 activity by H3K36me₃ [130-132], and recruitment of H3K36me₃ demethylase NO66 at transcribed regions follows by H3K27me₃ deposition [133-135]. Thus, it seems likely that in addition to promote PRC2 binding at H3K36me₃ chromatin regions, PCL protein stabilize local interactions with H3K36me₃ to modulate PRC2 activity (Figure I7A). In addition, PHF1 interaction with CGIs regions increased 2- to 3-fold PRC2 stability at chromatin, which might enhance H3K27me₃ deposition at CGIs [122].

On the other hand, the mutually exclusive EPOP and PALI1 may function to link PRC2.1 with other transcriptional regulators. EPOP serves as a scaffolding protein to the Elongin complex component EloB/C, which may stimulate the elongation activity of RNA

Introduction

polymerase II [136-138]. This mechanism might help to inhibit PRC2 chromatin binding and maintain basal level of PRC2-targets transcription. Accordingly, the removal of either EPOP or EloB/C results in increased PRC2 and H3K27me3 occupancy [136]. Finally, PALI1 interact with the CTBP1/2 co-repressors or the H3K9 methyltransferases EHMT1/2 [139-141], which may explain the moderate H3K37me3 decrease in PALI1 KO mESCs.

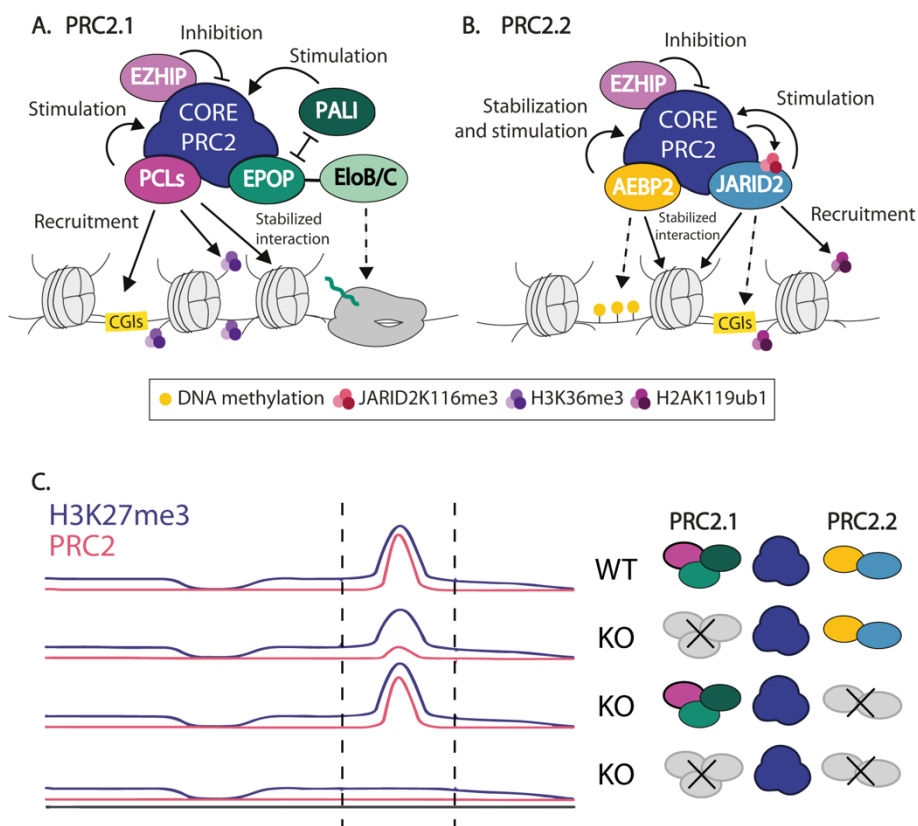


Figure I7. Cofactors regulating PRC2 recruitment and activity (A). In PRC2.1, all PCLs can recruit core PRC2 to unmethylated CGI and associate with H3K36me3 for specific targeting. PHF1 can extend the residence time of PRC2 and stimulate its catalytic activity. PALI1 can stimulate the catalytic activity of PRC2.1, whereas EPOP is likely to associate with EloB/C to maintain low levels of transcription. (B) In PRC2.2, AEBP2 and JARID2 can stimulate core PRC2 activity and increase its binding affinity to nucleosomes. AEBP2 stabilize PRC2.2 and can bind to methylated DNA *in vitro*. JARID2 can facilitate the recruitment of PRC2.2 through interaction with H2AK119ub, and also binds to GC-rich DNA *in vitro*, albeit the

function of this binding remains unclear. PRC2 can methylate JARID2, which in turn allosterically activate the enzymatic activity of PRC2. EZHIP exists as a PRC2 inhibitor (adapted from [142]). (C) Scheme depicting certain redundancy between PRC2.1 and PRC2.2 for H3K27me3 deposition (adapted from [96]).

2.3.3. PRC2.2

PRC2.2 binding is mostly mediated by JARID2, which binds DNA with a preference to CGIs *in vitro* [122, 143] and regulates the PRC1-recruitment of PRC2 by recognizing H2AK119ub1 (Figure I7B) [102, 144, 145]. This mechanism appears to contribute to create a positive-feedback loop between PRC2 and PRC1 that keeps genes from exiting PcG repressed state and allows for spreading of H3K27me3 [144]. Similar to PRC2.1, disruption of PRC2.2 function through the loss of JARID2 or H2AK119ub depletion release PRC2.2 from chromatin, but did not disrupt targeting of PRC2.1 [97]. Indeed, the combined loss of both PRC2.1 and PRC2.2 by KO of PCL1-3 and JARID2 is required to induce global mislocalization of PRC2 from highly enriched PRC2 sites to regions that previously contained little or no binding of core PRC2, which come along with a reduction on H3K27me3 and an increase on H3K27me2 [97]. This supports the model in which the specific accessory proteins within PRC2.1 and PRC2.2 cooperate to direct and stabilize PRC2 core proteins via both synergistic and independent mechanisms, while core PRC2 alone mediate more transit and unspecific chromatin interactions that are not strong enough to promote H3K27me3 (Figure I7C).

Of particular interest, JARID2 has been also identified as a nonhistone substrate of PRC2, which results in JARID2K116me3. JARID2K116me3 mimics H3K27me3 and binds EED to promote

EZH2 allosteric activation, which may account for a mechanism to confer PRC2 with a pioneering function that can initiate silencing regardless of H3K27me3 (Figure I7B) [108, 146]. Additionally, *in vitro* biochemical assays suggest a secondary modest contribution of both JARID2 and AEBP2 in enhancing EZH1/2 activity by increasing the ability of PRC2.2 to bind oligonucleosomes, specially at H2AK119ub1-containing H2A-B histones [123, 144]. Finally, JARID2 and AEBP2 might also challenge H3K36me3 and H3K4me3 PRC2.2 inhibition, as both cofactors can mimic histone tails and compete for their binding to EZH2. Interestingly, this competition underscores a mechanism by which PRC2.2 autoregulates its activity to override the epigenetic context of the targeted chromatin.

2.4. Chromatin components in mediating PRC2 function

Beyond PRC2 composition itself, the chromatin environment modulates PRC2 recruitment and catalytic activity through different mechanisms involving DNA conformation, short and long non-coding RNAs (lncRNAs), histone PTMs, or additional chromatin-bound proteins, among others [94].

2.4.1. DNA

PRC2 subunits have demonstrated sensitivity to particular DNA features. For instance, the length of linker DNA modulates EZH2 activity, which is 5 to 6-fold higher at 40 bp than on 30 bp linker DNA, while beyond 40 bp EZH2 activity start to decline (Figure I8A) [103]. This could be related to the specific EZH2 enhanced HMT

activity at di-/oligonucleosomes as compared with mononucleosomes, suggesting that the simultaneous interactions of PRC2 with two adjacent nucleosomes is required for optimal catalytic activity [107, 147]. In addition, MTF2 have demonstrated to preferentially recognize a particular unwound DNA helix shape within unmethylated CGIs, therefore narrowing down the potential targets and providing a potential explanation to the large fraction of PRC2-unbound unmethylated CGIs [148]. Hence, DNA shape may represent the head effector of PRC2 recruitment at unmethylated CGIs, rather than a specific primary sequence. Another critical feature is DNA methylation, as PRC2 binds exclusively at unmethylated DNA, and suggest a DNA methylation function in focusing PRC2 targeting [149-152].

Finally, Polycomb has also been implicated in more long-range interactions, named Polycomb bodies, with strong contacts between separate H3K27me3 specifically marked loci. This 3D genome organization have been suggested to be dependent on PRC1, H3K27me3 and DNA methylation [153-159].

2.4.2. RNA

PRC2 extensively binds long noncoding RNAs (lncRNAs) and pre-messenger RNAs (pre-mRNA) both *in vitro* and *in vivo*, although the direct role and biological relevance of these interactions remain disputed [160-162]. Several lncRNAs selectively recruit PRC2 at specific chromatin regions, as HOTAIR mediated PRC2 recruitment at Hox genes or PRC2 contribution to X-chromosome silencing via Xist [163]. On the other hand, pre-mRNAs associate with high

Introduction

affinity to PRC2 in a promiscuous or non-specific manner [164-167], although certain preference for G-short repeats as G-quadruplexes have been described. The functionality of pre-mRNA-PRC2 interactions over PRC2 recruitment and activity is a current source of discussion. The RNA eviction model suggest that actively transcribed genes counteract PRC2 silencing through the direct competition to interact DNA [168-170], whereas in the RNA bridges model, RNA act as a bridge between PRC2 and chromatin, increasing PRC2 co-localization with the chromatin environment and ultimately counteracting RNA PRC2-inhibition (Figure I8B) [164, 171, 172].

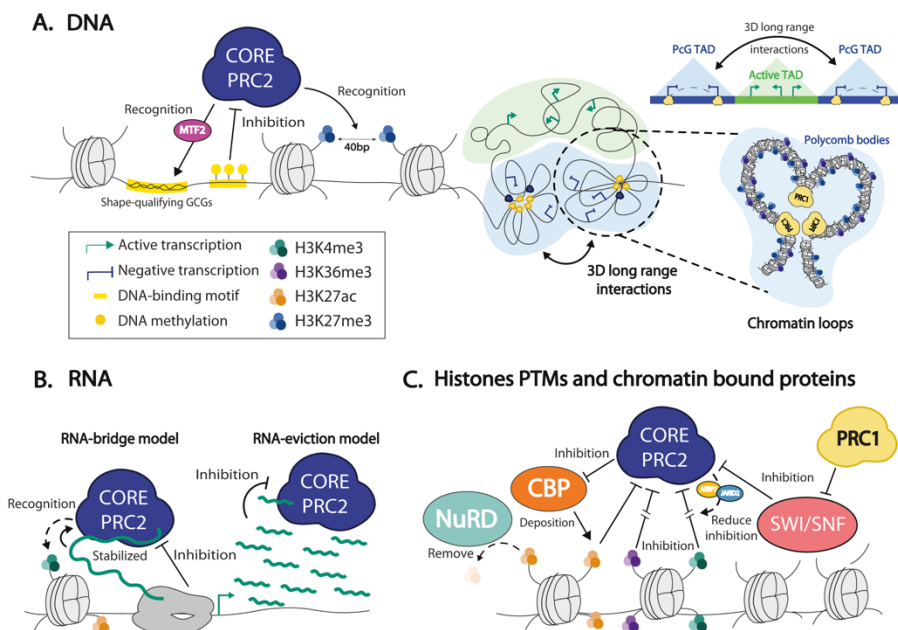


Figure I8. Chromatin scenario regulating PRC2. (A) DNA regulation of PRC2. MTF2 specifically recognize a DNA helical shape within CGIs to mediate PRC2 recruitment. DNA methylation disrupt PRC2-DNA interactions. EZH2 displayed higher affinity for 40bp internucleosomal DNA. (B) RNA regulation of PRC2. In the RNA-bridge model, PRC2 bound to the RNA in very close proximity to the chromatin, allowing PRC2 to interact with the chromatin environment and potentially deposit H3K27me3. In the RNA-eviction model RNA is transcribed at a very high rate, inhibiting and preventing PRC2 to bind chromatin. (C) Other histone marks and proteins regulation of PRC2. (adapted from [142])

2.4.3. Histones PTMs

Besides H3K27 methylation, other chromatin modifications and therefore their ‘writers’, ‘erasers’ and ‘readers’ proteins can influence PRC2 function. The Nucleosome Remodeling and Deacetylase (NuRD) complex mediate histone deacetylation of H3K27, which have been proven essential for PRC2 posterior association and repression on a subset of NuRD target genes in mESCs and human cancer (Figure I8C) [173-177]. Interestingly, on the other direction, the global loss of H3K27me2 results in an aberrant diffuse chromatin H3K27 hyperacetylation mediated by CBP and p300 activity, which has been associated with developmental failure [178, 179].

Furthermore, other active histone marks such as H3K4me3 and H3K36me2/3 strongly inhibit PRC2 activity *in vitro*, providing a rationale for the lower PRC2 activity in active chromatin environments [132]. However, PHF19 has been shown to interact directly with the H3K36me2/3 mark and induce the recruitment of H3K36me2/3-specific demethylases as KDM2B or NO66. Thus, KDM2B and NO66 may facilitate the removal of H3K36me3 and establish de novo H3K27me3 to target gene silencing during differentiation [133] [135, 180]. In addition, as mentioned above, AEBP2 and JARID2 have been shown to partially reduce H3K4me3 and H3K36me3 PRC2 inhibition through competition for EED binding. Finally, PRC2 has also demonstrated to interfere with repressive HTMs, such as JARID2 interaction with the H3K9 methyltransferase complex, which is essential to maintain H3K9 methylation levels [181]. Finally, histone demethylases such as UTX/KDM6A and JMJD3/KDM6B, and deubiquitinases including

BAP1, USP16, and USP21, can erase H3K27me3 and H2A K119ub1 marks to indirectly modulate PcG synergistic activity [94].

2.4.4. Other chromatin proteins

In addition to the abovementioned histone modifiers, other PRC2-independent proteins have been described to interact with PRC2 to modulate its function. The SWI/SNF ATP-dependent CRC have been suggested to drive a transition from inaccessible higher-order chromatin structure toward accessibility, and that this transition is due to the direct eviction of both PRC1 and PRC2. Bidirectionally, PRC1 has proved to inhibit SWI/SNF chromatin remodelling *in vitro*, which may contribute to the balance between SWI/SNF and Polycomb activity [182]. The alpha-thalassemia X-linked mental retardation (**ATRX**) chromatin remodeler has also been described to associate with PRC2 to modulate its interaction with Xist lncRNA [183] and some specific regions of the genome in mESCs [184]. In addition, multiple TF have been found to interact with PRC2 and mediate its chromatin recruitment, as E2F6, **REST**, RUNX1, ZNF281 or SALL4 [185-187].

3. PRC2 function in development and as a driver of disease

Multicellular organisms depend on the precise orchestration of chromatin organization to regulate transcriptional programs directing embryonic development and to maintain tissue homeostasis through their life spans [93]. The importance of PRC2 in chromatin regulation is underscored by the embryonic lethality of mice lacking a functional PRC2 complex and the large number of PRC2 alterations identified in human disease, especially cancer (see sections below). Therefore, a comprehensive characterization of the mechanisms controlling PRC2 activity will contribute to a better understanding of human pathologies and provide an opportunity for the development of new therapeutic strategies.

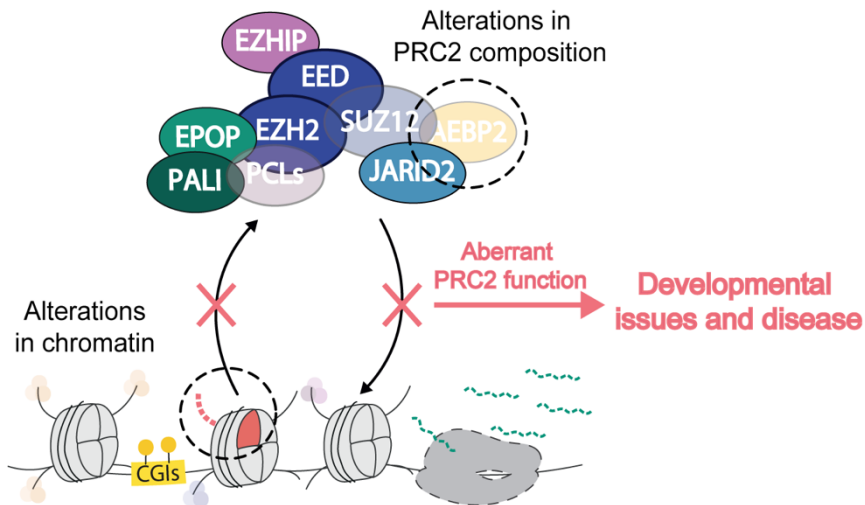


Figure I7. Alterations in PRC2 composition, by subunits deletions or mutations, and/or alterations in PRC2 interaction with chromatin, by aberrant changes in the chromatin landscape, impair PRC2 function and have been associated with several developmental problems and diseases.

3.1. Alterations in PRC2 composition affecting PRC2 function

3.1.1. Homozygous PRC2 KO mice

Homozygous mice germline KO for PRC2 core components, EZH2, EED, or SUZ12, lead to early embryonic lethality around gastrulation, day E7.5-E8.5 (Table I1) [188-191].

On the other hand, homozygous mice KO for PRC2 accessory proteins display delayed or less severe phenotypes than core components, which may be due to the partial compensation between PRC2.1 and PRC2.2 function (Table I1), still to be confirmed by generation of double-KO (DKO) mice. PRC2.1 and PRC2.2 accessory proteins demonstrate specific roles in mouse development and tissue homeostasis. Thus, PRC2.1 subunit PCL2/MTF2 KO mice display delayed embryonic lethality (around day E15.5) due to severe anaemia, whereas PALI1 KO leads to mouse perinatal mortality for still unknown reasons [192]. Conversely, KO mice for PRC2.1 subunits EPOP, PCL1/PHF1 [193] and PCL3/Phf19 are reported viable [194], although we have recently described pathogenic alterations in the embryonic and adult hematopoietic stem cell (HSC) compartment in PCL3/Phf19 homozygous KO mice (see in Results CHAPTER I). Regarding PRC2.2, JARID2 KO mice do not develop beyond E11.5 due to cardiac and neurodevelopmental defects [195-197], while AEBP2 KO mice models displayed different phenotypes depending on mouse strain, demonstrating either early embryonic lethality (E10.5) due to aberrant neuronal-crest differentiation, or pre-natal death (E15.5-18.5) due to cardiac defects [198, 199].

3.1.2. Heterozygous, or conditional PRC2 KO mice

The embryonic lethality of the homozygous germline KO mice for the PRC2 core components compromised further elucidation of PRC2 function in late development and adult tissue. Nonetheless, the deeper characterization of their predecessor heterozygous KO models early highlighted recurrent fails at the **hematopoietic system**, such as an increase in HSC proliferation, an HSC biased differentiation, and an ultimately increased incidence of hematopoietic malignancies (Table I1) [200-202]. As a consequence, most of the efforts aimed to elucidate PRC2 function *in vivo* in adult tissues have been focus on the hematopoietic compartment, so the generation of hematopoietic-lineage-specific conditional germline KO (cKO) mice have been done for all PRC2 core subunits. Notably, these PRC2 cKO mice display lethality either at embryo stages or immediately after birth due to impaired HSCs self-renewal (Table I1) [200, 203-205]. In addition, EZH1 and JARID2 cKO at the hematopoietic system develop successfully but demonstrated impaired HSC proliferation and differentiation, and increased incidence of hematopoietic diseases [205, 206].

Conditional deletion of PRC2 components in other adult tissues have also reported, although to a minor extent. For instance, EED specific deletion at intestinal or neuronal tissue also decrease the inherent tissue adult stem cells proliferation and impair differentiation capacities. This results in a compromise regeneration of the intestinal epithelium upon extensive damage and to premature death due to impaired neurogenesis [207-209]. Similarly, EZH2 skin-specific deletion induces acceleration of embryonic epidermal stem cells

Introduction

differentiation, which decrease their postnatal proliferation capacities, although do not seriously compromise epithelial tissue [210]. Finally, conditional ablation of JARID2 in pancreatic progenitors results in reduced endocrine cell area at birth due to impaired prenatal endocrine cell differentiation and proliferation [211], whereas early deletion of JARID2 in cardiac progenitors, but not in posterior differentiated cardiomyocytes, compromise normal cardiomyocyte lineage determination and lead to cardiac malformations and neonatal lethality [211, 212].

Overall, PRC2 function in late development and adult tissues appear to particularly compromise the multipotent progenitors that sustain the late stages of embryonic development and maintain tissues homeostasis during lifespan. [212]

PRC2 subunit	Homozygous KO	Heterozygous, conditional, inducible KO	Role in genetic syndromes and GoF and LoF mutations in cancer
EZH2	E7.5-E8.5 lethal [189]	Hematopoietic system cKO mice die immediately after birth due to impaired hematopoiesis [193]. Heterozygous KO mice display impaired HSCs homeostasis and high incidence of hematopoietic diseases [193]. Epidermal cKO impaired epidermal stem cells proliferation [203, 204]	Mutations in Weaver syndrome [213]. Increase expression in ataxia telangiectasia neurodegenerative disease [214]. Decrease expression in neurodegeneration in dementia [215]. Mutations in myelodysplastic syndromes [216, 217]. GoF and LoF mutations in several cancers (e.g. prostate cancer, T cell acute lymphoblastic leukaemia, DLBCLs etc)[218-227]
EZH1	Viable [188]	Hematopoietic system cKO mice display impaired HSCs homeostasis and myeloproliferative syndromes [206]. Heterozygous KO mice do not display major phenotypes.	To be determined (TBD)
EED	E7.5-E8.5 lethal [191]	Hematopoietic and neuronal tissues cKO mice die immediately after birth due to impaired hematopoiesis and neurogenesis. Heterozygous KO mice display impaired HSCs homeostasis and a high incidence of hematopoietic diseases [192]. Intestinal cKO mice display impaired intestinal stem cells and compromise regeneration of the intestinal epithelium. [200, 204]	Mutations in Weaver and Cohen-Gibson syndromes [213, 228] Mutated and misregulated in several cancers (e.g., leukaemia, lymphoma, etc.) [229-231]
SUZ12	E7.5-E8.5 lethal [190]	Hematopoietic system cKO mice die at E13.5. [232]. Heterozygous KO mice display impaired HSCs homeostasis and an	Fused in chromosomal translocation at endometrial stromal sarcoma, LoF mutations in specific MPNSTs tumours [233] and T cell acute lymphoblastic leukaemia [229, 234-237]

Introduction

		increased incidence of hematopoietic diseases.	
RBBP7	Viable [193]	TBD	GoF mutations in specific tumours [233, 238]
RBBP4	Pre- weaning lethality [193]	TBD	TBD
JARID2	E11.5 lethal [195-197]	Hematopoietic-specific cKO mice display impaired HSCs homeostasis and a high incidence of hematopoietic diseases [205]. Conditional cardiac and pancreatic-specific KO results in neonatal lethality and impair pancreas development respectively [211, 212]	Heterozygous deletions in intellectual disability [239]. Mutated and misregulated in hematopoietic cancers [205]
EPOP	Viable	TBD	LoF mutations in lung and thyroid tumours [99].
PAL1	Perinatal lethality[192]	TBD	TBD
AEBP2	E10.5 lethal or pre-weaning lethality [198, 199].	Heterozygous KO mice display a similar phenotype to Hirschsprung and Waardenburg syndromes [198].	TBD
PHF1	Viable [193]	TBD	Fused in chromosomal translocations at endometrial and ossifying fibromyxoid tumours [240, 241]
MTF2	E15.5 lethal due to severe anaemia [195, 196]	TBD	LoF mutations in leukaemia [242]
PHF19	Viable with aberrant HSC homeostasis.	TBD	Mutated in thyroid tumours and oesophageal squamous cell carcinomas [99]. LoF mutations in prostate cancer cells; GoF mutations in multiple myeloma [243, 244]
EZH1	Infertile. Oogenesis defects [128]	TBD	Fused in chromosomal translocation (endometrial stromal), highly expressed in PFA ependymoma [127, 245]

Table II. PRC2 proteins and their roles in development and disease (expanded from 77). TBD = to be determined. GoF = Gain of Function; LoF = Loss of Function; DLBCLs = Diffuse large B-cell lymphomas

3.1.3. Loss-of-function and gain-of-function PRC2 mutations in human disease.

In humans, mutations in PRC2 components have been related to different neurodevelopmental syndromes, including intellectual disability, to several hematopoietic malignancies, as myelodysplastic syndrome (MDS), to neurodegenerative diseases, like Huntington's disease, and especially to cancer [215, 239, 246-248]. Indeed, the most common PRC2 mutations in human somatic diseases concern EZH2 in **hematopoietic cancer malignancies**, with both gain (GoF) and loss (LoF) of function mutations being identified, suggesting that PRC2 can act either as a tumour suppressor or an oncogene [249]. This evidence highlights again the special link between the hematopoietic system and PRC2 function *in vivo*

GoF EZH2 mutations have been identified in 22% of Diffuse large B-cell lymphomas (DLBCLs) and 10% of follicular lymphomas (FLs), which lead to aberrantly high global levels of H3K27me3 at the promoters of genes required for cell cycle exit and terminal B-cell differentiation [227, 250, 251]. The high penetrance of EZH2 mutations on those cancers, the fact that these mutations were found to be an early mutational event in the emergency of the disease [252], and the development of lymphoma as a result of the GoF EZH2 mutant expression in mouse B-cells, suggest that increased PRC2 activity could act as a driver mutation. In addition to hematopoietic cancers, GoF EZH2 mutant has also been found in solid tumors, as sporadic parathyroid adenomas [253], Erwing sarcoma, and in melanomas [254].

Introduction

On the other hand, LoF mutations on EZH2 lead to spontaneous development of T-cell lymphoblastic leukaemia (T-ALL) and are present in almost 2% of acute myeloid leukaemia (AML) patients, which harbours the lowest mutation frequency of adult cancers (0.28 mutations per megabase), suggesting that LoF EZH2 mutations could also be a driving event in these tumours [203, 224]. In addition, LoF EZH2, SUZ12 or EDD mutations have been identified in other leukaemias, malignant peripheral nerve sheath tumours (MPNST), melanoma, and glioblastoma multiforme (GBM).

PRC2 accessory proteins display lower mutation rates in cancer, probably due to their cell type-specific expression and partial redundancy. LoF mutations in JARID2 and AEBP2 are present in up to 6.5% and 5.2% of AML [255, 256]. LoF mutations in EPOP have been observed in lung and thyroid tumours [99]. In addition, mutations in PHF19 have been identified in thyroid tumours and oesophageal squamous cell carcinomas (SCC), although the consequences of these mutations are not determined [99]. Notably, even if PRC2 accessory proteins display lower mutation rates, they have been recurrently found misregulated in several cancers. For instance, PHF19 overexpression has been reported to control the switch between proliferative and invasive states in melanoma and prostate cancers [243, 257], and to correlates with enhanced cancer cell proliferation in gastric cancer [258] and glioblastoma [259]. Finally, PHF19 overexpression has been suggested to increase the malignant progression of multiple myeloma (MM) to AML [244].

Overall, based on these observations, a lot of effort has been put in developing targeting strategies to inhibit the activity of PRC2, or associated mechanisms, to treat several cancers. Accordingly, several small-molecule inhibitors of EZH2 have been developed and are either being tested in clinical trials or already in use in clinical practice, such as for AML therapy [142]203[260].

3.2. Alterations in the chromatin environment affecting PRC2 function.

Perturbations in the chromatin environment can also impact PRC2 function and contribute to its role in human pathogenesis. For instance, SWI/SNF complexes are mutated in a large number of human neurologic diseases and more than 20% of all human cancers, and its mechanisms of pathogenesis appear to be related to the opposition between SWI/SNF and PcG complexes [261]. Thus, in atypical teratoid rhabdoid (ATRT) **pediatric brain tumors**, which harbor in 95% of the cases LoF mutations in the SWI/SNF subunit SMARCB1, PRC2 binds at 80% of the promoters that are SMARCB1 targets in healthy tissue, and that become aberrantly repressed in ATRT [262]. These promoters are also enriched in REST signal, a TF whose primary function is to suppress neuronal gene transcription in neuronal precursors and nonneuronal cells [263]. This suggests that REST may direct the PRC2 complex to repress neuronal genes, which could impair differentiation of neuronal precursor cells and drive cancer formation. Interestingly, similar cooperation between EZH2 and REST TF to repress neuronal genes have been described in paediatric neuroblastoma (NB) harbouring In-Frame Fusion (IFF)

mutations in the SWI/SNF-like chromatin remodeler ATRX [264]. As a consequence, the inhibition of EZH2 as a therapy for ATRT and NB have been suggested and is currently evaluated as a treatment for ATRT in different clinical trials (Clinical Study Identifier: NCT02601937; NCT02601950).

Recently, the identification of high penetrance mutations in all four core histone genes, also known as ‘oncohistones’, has marked a new chapter in cancer epigenomics [265]. Notably, the first ‘oncohistone’ was identified in up to 80% of paediatric diffuse intrinsic pontine gliomas (DIPG) and involve the substitution of lysine 27 by methionine in histone 3 variant 1 or 3, also listed as H3.1/3K27M [266]. Interestingly, although H3K27M mutations occur in each case at only one of the 16 genes coding for H3, representing only 3%–17% of the H3 pool, early works demonstrated that H3K27M leads to a H3K27me3 global reduction and an aberrant reorganization of chromatin features such as a global increase in H3K27ac [266-270]. This suggests a dominant role for H3K27M that have been mostly linked to the direct alteration of EZH2 activity [267, 270-275]. Hence, in addition to being the first oncohistone identified and the absence of effective therapies for DIPG, H3K27M gain lots of attention due to the potential of these tumours as a means for understanding PRC2 function in chromatin.

3.2.1. H3K27M in mediating PRC2 function in DIPG

Despite great efforts that have been done to elucidate the molecular mechanisms involved in H3K27M mutation, these remain highly controversial [267, 270, 276-280]. The consensus is that H3K27M

induce a general loss of H3K27me3 all over the genome [266], with the exception of certain unmethylated CGIs considered strong PRC2-targets [267, 270, 277-280].

Initially, H3K27me3 reduction was associated with increased retention of PRC2 at H3K27M containing nucleosomes [277], which may sequester PRC2 and lead to restricted EZH2 activity (Figure I9A). This model is consistent with: i) the PRC2 higher affinity for H3K27M; ii) an H3K27M dose-dependent inhibition effect [280]; iii) with single-cell imaging evidences suggesting a 1.5-fold increase of EZH2 residency on H3K27M chromatin [240]; and iv) a partial correlation between H3K27M, PRC2 and retained H3K27me3 loci in DIPG cell lines [277, 281].

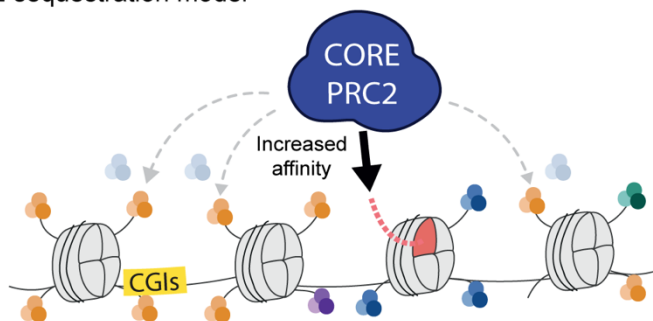
However, some studies sustain a mutually exclusive localization of PRC2 and H3K27M in DIPG [279]. In addition, following H3K27M induction in oligodendrocyte precursors (OPCs), H3K27me3 lost comes first at H3.3K27M specific sites, which is also contradictory with the PRC2 sequestration model [282]. In this line, recent data also suggest an alternative model in which H3K27M excludes PRC2 binding as the result of an aberrant accumulation of H3K27ac on the other wildtype copies of histone H3 in H3K27M-wild type (WT) heterotypic nucleosomes (Figure I9B) [279].

In keeping with both models, a recent study suggested that H3K27M sequestration of PRC2 is transient, although PRC2 inhibition persisted after being released from H3K27M suggesting a lasting impact of H3K27M on PRC2 [280]. However, this study was

Introduction

performed using 293T cells expressing exogenous H3K27M, so it remains uncertain as to whether these results can be extrapolated to DIPG. Overall, the differential PRC2 distribution and H3K27me3 retention in the presence of H3K27M remains an open question, most probably due to the differences between study models.

A. PRC2 sequestration model



B. PRC2 exclusion model

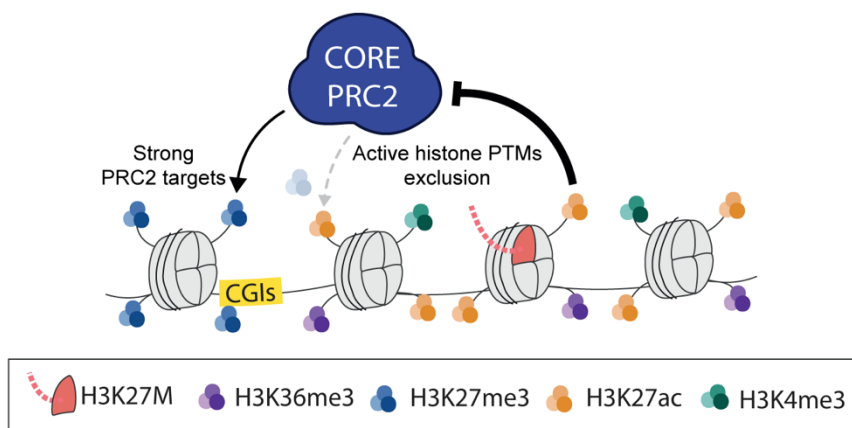


Figure I9. Schematic representation of both H3K27M models for PRC2 inhibition. (A) The PRC2 sequestration model sustains that the increased affinity for H3K27M retains PRC2 at H3K27M regions, which results in genome-wide H3K27me3 depletion outside of H3K27M regions. (B) The PRC2 exclusion model sustains that the increase of active histone PTMs at H3K27M containing nucleosomes results in PRC2 eviction from chromatin at H3K27M regions, which in turn leads to genome-wide H3K27me3 depletion with the exception of strong PcG targets.

Interestingly, in addition to a general increase in H3K27ac [266, 279], several studies reported further chromatin alterations dependent on H3K27M. H3K4me3 showed a slight increase and redistribution to bivalent chromatin regions [283, 284], where there is also an enrichment of H3.3K27M variant deposition. H3K9me3 increase at the regions losing H3K27me2/3 marks which, together with a still unclear general loss and redistribution of DNA methylation, might compensate for the absence of H3K27me3 repressive effect [270]. H3K36me2 is increased all over the genome, although the presence of normally abundant H3K27me2-H3K36me2 combinatorial domains decrease [280, 285]. Notably, this loss of H3K27me2-H3K36me3 domains is not only due to a reported global reduction on H3K27me2 but also to a predominant redistribution of H3K27me2 to previously H3K27me3-occupied regions, which classically are absent of H3K36me2/3. Indeed, interestingly, although global levels of H3K27me1 seems unaffected, there is a drastic redistribution from gene bodies to intergenic regions, mimicking normal H3K27me2 patterns. All at once, the changes in the distributions of all three K27 methylation marks may be explain by the reduction of one K27 methyl group throughout the genome in the presence of HK27M (K27me3 \rightarrow K27me2; K27me2 \rightarrow K27me1; K27me1 \rightarrow K27me0).

Overall, the extensive reorganization of the chromatin landscape stresses that several chromatin regulatory mechanisms are involved in sustaining DIPG tumorigenesis. This could partially explain how, despite the dramatic loss of H3K27me3, there are relatively modest transcriptome changes in DIPG [254, 255][283]. In this line, several studies suggest that H3K27M transcriptional changes correlate with

Introduction

H3K27ac gain at a few super-enhancers, which associate with upregulation of genes involved in neural differentiation and developmental pathways [286]. In this line, in addition to therapeutic strategies using EZH2 inhibitors, new therapies targeting other H3K27M chromatin alterations has demonstrated antitumor activity, as inhibition of the bromodomain (BRD) family of proteins or different deacetylases [266, 286, 287]. Therefore, despite DIPG chromatin misregulation remains largely unexplored beyond H3K27me3, its further characterization could be essential for the development of targeted therapeutic strategies for DIPG.

AIM

This thesis project was divided into two complementary aims that pursue to understand the consequences of misregulation of PRC2 composition, and the interplay between PRC2 and chromatin, in adult tissue homeostasis and disease. These aims were enclosed in two different chapters, which in turn include different objectives:

CHAPTER I: Alterations in PRC2 composition.

The main aim of the first chapter was to investigate the consequences *in vivo* of the depletion of PHF19 in the hematopoietic stem cell compartment. In an attempt to address this aim the following objectives were defined:

1. Investigate *in vivo* the consequence of PHF19 depletion on the transcriptome of the hematopoietic progenitor cells.
2. Investigate *in vivo* the consequence of PHF19 depletion on the epigenome of hematopoietic stem cells.
3. Characterize *in vitro* the consequence of PHF19 depletion on the proliferative and differentiation capacities of hematopoietic stem cells.
4. Characterize *in vitro* the molecular mechanism of PHF19 in hematopoietic stem cells.

CHAPTER II: Alterations in the interplay between PRC2 and chromatin.

The main aim of the second chapter was to identify acquired epigenetic vulnerabilities in DIPGs for selective targeting of tumoral cells. In an attempt to address this aim the following objectives were defined:

Aim

1. Identify proteins essential to maintain DIPG cells viability using a CRISPR-Cas9 dropout screening.
2. Characterize the total proteome and the chromatin-bound proteins of DIPG cells.
3. Investigate the potential as therapeutic targets of chromatin-bound proteins enriched in DIPG cells.

RESULTS

CHAPTER I

Characterization of the consequences *in vivo* of the depletion of PHF19 in the hematopoietic stem cell compartment.

Pedro Vizán*, Arantxa Gutiérrez*, Isabel Espejo, Marc García-Montolio, Martin Lange, Ana Carretero, Atefeh Lafzi, Luisa de Andrés-Aguayo, Enrique Blanco, Roshana Thambyrajah, Thomas Graf, Holger Heyn, Anna Bigas, and Luciano Di Croce. [The Polycomb-associated factor](#)



1. Previous generation and phenotypical hematopoietic characterization of a *Phf19* KO mice.

The haematopoietic system is the collection of organs and tissues, including the bone marrow, spleen, thymus and lymph nodes, involved in hematopoiesis [288]. Hematopoiesis is the process by which both fetal and adult organisms produce the diverse repertoire of blood cells from a single multipotent progenitor cell type, **the hematopoietic stem cells (HSCs)** (Figure R1) [288, 289]. The hematopoiesis process required the precise contributions of different classes of chromatin regulators to fine-tune HSC self-renewal and differentiation during development and adult lifetime. PcG proteins have been identified as major regulators in the safeguard of the fetal and adult HSC transcriptional programs, as well as in the dynamic chromatin reorganization responsible of lineage commitment [249]. Accordingly, failures in hematopoiesis are among the most common phenotypes found in PcG KO mice and PcG mutations in adult cells have been mostly reported in hematopoietic cancers (Figure R1) [249].

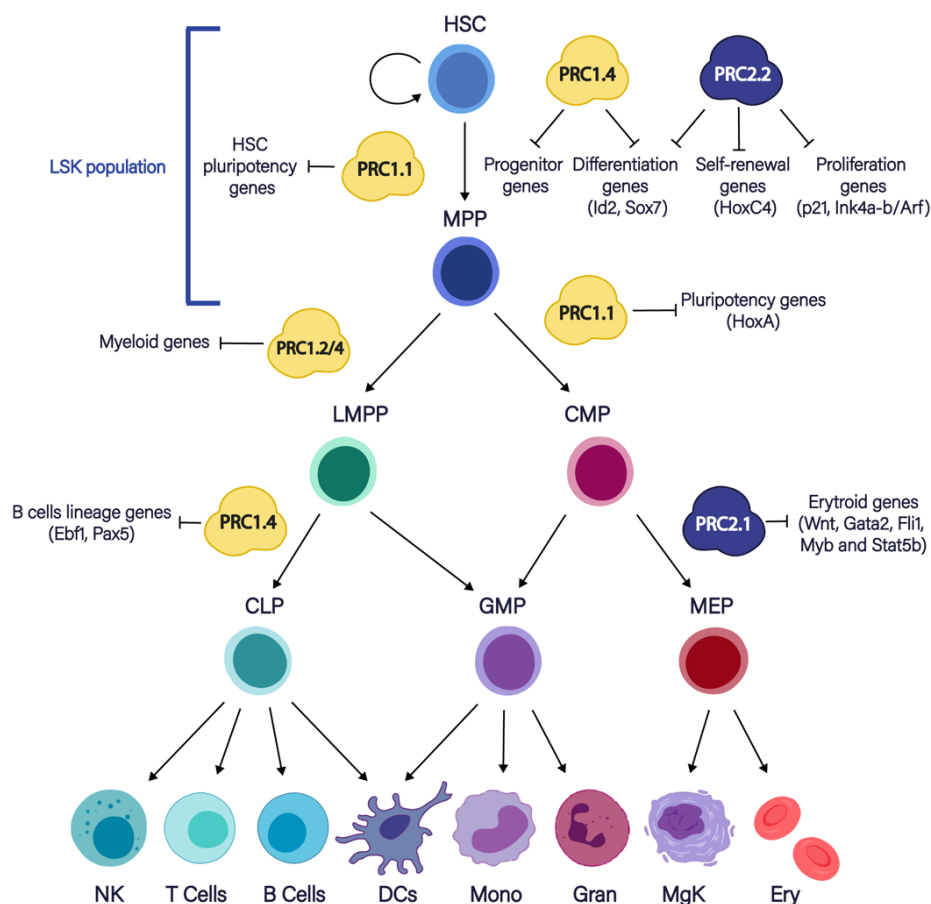


Figure R1. Scheme depicting the canonical stepwise hematopoiesis model and the described PcG proteins function in this pattern (adapted from [249]). The cellular bone marrow composition is extremely heterogeneous and includes different populations of progenitors and differentiated blood cells that can be discriminate according to the expression of specific cell surface markers. Traditionally, these different populations have been hierarchically organized according to their self-renewing and differentiation potential, although recent single-cell studies have challenged this classical step-wise model [290]. PRC = Polycomb Repressive Complex. HSC = Hematopoietic Stem Cells; MPP = Multipotent Progenitors; LSK = Lin⁺ Sca1⁻ c-Kit⁺ Cells; LMPP = Lymphoid-Primed Multipotent Progenitors; CMP = Common Myeloid Progenitors; CLP = Common Lymphoid Progenitor; GMP = Granulocyte-monocyte Progenitors; MEP = Megakaryocyte/ Erythrocyte Progenitor; NK = Natural Killer; DCs = Dendritic Cells.

The role of PH19, a substoichiometric PRC2.2 cofactor, have been extensively addressing *in vitro* by several laboratories including ours [130-132]. Recently, the generation and characterization of germline

Phf19 KO mice were reported for the first time by our laboratory [194]. This section summarizes the main phenotypic findings published (a full detailed version of the results could be found in Annex I), which are essential and previous to this thesis contribution regarding the multi-omic characterization of this model.

Phf19 KO mice were reported to be viable and breed normally, but upon ageing (>60 weeks) *Phf19*-KO mice displayed a high penetrance of splenomegaly, an enlargement of the spleen frequently associated with hematopoietic malignancies, such as leukemia or MDS (Figure R2A) [194]. The analysis of the hematopoietic system in *Phf19* KO versus control *Phf19* floxed adults mice shown a significant reduction in whole bone marrow (BM) cellularity, HSCs and LSK progenitors (see Figure R1), and a decrease proportion of proliferative cells within the adult HSC compartment (Figure R2B) [194]. These data suggest a significant increase in HSC quiescence in homeostasis in *Phf19* KO compared to control *Phf19* floxed, which may lead to a decrease in HSC, LSK, and total BM cellularity in adult mice.

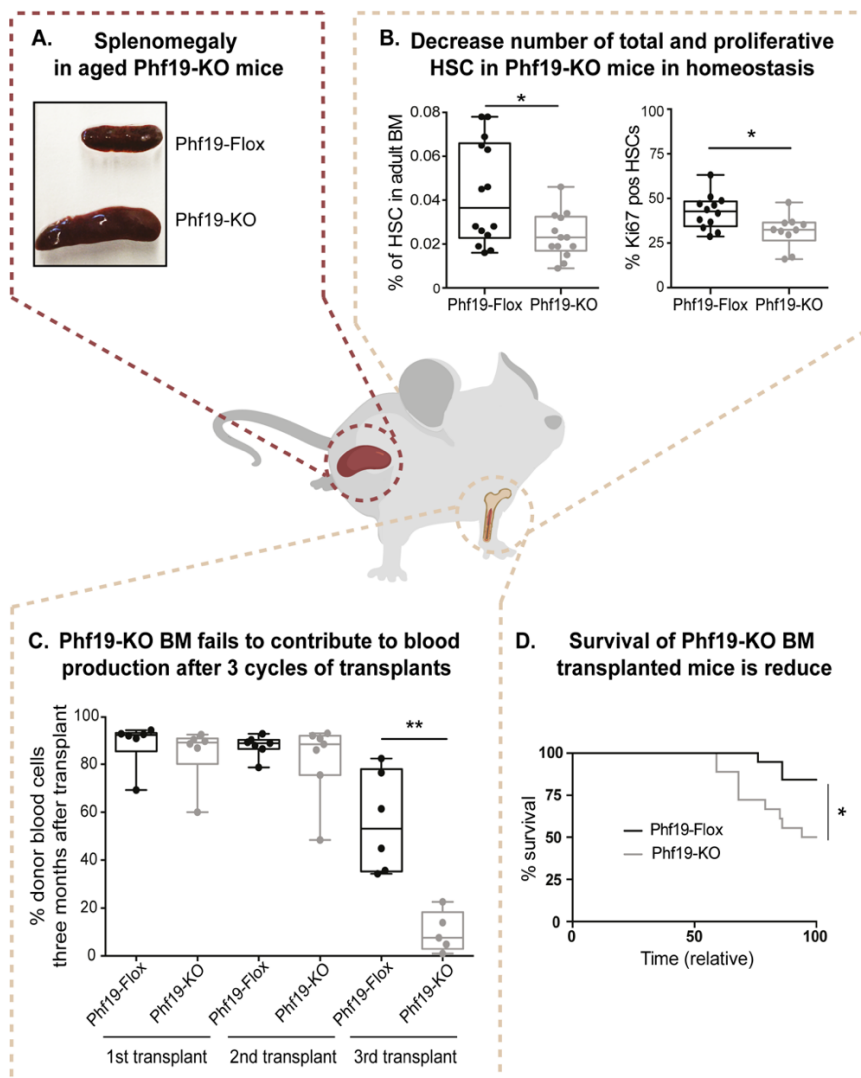


Figure R2. Summary of the main results extracted from the characterization of the hematopoietic system of *Phf19* KO mice (for a full description see Annex I) [194]. (A) Picture depicting splenomegaly in aged *Phf19* KO mice (>60 weeks). (B) Surface markers defined HSC quantification from 14 control *Phf19* floxed and 13 to 15 *Phf19* KO mice, with percentages shown for the live cells measured as DAPI negative. Percentage of Ki67+ HSC cells in 12 mice for *Phf19* floxed and 10 mice for *Phf19* KO. (C) Percentage of donor-derived blood cells in 3 months after transplant recipient mice. (D) Survival curve of all mice *Phf19* floxed versus *Phf19* KO after their last transplant. Time normalized for each experiment to final sacrifice time of all animals.

Serial BM transplantation assay is the gold standard technique to determine *in vivo* the capacity of the hematopoietic precursors from a donor mouse to repopulate the hematopoietic compartment of BM-depleted recipient mice [291]. The increasing number of serial transplants results in a decrease in the self-renewal and differentiation capacity of HSCs [291]. HSCs from young WT mice normally can reconstitute recipient bone marrow for 4–5 cycles of transplantation. Notably, after 3 cycles of transplantation, mice recipient from a *Phf19* KO donor BM displayed: i) an increase in LSK, but a decrease contribution to blood production; ii) a differentiation bias toward myeloid lineage, and; iii) a decrease survival of recipient mice after their last transplant. Overall, these results pointed to a compromise HSC long-term repopulation capacity of *Phf19* KO HSCs (Figure 2C-D) (see Annex I) [194].

Similarly, aged *Phf19* KO mice also displayed an increase in LSK, and an increased myeloid differentiation bias when compared to aged *Phf19* floxed mice (see Annexe I Figure S4A). Recipient mice transplanted with BM from aged *Phf19* KO mice shown an exacerbated phenotype, with a decrease contribution to blood production after only 1 cycle of transplantation, an increased incidence of splenomegaly, and a dramatic decrease in survival (Figure R2D) [194]. Flow cytometry characterization of their BM and spleen showed the presence of uncharacterized highly proliferative populations, suggesting malignant haematopoiesis (see Annexe I Figure 5). In this line, the histological analysis of the spleen

revealed an aberrant cell population compatible with a potential myeloid dysplastic syndrome (see Annexe I Figure 5) [194].

Altogether, these previous results from our laboratory indicate that PHF19 may be necessary to maintain a correct hematopoietic balance upon lifespan and that its depletion induces an aberrant state priming HSCs for malignant progression. However, the molecular mechanisms underlying this observed phenotype remained unaddressed. Therefore, to gain insights into the observed phenotype of *Phf19* KO mice, this first chapter aimed to investigate the molecular function of PHF19 in regulating the hematopoietic progenitor cells.

2. *Phf19* KO HSCs display an increased HSC identity and a highly quiescent transcriptome.

To gain insight into how PHF19 deletion impact correct haematopoiesis, and more specifically the HSC compartment function, we decided to perform RNA-seq analysis in two different hematopoietic progenitor populations, LT-HSC and LSK. These populations can be discriminated using different combinations of cell surface markers. The **Lin⁻Sca-1⁺c-Kit⁺** combination of markers, also known as **LSK**, encompass the whole diversity of hematopoietic progenitors, which are classically defined as a group of multipotent populations that differ in their self-renewal capacities (Figure R3A). Moreover, using a series of surface markers known as SLAM, more stem populations can be discriminated within LSK. The SLAM family CD150 and CD48 cell surface receptors allow to isolate **LT-HSC (LSK CD150⁺ CD48⁻)**, which are at the top of the hematopoietic system hierarchy and harbour the highest ability to self-renew and generate a continuous turn-over of blood cells [292, 293]. Importantly, single-cell techniques have demonstrated that the LT-HSCs population is still a heterogeneous and multifaceted entity, which comprise cells that are more quiescent and multipotent, namely **dormant LT-HSCs (dHSC)**, and others that are cycling and can be partly primed toward myeloid or lymphoid phenotypes, namely **active LT-HSC (aHSC)** (Figure 3B). However, there are not surface markers to solve this heterogeneity [294-299].

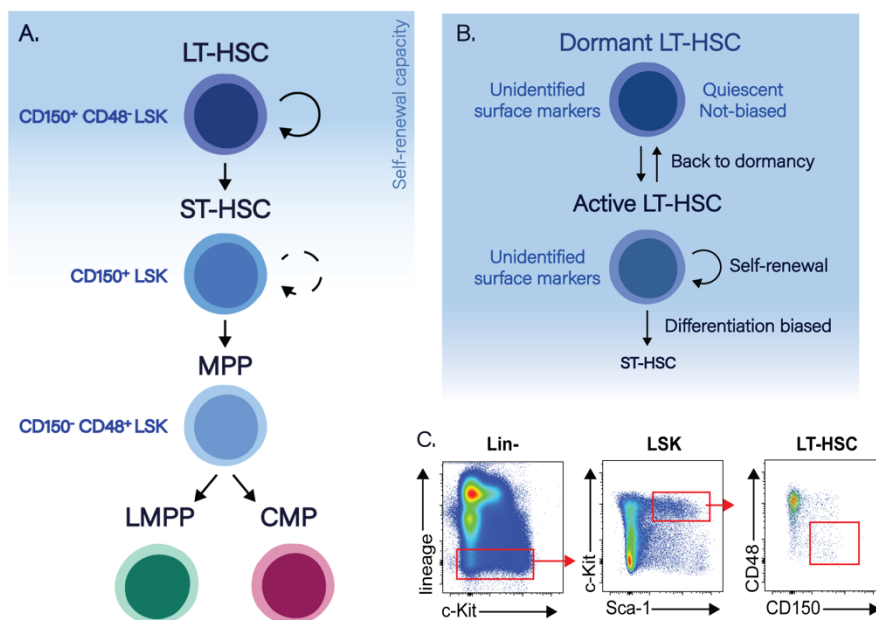


Figure R3. (A) The main multipotent populations within the LSK compartment according to classical surface markers segregation strategies. Besides their surface markers, these populations differ in their self-renewal capacities (blue gradient). (B) The LT-HSC compartment populations. aHSCs can contribute to homeostatic differentiation during steady-state or to self-renewal, while dHSCs acts as a reserve pool to produce differentiated cells only in response to disturbances. (C) BM gating strategy for isolation of defining LSK, and LT-HSC to perform RNA-seq analysis.

LT-HSCs are a really rare population with an estimate frequency of 0.01% of total BM nucleated cells [300], so we managed to extract an average of 1000 LT-HSC per animal. Therefore, we decided to perform Smart-Seq2, a protocol designed for single-cell RNA-seq libraries [301], to perform our RNA-seq analysis using 5000-10,000 **LT-HSCs** and **LSK** cells extracted from young (8 weeks) *Phf19* floxed and *Phf19* KO mice (Figure R3C). We performed duplicates for LT-HSCs and a single replicate for LSK cells. The library preparation, the sequencing run, and the quality control of the sequencing were performed by the CRG Genomics facility. The read mapping and counting were performed by Dr. Enrique Blanco.

First, in order to elucidate the consequences of PHF19 depletion from the outset of the haematopoiesis process, we conduct differential gene expression (DE) analysis between *Phf19* floxed and *Phf19* KO LT-HSC RNA-seq data considering both replicates. We identified 901 upregulated (fold change (FC) >1.5) and 556 downregulated (FC <-1.3) genes in *Phf19* KO LT-HSC (Figure R4A). Gene Ontology (GO) term enrichment analysis of these genes identified within the top upregulated categories retinoic acid receptors, and calcium activity, and within the top downregulated the ribosomes biosynthesis, mRNA transcription, energy production, cell cycle, and HSC differentiation (Figure R4B). Interestingly, the top categories differentially enriched in dHSC versus aHSCs are an increase in retinoic acid receptors signalling and calcium activity, and a general downregulation of biosynthesis, energy production and cell cycle, [294, 296, 297]. Thus, our results suggested that the transcriptome of *Phf19* KO LT-HSCs could be similar to the transcriptome of dHSCs.

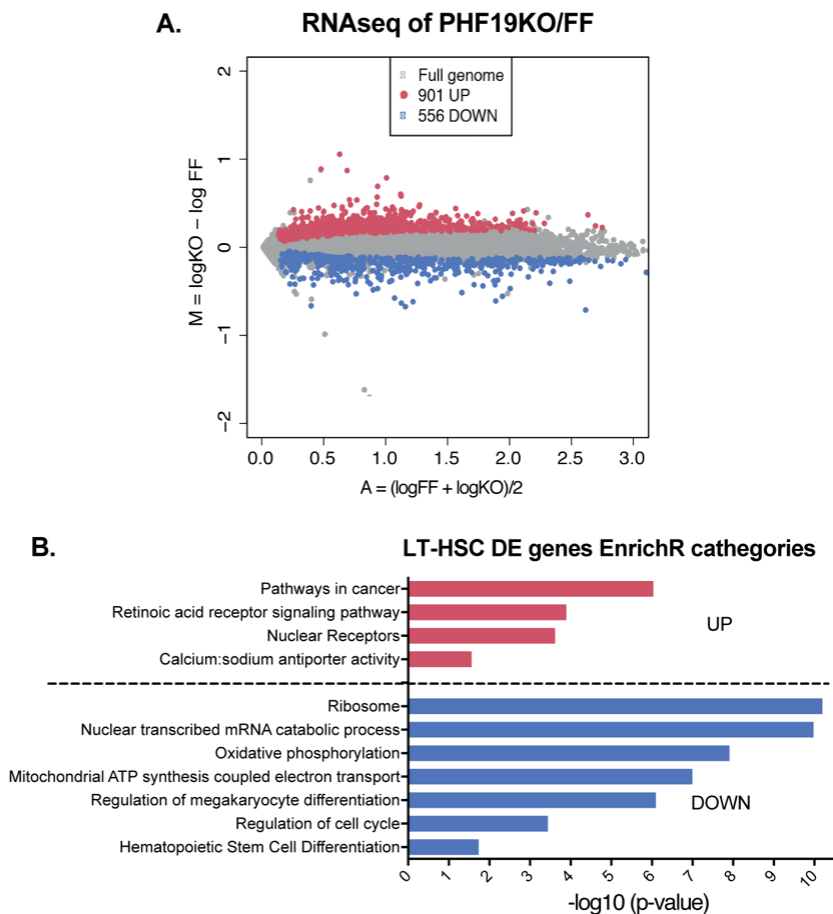


Figure R4. (A) MA plot for differential expression analysis in *Phf19* floxed and *Phf19* KO LT-HSC RNA-seq replicates. Every dot represents a gene, the x-axis indicated \log_2 of the average *Phf19* floxed and *Phf19* KO RPKM values from both replicates, and the y-axis indicates \log_2 FC of *Phf19* KO over *Phf19* floxed RPKM average values from both replicates. Significant upregulated (FC>1.5) and downregulated genes (FC<-1.3) are highlighted in red and blue respectively. (B) Term enrichment analysis of DE genes using EnrichR, a web-based gene ontology tool developed by the Ma'ayan Laboratory [302, 303]. Relevant categories are selected and ranked by p-value.

2.1. *Phf19* KO HSCs transcriptome display similarities with dormant HSC.

To further support the potential increase of the transcriptional programs sustaining dHSC identity in *Phf19* KO LT-HSC, we

performed gene set enrichment analysis (GSEA) using the gene signatures most commonly associated with HSC and dormancy. Interestingly, HSC-fingerprint genes set extracted from Chambers et al. and Gazit et al. were highly enriched in *Phf19* KO LT-HSC, indicating that HSC identity is enhanced upon PHF19 depletion (Figure R5A) [304, 305]. Further, we generated a signature using the genes DE between dHSC and aHSC according to Lauridsen et al. [297], and we found a significant enrichment of dHSCs transcriptional identity in *Phf19* KO LT-HSC (Figure R5B). We also used the gene signatures described by Cabezas-Wallscheid et al. to be characteristic of dHSCs transcriptome [296], i.e. a decrease in the Myc network, ribosome, and mRNA processing, and an increase in retinoic acid targets. Notably, we obtained even a more significant enrichment for these gene sets in *Phf19* KO LT-HSCs comparing to *Phf19* floxed LT-HSCs than their dHSCs versus aHSCs comparison results (Figure R5C). Altogether, the further transcriptomic analysis suggested that *Phf19* KO LT-HSCs display an increase in HSC- and dormant identity in homeostasis.

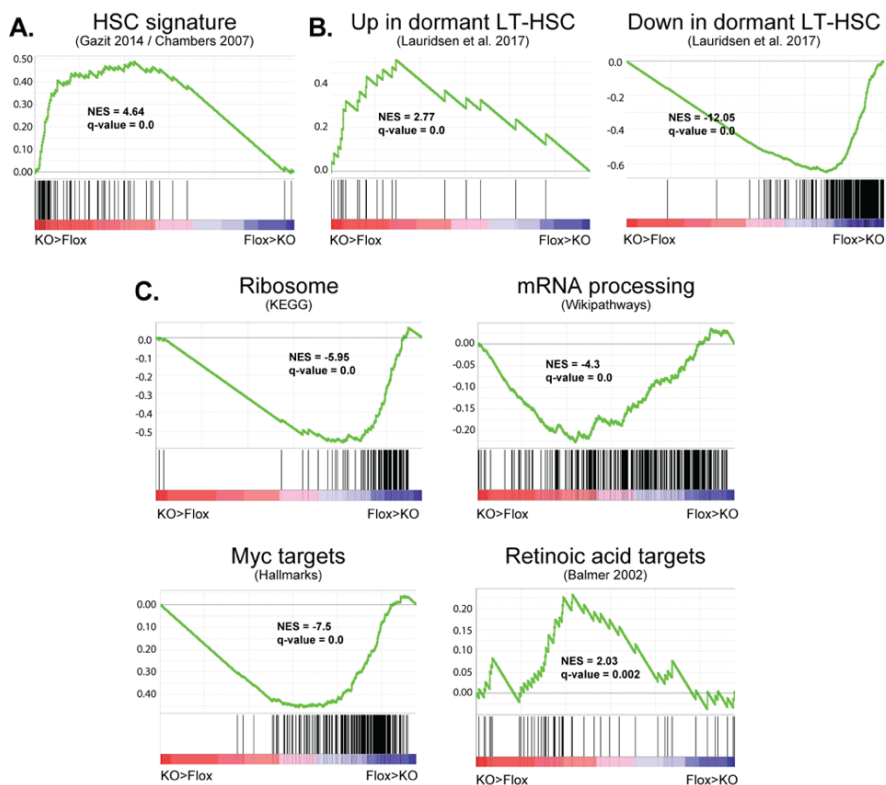


Figure R5. (A) GSEA showing positive enrichment in *Phf19* KO transcriptome for HSC gene set [304, 305]. (B) GSEA showing positive and negative enrichment in *Phf19* KO transcriptome for reported up and down-regulated, respectively, dormant HSC gene signature [297]. (C) GSEA showing negative enrichment for Myc targets, Ribosome and mRNA processing gene sets, and positive enrichment for retinoic acid targets gene set [296].

Importantly, in addition to a high quiescence and HSC identity, dHSC has been associated with a high long-term sustained multipotency and blood production capacity [294-299]. These HSC capacities were in apparent conflict with the phenotype of *Phf19* KO mice since i) *Phf19* KO BM displayed a differentiation bias toward myelopoiesis upon aging, and ii) a decrease contribution to blood production upon serial transplants (see Annexe I Figure1E and 4) [194]. Thus, we speculate that the transcriptome of LT-HSCs in homeostasis did not reflect future HSC response upon mobilization,

such as upon stress induction by serial transplantation assays. In this line, we reasoned that *Phf19* KO LT-HSC may display a higher quiescence in homeostasis similar to dHSCs, but a decrease capacity to sustain haematopoiesis when required.

2.2. *Phf19* KO HSC transcriptome display similarities with aged HSC

Interestingly, multiple studies have established that aged HSC displayed an increased quiescence, but a decreased capacity to undergo activation, which leads to a decrease long-term repopulation capacity [306, 307]. Moreover, aged HSC that succeeded to exit quiescence and trigger differentiation display a subsequent bias toward myeloid lineage differentiation, which is also reminiscent of our *Phf19* KO phenotype [308, 309]. To evaluate whether *Phf19* KO LT-HSCs could be similar to aged HSCs, we generate an aged HSCs gene signature using the genes DE between young (4 weeks) and aged HSCs (24 weeks) reported by Sun et al. [298], and we used it to interrogate our LT-HSC transcriptome data. Remarkably, there was a high correlation between our *Phf19* KO LT-HSC transcriptome and those of aged HSCs (Figure R6). These results suggest that the young *Phf19* KO LT-HSCs transcriptome resemble those from aged HSCs in homeostasis, therefore reinforcing our hypothesis.

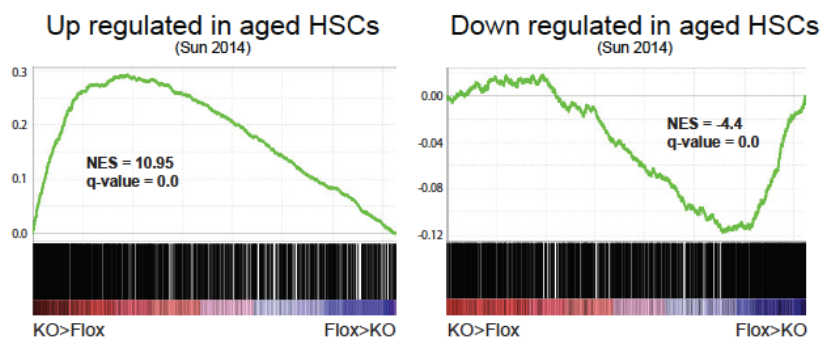


Figure R6. Gene set enrichment analysis (GSEA) showing positive and negative enrichment for reported up and down-regulated genes, respectively, in aged vs young HSCs.

In order to understand the myeloid differentiation bias in *Phf19* KO mice, we decided to analyse whether the LSK population, which have been successfully activated and shown lineage commitment, presented a transcriptional signature for this lineage differentiation. For that, we evaluated the LSK RNA-seq data, and we identified 278 upregulated ($FC > 1.5$) and 418 downregulated ($FC < -1.5$) genes in *Phf19* KO LSKs. Interestingly, GO term enrichment analysis shown several myeloid differentiation categories among our set of upregulated genes (Figure R7A). To further support a potential myeloid bias in *Phf19* KO LSK, we took advantage of the specific common myeloid (CMP) and lymphoid progenitors (CLP) gene signatures employed by Yu et al. to hallmark biased to myeloid or lymphoid differentiation hematopoietic progenitors [310]. Notably, *Phf19* KO LSK cells shown an increase expression of genes associated with CMP, while displaying a concomitant decrease in genes associated with CLP (Figure R7B). These results suggest that PHF19 depletion in hematopoietic progenitors promote a transcriptional bias towards myelopoiesis.

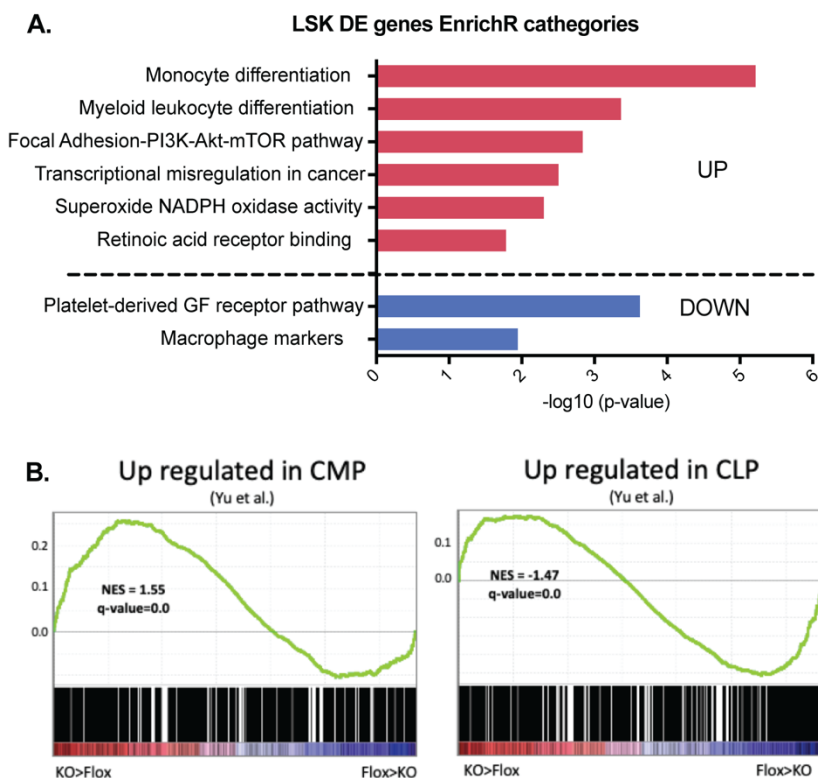


Figure R7. (A) GO term enrichment analysis of DE genes in LSK RNA-seq data using EnrichR. Relevant categories are selected and ranked by p-value. (B) GSEA showing positive enrichment in *Phf19* KO LSK transcriptome of genes up-regulated in HSC clones with a future differentiation toward myeloid lineage, and opposite trend for HSC clones deriving in the lymphoid branch.

Finally, we were also interested in understanding if PHF19 depletion promotes transcriptional changes in young animals that could explain the higher incidence of hematopoietic malignancies in old *Phf19* KO mice. For that, we searched for gene signatures associated with HSC in a pre-malignant or malignant state, and we found a signature associated with leukemic stem cells (LSC). LSC originates from particularly highly quiescent HSCs that progressively accumulate early mutations that enhance their self-renewal potential and, often, impair their differentiation. These LSCs remains latent in

homeostasis until a late mutational event induce exit from quiescence and trigger massive proliferation, full differentiation block, and drive the development of leukemias. Notably, we found an increased expression of this LSC signature in our *Phf19* KO LT-HSCs transcriptome. Based on these results we speculate that PHF19 depletion may promote a highly quiescent HSC population that is likely to accrue a high rate of mutations and develop hematopoietic malignancies during ageing (Figure R8).

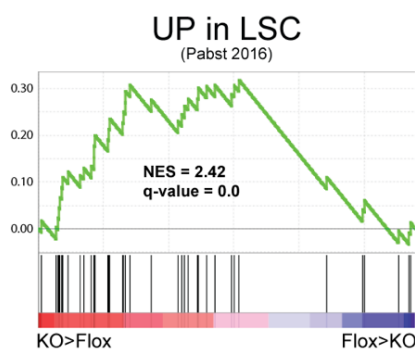


Figure R8. GSEA showing positive enrichment in *Phf19*-KO transcriptome for leukemic stem cell (LSC) differentially up-regulated genes.

Overall, our transcriptome analysis point towards an increase HSC identity and quiescence in *Phf19* KO LT-HSC under homeostatic conditions, which is similar to those of highly quiescent populations as dHSC and aged HSC. In line with an enriched aged HSC transcriptomic identity, LSK transcriptome displayed a bias toward myeloid lineage, and young *Phf19* KO HSCs already mimic the transcriptional state of HSC that have accumulate malignant mutations during lifespan.

3. *Phf19*-KO HSCs present a chromatin accessibility unfavourable to undergo lineage differentiation.

The similarity of *Phf19* KO LT-HSCs transcriptome with both dHSC and aged HSC, suggests that the transcriptome of LT-HSCs in homeostasis do not reflect how LT-HSCs would respond upon stress induction. Indeed, multiple studies demonstrated that the transcriptome is not consistently correlated with HSCs future trajectory, whereas epigenetic features such as histones PTMs and chromatin accessibility demonstrated a more confident association with the subsequent function of specific HSC clones [310, 311]. This suggests that epigenetic memory represents an important factor in guiding the heterogenous HSCs functional decisions. In order to better understand how PHF19 depletion could compromise HSCs function and downstream haematopoiesis, and considering the association of PHF19 with PRC2 activity, we decided to conduct H3K27me3 ChIP-seq and ATAC-seq analyses on HSCs.

3.1. *Phf19*-KO HSCs increase H3K27me3 deposition at HSC repressed genes.

In order to perform H3K27me3 ChIP-seq in a maximum of 10,000 HSC (LSK CD150⁺) from *Phf19* floxed and *Phf19* KO mice we tested several protocols that enable to perform ChIP-seq with a low number of cells, such as ChIPmentation [312]. Finally, we obtain the best quality results using a commercial kit for a low number of cells (Low Cell ChIP-seq kit Active Motif). The library preparation, the sequencing run, and the quality control of the sequencing were

performed by the CRG Genomics facility. The read mapping and counting were performed by Dr. Enrique Blanco.

Our first approach to perform H3K27me3 ChIP-seq shown a remarkably genome-wide enrichment of H3K27me3 deposition in *Phf19* KO HSC. This was a bit confusing, as after the removal of a PRC2.2 cofactor as PHF19, we were expecting a reduction of H3K27me3. To further consolidate this finding, we conducted a second H3K27me3 ChIP replicate, in this case mixing the mouse chromatin with chromatin *D. melanogaster* cells as spike-in, for normalization. The spike-in chromatin enables to monitor the technical variations during the ChIP-seq procedure for a subsequent computational normalization of the data [313]. Notably, after spike-in correction, we still obtained a significant genome-wide increase of H3K27me3 deposition upon PHF19 depletion (Figure R9A).

In order to understand the functionality behind these changes, we crossed the H3K27me3 ChIP-seq with our LT-HSC transcriptome. Interestingly, we did not observed major transcriptomic changes in genes gaining H3K27me3 at the TSS in *Phf19* KO HSC (H3K27me3 target genes), as the majority of those genes were already silenced in *Phf19* floxed HSCs (Figure R9B). Moreover, we divided the subset of H3K27me3 targets by their FC H3K27me3 enrichment in *Phf19* KO over *Phf19* floxed HSCs, and compare the expression of their target genes in the *Phf19* floxed HSCs transcriptome. These results showed that the highest H3K27me3 enrichment (FC>2) was at the lowest expressed genes in normal HSC (Figure R9C).

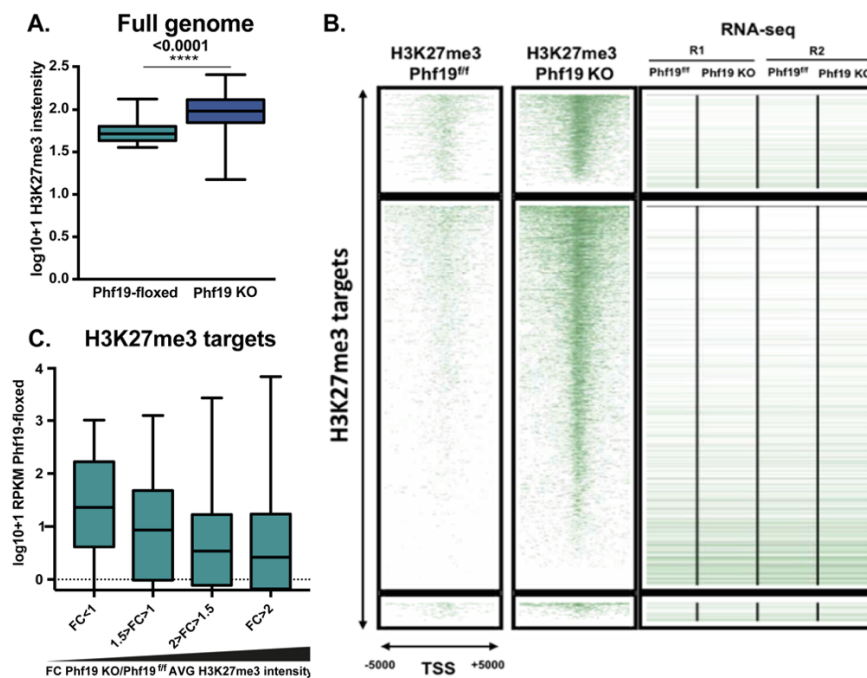


Figure R9. (A) Spike-in corrected ChIP-seq levels of H3K27me3 in the transcription start site (TSS) \pm 2 Kb of the whole genome. (B) Heatmap of all H3K27me3 peaks called and their depicting expression levels in LT-HSC RNA-seq replicates. (C) Gene expression levels in *Phf19* floxed HSC for the different subsets of H3K27me3 peaks based on H3K27me3 enrichment in *Phf19* KO over *Phf19* floxed HSCs.

3.2. *Phf19* KO HSCs increase H3K27me3 at genes required for HSC differentiation, but not HSC-identity.

Based on the evidence that cell fate is determined by specific gene expression programs [93], we reasoned that HSC-fingerprint genes should be active in HSCs, while genes associated to other tissues and to differentiated blood cells might be silenced. In this regard, we speculate that HSC-fingerprint genes may be protected from H3K27me3 deposition upon PHF19 depletion, while genes associated with other cell types may gain H3K27me3. Consistently,

GO term enrichment analysis of genes enriched for H3K27me3 (FC>1.5) displayed categories related to other differentiated tissues, as nervous system (i.e. axon development (p-value<3.3x10⁻⁷), or embryonic heart tube morphogenesis (p-value<5.9x10⁻⁴). In order to focus on the hematopoietic system, we used again the HSC-fingerprint gene set, and we define also from Chambers et al. and Gazit et al. a ‘differentiation signature’, comprising genes specifically expressed in different mature blood cells [304, 305]. In addition, we used a ‘TF signature’ extracted from Yang et al., which contains master regulators of hematopoietic lineage differentiation [314]. In line with our reasoning, we did not observe an increase of H3K27me3 in HSC-fingerprint genes, whereas we observed a significant increase of H3K27me3 upon *Phf19* KO in the differentiation and TF gene signatures (Figure R10).

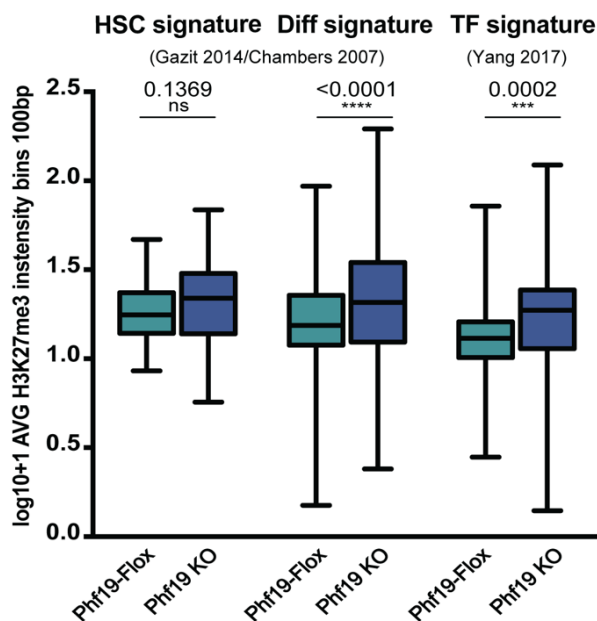


Figure R10. ChIP-seq levels of H3K27me3 in the transcription start site (TSS) ± 2 Kb of specific gene sets: the HSC gene set, the differentiation gene set, and for transcription factors associated with hematopoietic differentiation processes.

Altogether, these results suggest that H3K27me3 enrichment in *Phf19* KO HSC occurs at genes required for HSC differentiation, but not at those required to maintain HSC identity.

3.3. *Phf19* KO HSCs H3K27me3 profile is similar to aged HSCs.

Interestingly, Sun et al. reported that H3K27me3 targets are conserved upon aging, although H3K27me3 levels increase dramatically [298]. To further validate our hypothesis that *Phf19* KO LT-HSC may be closer to aged HSC, we crossed our data with the H3K27me3 ChIP-seqs performed by Sun et al. in HSC isolated from young (4 weeks) and old (24 weeks) mice [298]. Notably, we obtained a very high correlation between the H3K27me3 profile of aged HSC and *Phf19* KO HSC, whereas the H3K27me3 profile from young mice correlates better with *Phf19* floxed (Figure 11).

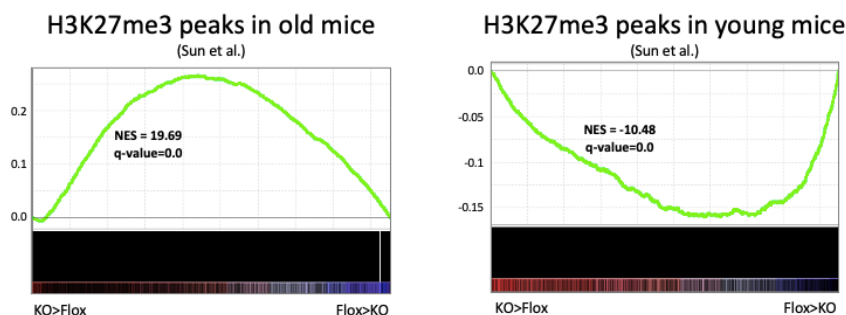


Figure R11. (A) GSEA showing a dramatic increase in *Phf19*-KO H3K27me3 levels in annotated old LT-HSC specific targets, and non-significant H3K27me3 changes in annotated CMP and CLP specific enhancers

These observations highlighted that the H3K27me3 changes occurring in young *Phf19* KO LT-HSC are similar to those occurring upon HSC ageing.

3.4. *Phf19* KO HSCs decrease chromatin accessibility at genes required for HSC differentiation.

Finally, to further support the global H3K27me3 enrichment in HSCs upon PHF19 depletion, and to associate these changes with an increase in facultative heterochromatin, we decided to study the chromatin accessibility profile of *Phf19* floxed and *Phf19* KO HSC. We performed OmniATAC-seq analysis using 10,000 HSCs (LSK CD150⁺) extracted from *Phf19* floxed and *Phf19* KO mice in two independent biological replicates. Consistent with the global increase in H3K27me3 deposition, we identified a significant genome-wide decrease in chromatin accessibility upon PHF19 depletion (Figure R12A). In addition, we analysed the ATAC-seq levels at the previously used gene sets. We observed a tendency towards a decrease in chromatin accessibility at genes essential for differentiation, but not at HSC specific genes. However, we only obtained statistically significant differences for one of our replicates (Figure R12B). Interestingly, we observed an increased H3K27me3 deposition and correlated decrease chromatin accessibility at essential lineage-specific TF genes, such as *Cebpa* (myeloid); *Pax5* and *Irf5* (B cells); *Tcf7* (T cells); members of Notch pathway as *Dll1* (T and B cells); and *Epo* (erythroid cells). These results show that PHF19 depletion gives rise to a global decrease in chromatin accessibility, and that this decrease is exacerbated at genes required for differentiation, but not for HSC identity.

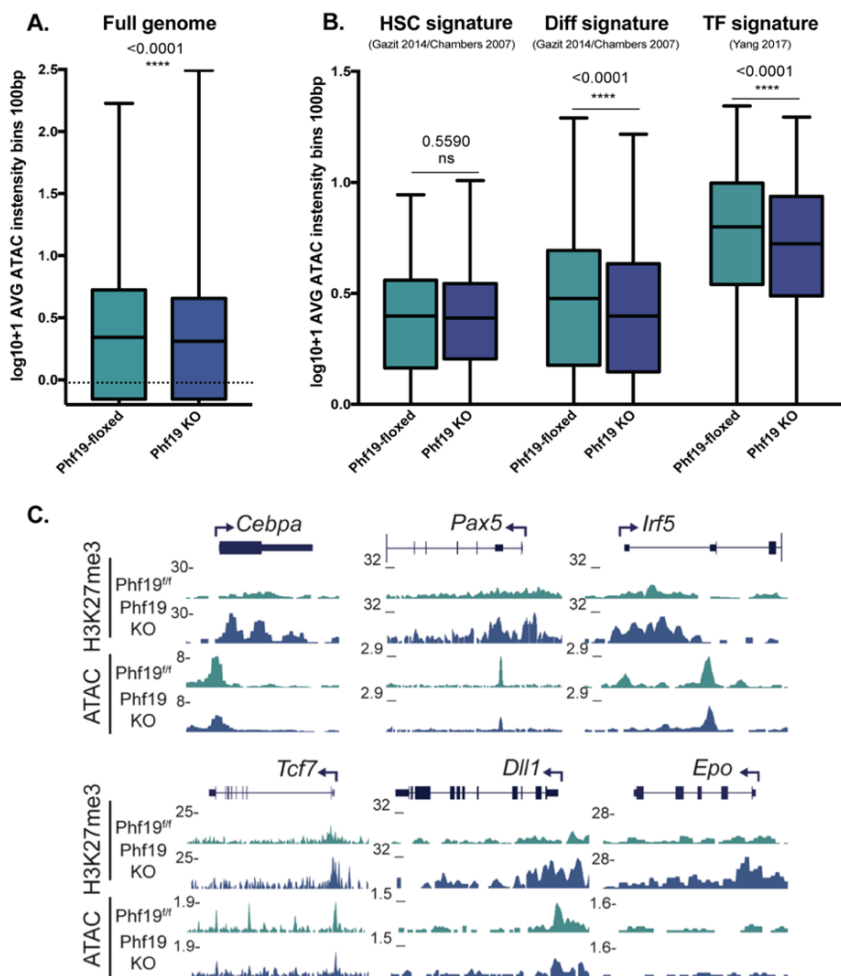


Figure R12. (A) ATAC-seq levels in the transcription start site (TSS) \pm 2 Kb of whole genome. (B) ATAC-seq levels in the transcription start site (TSS) \pm 2 Kb of specific gene sets associated with hematopoietic differentiation processes. (C) UCSC genome browser screenshots for H3K27me3 and ATAC of haematopoiesis master regulators.

Overall, our results suggest that *Phf19* KO HSC present an aberrant HSC chromatin structure in homeostasis, which might promote HSC identity and impair the dynamic chromatin reorganization responsible for lineage commitment. In addition, in line with our transcriptomic results, this chromatin structure is similar to the one of aged HSC.

4. *Phf19* KO HSC quit quiescence upon stress but display impaired lineage differentiation and a bias towards HSC self-renewal.

The phenotypic characterization, our transcriptome, and our chromatin profile results, suggested that the PHF19 depleted HSC compartment is dominated by a HSC clonal population i) with an increased HSC identity and quiescence, ii) with a chromatin structure that is unfavourable to allow differentiation, and ii) with high similarity with aged HSCs.

In order to test this working hypothesis, we aimed to elucidate if this population can exit quiescence and to trigger differentiation when required. Thus, in collaboration with Dr. Pedro Vizán, Arantxa Gutierrez, and Dr. Marc García from our lab, we sorted single *Phf19* floxed and *Phf19* KO LT-HSC on 96-well plates and we cultured them under conditions that promote HSC division and differentiation (Figure R13A). We then monitored and compared the number of cells and surface markers at different time points. First, during four days after sorting, we scored each well according to its division kinetics; as: i) proliferating (i.e., sustained cell growth over the 4 days), ii) nonproliferating (i.e., not able to divide more than twice during 4 days), and, iii) proliferating and stop (i.e., division that eventually stopped or reduced the number of cells). Under our culture conditions, most wells were scored as proliferating, both in *Phf19* floxed and *Phf19* KO plates, but a significant increase in the percentage of proliferating colonies in *Phf19* KO wells was observed

(Figure R13B). We maintained the colonies in culture for additional five days and then calculated the number of cells per well by fluorescence cytometer after propidium iodide (PI) staining (Figure R13C). Notably, not only *Phf19* KO wells had more proliferating colonies, but also these colonies individually contained a higher number of cells. On the other side, we measured the expression of lineage and HSCs markers after 2 weeks. Interestingly, we observed impairment of differentiation in the *Phf19* KO condition indicated by the increase number of Lin⁻ cells (Figure R13D). Altogether, these results demonstrate that more *Phf19* KO LT-HSC can exit quiescence upon force HSC activation, also displaying a higher proliferation capacity, but that these *Phf19* KO LT-HSC may present an impaired differentiation. This could be explained by a bias toward self-renewal, or by a slowdown or a blockage in the course of the differentiation pathway.

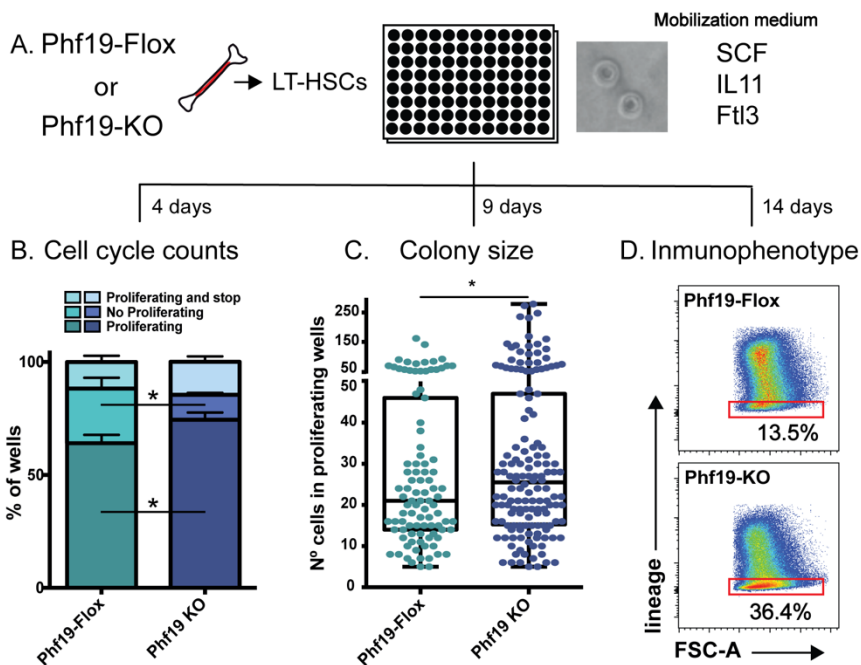
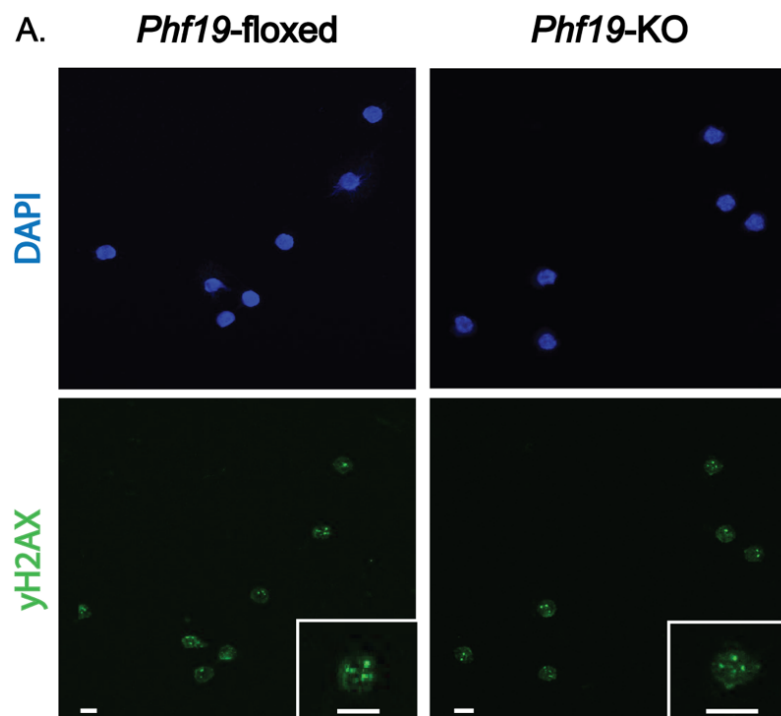


Figure R13. (A) Scheme a single-cell HSC *in vitro* functional assay. (B) Proportion of wells of proliferating, non-proliferating, and proliferating and then stopped HSCs in three independent experiments (mean + S.E.M.). (C) Number of cells per well in all wells of proliferating HSCs is shown from three independent replicates. (D) Percentage of lineage-negative and lineage-positive cells from three independent replicates (mean + S.E.M.) and gating strategy of a lineage-negative marker of pooled wells after 14 days in culture.

Multiple studies have provided evidence that quiescent HSCs present a decrease in DNA damage response and repair pathways, leading to a progressive accumulation of DNA strand breaks upon aging and in LSCs, and providing a potential explanation for the increase incidence of hematopoietic malignancies in aged mice [309, 315, 316]. To further support a potential increase of *Phf19* KO LT-HSC quiescence, we measured the levels of DNA damage in HSCs. For that, we performed γ H2AX immunostaining and alkaline comet assay to measure single- and double-strand breaks in purified HSCs from *Phf19* floxed and *Phf19* KO mice. The quantification of the fluorescence signal was performed by the CRG Advanced Light Microscopy unit. Intriguingly, we recurrently observed high levels of DNA damage (>3 foci average) in both conditions, without major differences between *Phf19* floxed and *Phf19* KO HSCs (Figure R14). This was unexpected, as young HSC have been previously reported to display less than 5% of cells with more than 3 foci [316]. Hence, our results could be due to HSCs stress upon sorting and extensive manipulation. In this line, comet assay results also demonstrated high levels of DNA damage in both *Phf19* floxed and *Phf19* KO HSCs (data not shown). Thus, we cannot validate our hypothesis about an increase in DNA damage in young *Phf19*-KO HSCs.

**B.**

Replicate	N ^o foci per cell	Average size	Average intensity
<i>Phf19</i> -floxed R1	3,71	0.655	166.920
<i>Phf19</i> -floxed R2	4	0.738	188.227
<i>Phf19</i> -floxed R3	3,4	0.705	179.674
<i>Phf19</i> KO R1	4,46	0.784	199.876
<i>Phf19</i> KO R2	4,16	0.630	160.624
<i>Phf19</i> KO R3	4,11	0.752	191.880

Figure R14. (A) Representative images of DAPI stained nuclei (top) and γ H2AX stained foci (bottom) in *Phf19* floxed and *Phf19* KO LT-HSCs. (B) Quantification of γ H2AX foci in *Phf19* floxed and *Phf19* KO LT-HSCs as an absolute number of γ H2AX foci per cell, average size per foci, and average fluorescent signal intensity per foci.

5. HPC-7 cell line is not a good alternative to study PHF19 mechanism in LT-HSCs.

To understand how PHF19 lead to these transcriptional changes, to the genome-wide increase of H3K27me3 and, even more importantly, to a differential H3K27me3 enrichment between HSC and differentiation gene sets, we aimed to perform ChIP-seq analysis of PHF19 and other PRC2 subunits in HSCs. However, this was not feasible due to the cell input limitations, which rendered it impossible to perform lowly expressed TF ChIP-seq with less than millions of cells. We thus decided to use an alternative *in vitro* model that allows us to obtain high numbers of cells. HPC-7 are an immortalized mouse hematopoietic stem cell line which has been recurrently used in the literature as an alternative to mouse isolated HSCs [317]. In order to test, whether we can use HPC-7 *in vitro* as a good model to evaluate the function of PHF19, we performed *Phf19* knockdown. For that, we infected HPC-7 cells either with a control empty shRNA pLKO.1 or with an anti-*Phf19* shRNA pLKO.1 vector. Next, after puromycin selection and validation of *Phf19* KD by RT-qPCR, we performed a comparative analysis in HPC-7 similar to the one that we performed for LT-HSCs *in vivo* so that we characterized the effect of *Phf19* KD over HPC-7 phenotype, transcriptome, and H3K27me3 profile. In addition, we performed PHF19 and MTF12 ChIP-seq.

Importantly, our results demonstrated major differences between HPC-7 cells and LT-HSCs, which suggest that HPC-7 is unlikely to be a good model of LT-HSCs. For instance, under control conditions, empty HPC-7 displayed a higher expression of differentiation genes

than HSC fingerprint genes, which suggest that HPC-7 is a more differentiated hematopoietic population than LT-HSCs (Figure R15A). In this line, upon PHF19 depletion, several results pointed toward an increase in HPC-7 differentiation, as a downregulation of HSC-fingerprint genes and an up-regulation of differentiation genes, and a decrease proliferation together with an increase of multi-lobed nuclei, a phenotype associated with differentiated blood cells (Figure 15B). As a consequence, we decided to do not to go further with our HPC-7 characterization, as we could not extrapolate our conclusions to our *in vivo* observations.

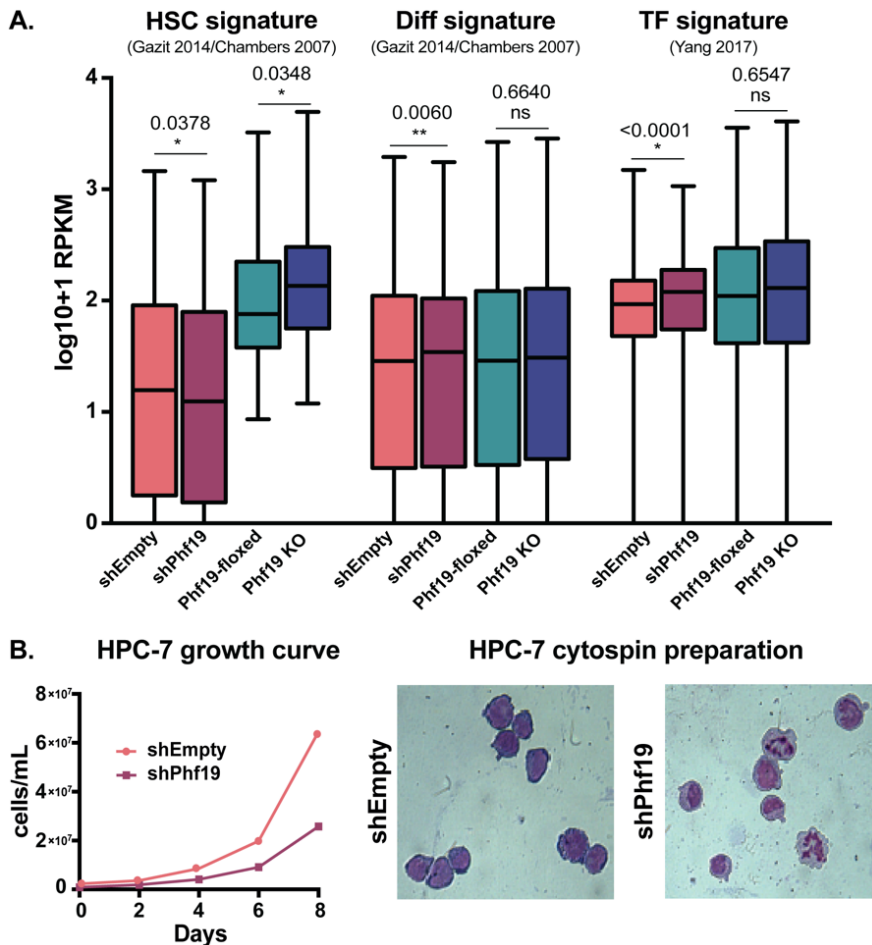


Figure R15. (A) RPKM values for HPC-7 (warm) and LT-HSC (cold) of specific gene sets the HSC-fingerprint gene set, the differentiation gene set, and the transcription factors required for hematopoietic differentiation. (B) Growth curve of HPC-7 cells over 8 days in Empty and PHF19 depleted conditions. (C) Cytopsin-stained images showing distinctive changes in both cytoplasmic and nuclear morphologies in Empty and PHF19 depleted HPC-7. Morphology changes such as myeloid (large cells with segmented nuclei or single nuclei at centre), erythroid (small cells with eccentric nucleus, poly-chromatophilic cytoplasm), and megakaryocytic (large cells with multi-segmented nucleus) are observed upon PHF19 depletion.

Nevertheless, although our data in HPC-7 would require further validation and we decided to do not to go deep into our analysis, it is worthy to note that we reported again a global increase of H3K27me3 (Figure R16A) and a higher H3K27me3 enrichment among lowly expressed genes in control conditions (Figure R16B). In addition, we succeed to perform PHF19 and MTF2 ChIP-seq and we observed that they share multiple targets in HPC-7 cells (Figure R16C). Interestingly, PHF19 depletion appears to increase MTF2 occupancy at PHF19 targets which, if further validated, suggest that they may compete and counteract to each other on their binding PRC2 and chromatin (Figure R16D-E).

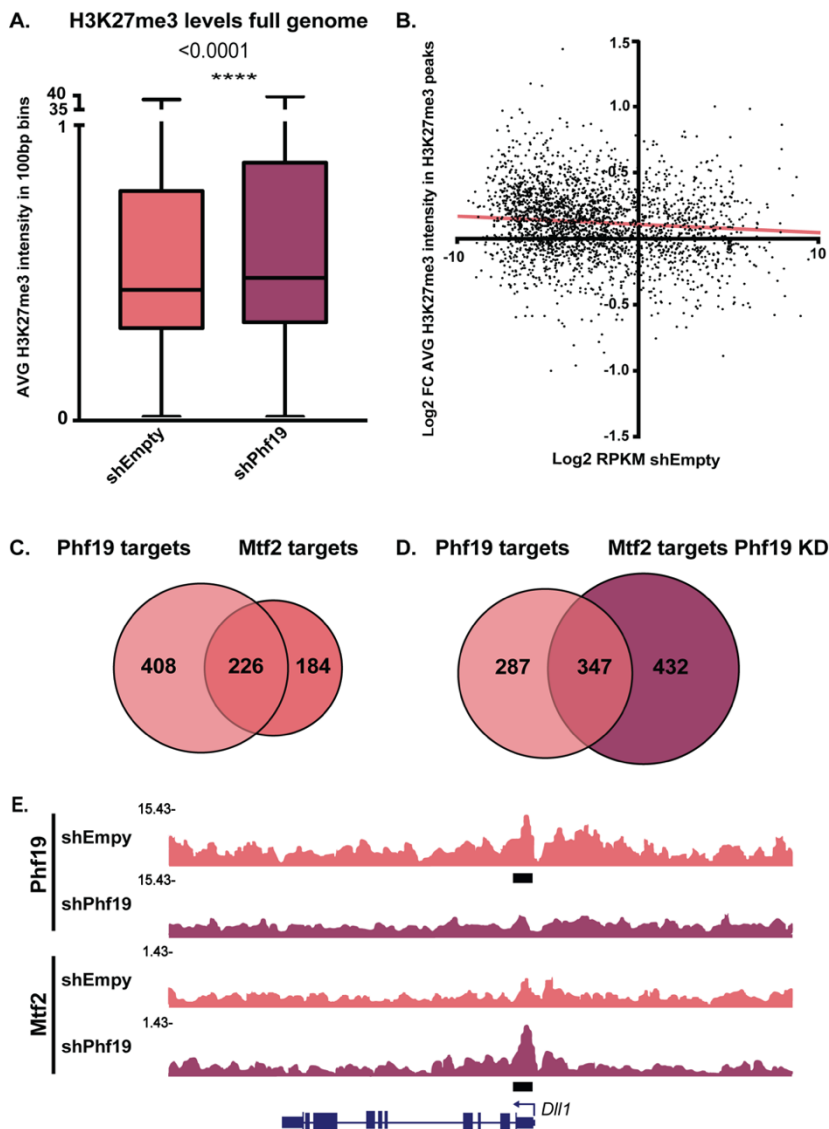


Figure R16. (A) H3K27me3 average ChIP-seq signal in all 100 bp bins of the mouse genome. (B) Every dot represents a target gene with an H3K27me3 peak in the region 2.5 Kbp upstream of their TSS. The x-axis indicated the log 2 of the RPKM values in shEmpty HPC-7 RNA-seq data, and the y-axis indicates the log2 fold change of the average ChIP-seq signal in shPhf19 over shEmpty H3K27me3 ChIP-seq data. (C) Venn diagram showing the intersection between Phf19 and Mtf2 targets in control HPC-7. (D) Venn diagram showing the intersection between Phf19 targets in control HPC-7 and Mtf2 targets in Phf19-KD HPC-7. (E) UCSC genome browser screenshots for Phf19 and Mtf2 ChIP-seq profiles in shEmpty and shPhf19 HPC-7

CHAPTER II

**Identification of acquired epigenetic vulnerabilities in
DIPGs for selective targeting of tumoral cells.**

1. The relevance of chromatin and epigenetics in driving Diffuse Intrinsic Pontine Gliomas (DIPG).

Diffuse Intrinsic Pontine Gliomas (DIPG) are the most aggressive and unconditional fatal type of paediatric glioblastoma. DIPG correspond to the 10-20% of childhood tumours, with an incidence of 2 cases per 100.000 population [318]. The average age at the time of diagnosis is 6.5 years old and the median overall survival is 10.8 months (Figure R17A) [275]. The tumours specifically arise from the ventral pons and midline area of the brain stem but early display a highly devastating diffuse and invasive spread throughout the nervous system. This diffuse nature hamper DIPG treatment, as render impossible a tumour surgical resection and ineffective a focal radiation therapy to the pons, which remains standard of care [319].

As mentioned on the Introduction, a highly recurrent H3K27M mutation in genes encoding either H3.1 or H3.3 is found in 80% of all DIPGs, [320]. Beyond H3K27M, DIPGs have shown to harbour few large-scale chromosomal alterations and somatic driver mutations, which also present a lower incidence than H3K27M [274]. The most common somatic driver mutations involve loss of TP53, loss of ACVR1 (particularly associated with H3.1K27M), amplification of PDGFRA, MYC, MYCN and CDKN2A/B, and amplification or activating mutations in the MERK/ERK and PI3K/Akt pathways (Figure R17B).

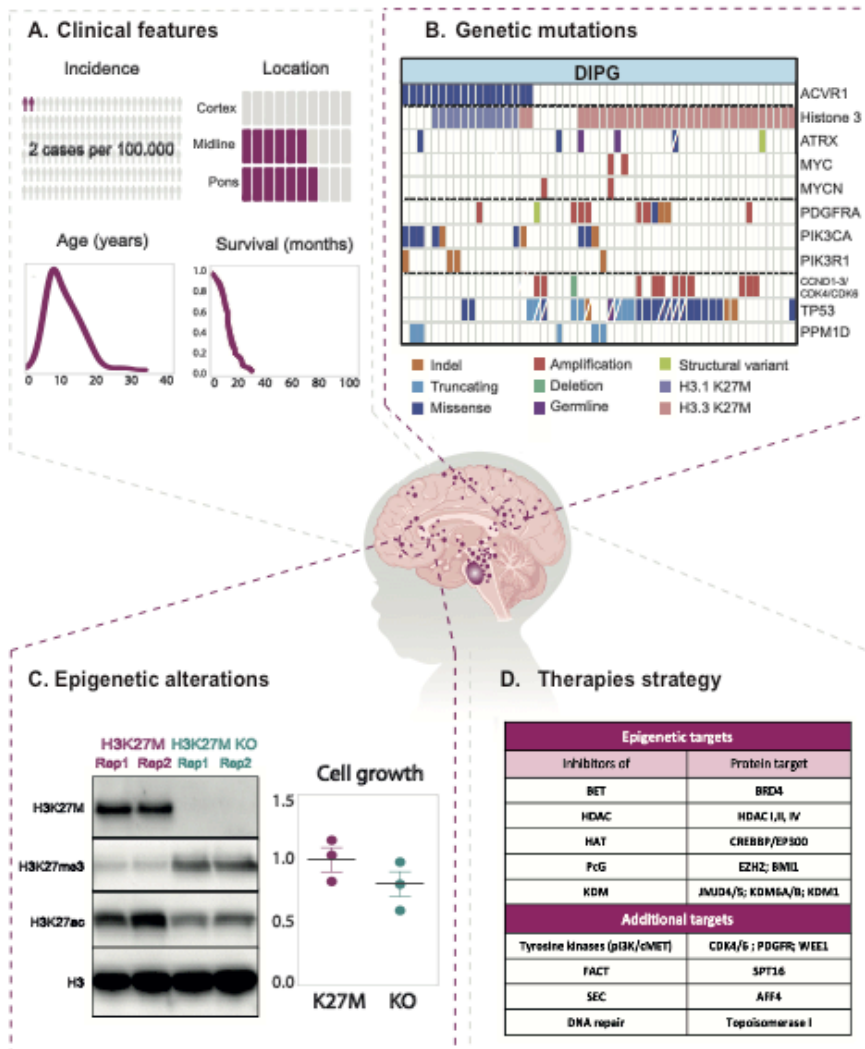


Figure R17. (A) Summary of the main DIPG distinctive clinical features. (B) Recurrent genetic alterations in DIPG. Genetic alterations detected in 11 genes are displayed according to the colour key shown below (adapted from [274]). (C) Western blot depicting H3K27M-dependent alterations in H3K27me3 and H3K27ac. Decrease proliferation of H3K27M-KO over H3K27M DIPG cells (extracted from [286]). (D). Summary of the main identified epigenetic and biological targets in DIPG cell line models.

However, somatic mutations have demonstrated insufficient in triggering DIPG *in vivo*, since DIPG initiation depends on H3K27M induction and TP53 depletion within a **specific cell context** [321-

323]. Therefore, *in vitro* H3.3K27M induction and TP53 KD in neuronal stem cells (NSCs) give rise to DIPGs tumours in xenograft models, but the same study fails to develop DIPG mice tumours from oligodendrocyte precursors (OPCs) carrying H3.3K27M and TP53 KD [284]. Similarly, *in vivo* H3.3K27M need to occur during a specific window of mouse development to give rise to tumours recapitulating human DIPG [324].

In line with the low prevalence of somatic alterations and the influence of cell fate in DIPG initiation, PRC2 misregulation has demonstrated to be essential for DIPG oncogenic transformation and maintenance, which definitely highlight the pivotal role of **chromatin** and **epigenetics** as driving forces in DIPGs [277, 285, 286]. Importantly, H3K27M KO has shown to revert H3K27me3 and H3K27ac distribution similar to non-pathological levels, which in turn induce decrease DIPG cells proliferation (Figure R17C) [286]. Therefore, several efforts have been directed to identify potential epigenetics targets for DIPG based on both, i) the relevance of epigenetic alterations to sustain DIPG tumorigenicity, and ii) the intrinsic reversible nature of these epigenetic alterations (Figure R17D). However, to date, no therapy has been proved to be effective. Consequently, the aim of this second chapter was to identify acquired epigenetic vulnerabilities in DIPGs for selective targeting of tumoral cells.

2. Identification of epigenetic vulnerabilities in DIPG primary cell lines for selective targeting of tumoral cells.

In order to identify chromatin-modifying factors as potential therapeutic targets in DIPGs, we collaborated with Dr Jaume Mora and Dr Ángel Carcaboso from Sant Joan de Deu Children's Hospital (HSJD) in Barcelona, who has been pioneers in the establishment of DIPG cell lines from patients biopsies. They provided us with a panel of 7 different patient established DIPG cell lines, all positive for H3.3 K27M mutations, displaying a diffuse spread phenotype, and biopsied from the pons (Table MM1). In addition, for comparative analysis, they also provided us with 2 cell lines established from different pediatric high-grade glioblastoma (pHGG) patients, that do not harbour H3K27M mutations and were obtained from well-defined hemispheric brain tumours (Figure R18A).

With this worldwide unique collection of DIPGs models, we aimed to perform a CRISPR drop-out screening in DIPG and pHGG cell lines using **ARUNA**, a **chromatin-focused sgRNAs library** developed in our laboratory. Simultaneously, we aimed to obtain a list of **chromatin-bound proteins** in DIPG and pHGG cell lines using **iPOTD**, a novel technology for mass-spectrometry (MS)-based quantitative chromatome profiling also developed in our laboratory. We aimed to combine the results of both approaches with the criteria that proteins targeted by sgRNAs depleted in DIPG cell lines but not pHGG, and also enriched in DIPG chromatin over pHGG, may

represent chromatin regulators essential for DIPG cell viability, and would serve to create a list of potential therapeutic targets (Figure R18).

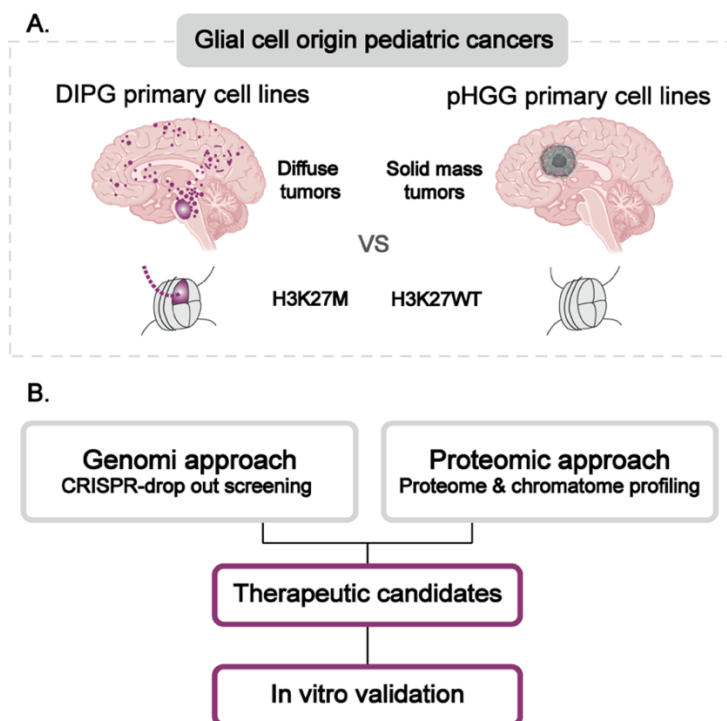


Figure R18. (A) Schematic representation of the main features differing between our models. (B) Schematic representation of the major steps involved in our strategy to identify new therapies for DIPG.

Based on the expertise of the people from HSJD, we decided to use HSJD-DIPG007 and HSJD-GBM001 cell lines as the most workable models to serve our strategy. We decided to compare DIPG with an H3K27WT paediatric glioblastoma in order to identify differences associated with the particular oncogenic properties of DIPG, rather than with the healthy tissue of origin. Finally, after the identification of therapeutic protein targets, we aimed to the extent the validation of these targets to all HSJD patient cell lines in order to draw joint conclusions that can be representative of DIPG disease.

3. CRISPR screen of 652 chromatin regulator factors to identify potential candidates that compromise DIPG cell proliferation and/or survival.

In order to perform our genomic approach, we need i) to generate HSJD-DIPG007 and HSJD-GB001 cultures harbouring an optimal ARUNA library representation, ii) to express Cas9, and iii) to extract genetic material to perform sgRNA unbiased qPCR amplification and sequencing, to ultimately detect changes in sgRNA representation during culture expansion (Figure R19A).

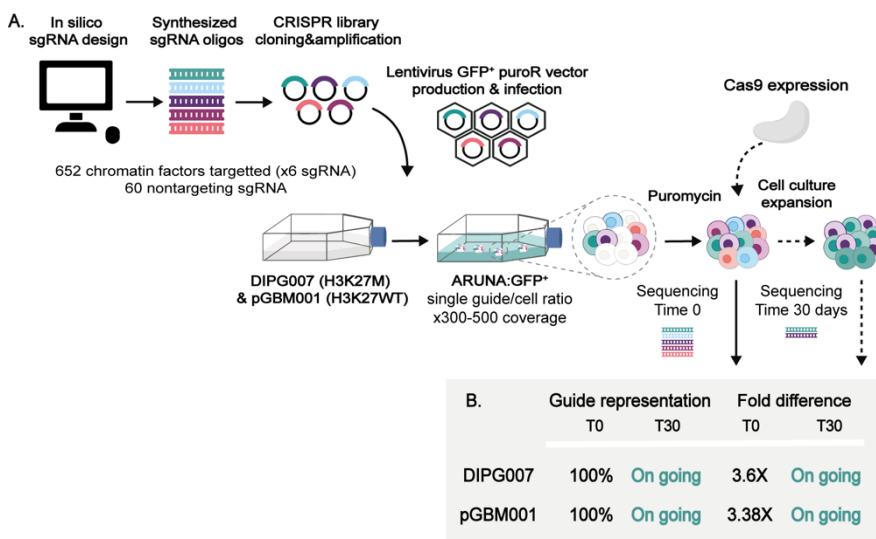


Figure R19. (A) Schematic representation of the major steps involved in our genomic strategy. (B) Summary of the sgRNA sequencing results from ARUNA⁺ cell lines prior Cas9 expression (time 0).

To generate HSJD-DIPG007 and HSJD-GBM001 ARUNA cultures we went through several optimization steps. The ARUNA library is built on a lentiviral vector with a green fluorescent protein (GFP) reporter and puromycin resistance (Figure R20A), and it contains

3937 different sgRNAs targeting 652 chromatin regulator factors and 60 non-targeted control sgRNAs. First, in order to associate an effect over cell viability with a specific targeted protein, we need to ensure that the majority of cells contains only one sgRNA. This is achieved by using a multiplicity of infection (MOI) lower than 0.3 [325], which means 15-30% of infected cells. To titrate our viral infection we take advantage of the GFP reporter of ARUNA, so we determined by flow cytometry the transduction conditions to obtain a 15-30% of GFP+ cells. Next, to minimize the impact of potential biases introduced while expanding the cultures or amplifying the sgRNA cassette, we aimed to obtain at least an initial representation of 300-500 cells per each sgRNAs, which in the case of ARUNA library means 2 million cells (3997 ARUNA sgRNAs x 500 cells). Hence, we finally transduced with a MOI of ~0.15-0.3% 12 millions of HSJD-DIPG007 and HSJD-GBM001 cells.

We next focused on generating homogenous HSJD-DIPG007 and HSJD-GBM001 ARUNA+ cultures. To select ARUNA+ cells we tried again to take advantage of the GFP reporter to select transduced cells using Fluorescence activated cells sorting (FACS). However, despite multiple optimizations to the procedure, HSJD-DIPG007 cells ended up dying in culture a few days after FACS. Finally, HSJD-DIPG007 and HSJD-GBM001 ARUNA+ cells were selected using puromycin, which allow us to obtain >90% of ARUNA GFP+ cells (Figure R20B).

To verify that we did not introduce significant sgRNAs bias during the generation of our ARUNA+ cultures or performing qPCR

amplification of the sgRNA cassette, we sequenced the amplified genomic material from 2 million HSJD-DIPG007 and HSJD-GBM001 ARUNA+ selected cultures (time 0). Notably, we obtained a total representation of the in silico designed sgRNAs and an optimal library uniformity, as we reported less than 10-fold difference between the 10% least and most abundant sgRNAs reads (Figure R19B) [326].

After successful generation of HSJD-DIPG-007 and HSJD-GBM-001 ARUNA+ cultures, we needed to express Cas9 protein in order to start our screening. However, after many attempts, achieving optimal levels of Cas9 protein expression in HSJD-DIPG007 cells demonstrated to be really challenging. A summary of the strategies followed and the problems encountered are shown in Table R1.

Strategy	Cas 9 expression	Expression system	Delivery strategy	Efficiency issues
Cas9:mCherry fusion protein	Protein transduction	Recombinant Cas9:mCherry protein	Lipofection (2000, 3000, lipostem)	>90% mCherry+ cells by fluorescence cytometry No signal of Cas9 activity mCherry signal at the cellular membrane by microscopy
Cas9:mCherry PX458 plasmid	Constitute expression of Cas9 driven by CBh promoter	Non-lentiviral plasmid	Lipofection (2000, 3000, lipostem) Nucleofection (Amaxa, 4D)	Low transfection efficiency (< 0.1% mCherry+ cells) Low transfection efficiency (< 50% mCherry+ cells)
Doxycycline iCas9:mCherry lentiviral vector	Inducible expression of cas9 driven by the minimal CMV promoter	Lentiviral vector	Lentiviral transduction	Cell death after sorting Cell death after high infection Uncomplete neomycin selection (>70%) Progressive promoter silencing
Constitutive Cas9:BFP lentiviral vector	Constitutive expression of Cas9:BFP driven by EF1 α promoter	Lentiviral vector	Lentiviral transduction	Uncomplete blasticidin selection (>70%)

Table R1. Summary of the different Cas9 expression strategies.

To minimize off-target effects, we first aimed to transiently express Cas9 protein. We began by transfecting a recombinant Cas9:mCherry fusion protein and optimizing different lipofectamine conditions until we reported >90% of mCherry+ HSJD-DIPG007 and HSJD-GBM001 cells using flow cytometry. Next, to evaluate the Cas9 efficiency, we co-transfected ARUNA+ cells with mCherry:Cas9 protein and several sgRNAs against the GFP reported of ARUNA. If the introduced mCherry:Cas9 is effective, we would expect a reduction in GFP intensity due to selective targeting of GFP. However, the % of ARUNA GFP+ cells remained unaltered and, as we suspected, fluorescence microscopy localized at the cellular membrane the mCherry fluorescent signal. Thus, based on the high

efficiency of viral transduction obtained with ARUNA lentiviral particles, we next decided to generate a doxycycline-inducible Cas9:mCherry lentiviral plasmid. However, we did not obtain successful induction of Cas9:mCherry expression using doxycycline in the resistant clones.

Then, we finally decided to express Cas9 fused to a blue fluorescent protein (BFP) constitutively, using a strong promoter (EF1alpha). Using this strategy, we obtained >70% of Cas9-BFP expressing cells for HSJD-DIPG007 and HSJD-GBM001 cultures (Figure R20B, left panel). Despite the presence of ~30% of non-expressing Cas9-BFP cells, and after all the initial unsuccessful attempts to obtain homogenous Cas9-expressing glioblastoma cultures, we decided to continue the functional CRISPR screening using the Cas9-BFP >70% expressing cultures.

For that, we initially transduced 24 millions HSJD-DIPG007 and HSJD-GBM001 Cas9-BFP expressing cultures with ARUNA library (MOI of ~0.15-0.3) (Figure 20B, middle panel). After puromycin selection, we obtained >90% ARUNA+ cells (Figure 20B, right panel), we collected 5 millions cells (time 0), which represents 1,250X coverage of our ARUNA library. Additionally, we expand our cultures by collecting cellular pellets every passage up to one month, in order to extract the genomic material to amplify the sgRNA cassette for subsequent sequencing.

A. ARUNA lentiviral vector



Cas9:BFP lentiviral vector



B. HSJD-DIPG007 First CRISPR screening

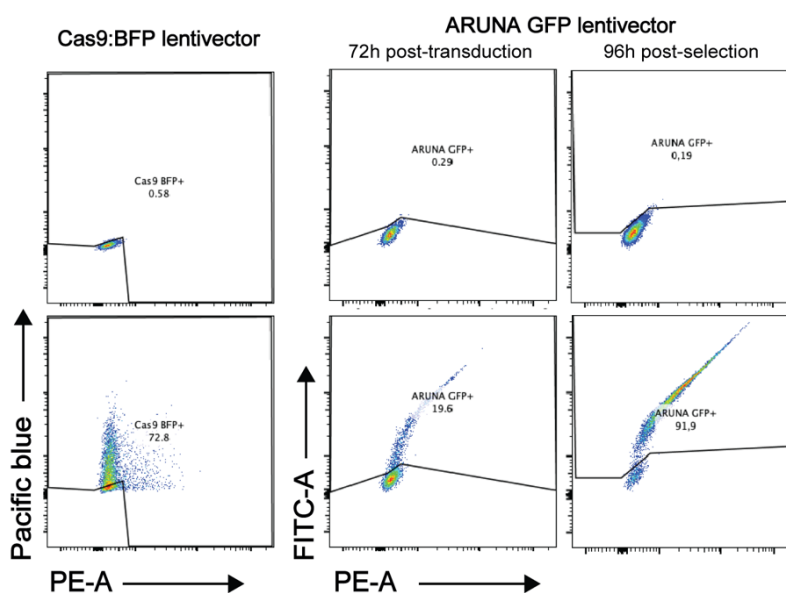


Figure R20. (A) Scheme of ARUNA sgRNA library lentiviral vector. (B) Scheme of the constitutive Cas9:BFP lentiviral vector. (C, Right panel) Flow cytometry analysis of the percentage of Cas9:BFP⁺ cells in control (top) and selected (bottom) HSJD-DIPG007 Cas9/ARUNA⁺ cells. (C, Left panel) Flow cytometry analysis of the percentage of ARUNA:GFP⁺ cells in control (top), after transduction with a MOI ~0.3 (bottom left), and after puromycin selection (bottom right) in HSJD-DIPG007 Cas9/ARUNA⁺ cells.

Unfortunately, during the process of writing this thesis, we are performing the genetic screening and, therefore, we could not include the results in this section.

4. Successful generation of the first proteome and chromatome data from DIPGs cell lines.

Considering that chromatin deregulation is a major driver for DIPG progression, we hypothesize that the enrichment of specific proteins in the chromatin of DIPG might sustain their tumorigenicity and might be considered potential vulnerabilities for pharmacological targeting. In order to identify the DIPG-enriched chromatin bound proteins that could serve us to select potential therapeutic targets, we first aimed to characterize the chromatin-bound proteome in HSJD-DIPG007 and HSJD-GBM001 cells.

We took advantage of iPOTD, a method recently published by our laboratory to profile the chromatome in a global and unbiased manner [91]. As mentioned on the Introduction, iPOTD is based on both, the DNA incorporation of EdU molecules directly from the medium and during DNA replication, and the ‘click chemistry’ to covalently link biotin to EdU-labeled DNA. By doing so, iPOTD relies on biotin streptavidin affinity to capture cross-linked sheared chromatin into streptavidin beads, which ultimately enable to specifically elute chromatin-bound proteins upon reversed chromatin crosslink. Thus, in combination with high-resolution liquid-chromatography-tandem mass spectrometry (LC-MS/MS), iPOTD provides a highly specific and sensitive characterization of the complete cellular chromatome (Figure R21A).

In order to perform iPOTD in HSJD-DIPG007 and HSJD-GBM001 we first evaluated the doubling time of our glioblastoma cultures. To

perform iPOTD in an unbiased manner, EdU molecules are expected to be represented across the whole genome. For that, we aimed to perform the continuous incubation with EdU for a period of at least two rounds of cell division. Therefore, in order to evaluate the doubling time of DIPG and GBM cells, we used a tracing fluorescent reagent (CellTrace CFSE Cell Proliferation Kit) that, after a short pulse, is incorporated into the cells and passively diluted during cell division. The dilution of the dye can be monitored by flow cytometry at different time points, thus enabling to quantifying the doubling time of your culture. By using this strategy, we identified 48h as the optimal incubation time, for which more than 90% of HSJD-DIPG-007 and HSJD-GBM-001 cells may undergo at least one cell division (Figure R21B).

Next, to evaluate the potential toxicity of EdU incubation during 48h, we used Trypan Blue dye exclusion assay to measure the number of viable cells incubated with or without 0.1- μ M EdU up to 72h. We could not observe any differences in cell viability after growing HSJD-DIPG-007 and HSJD-GBM-001 for 72h in a medium supplemented with 0.1- μ M EdU (Figure R21C).

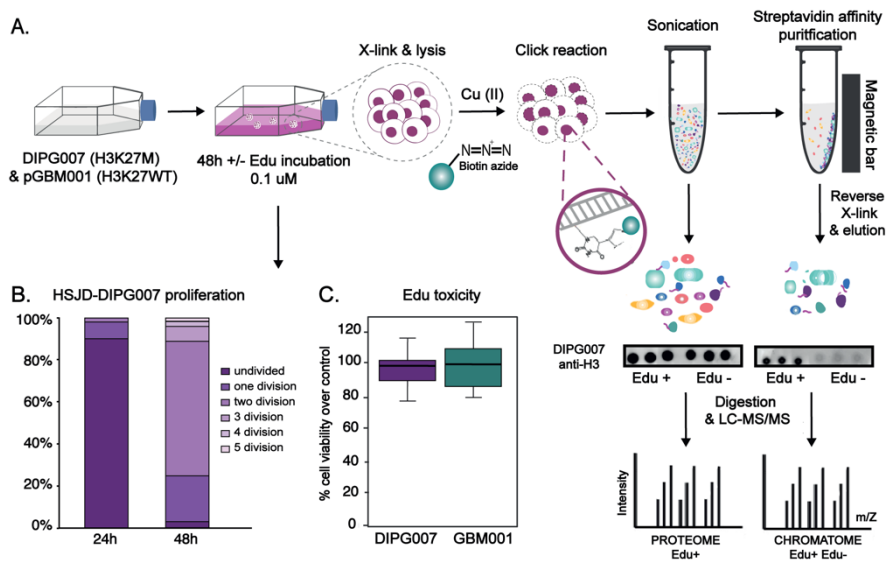


Figure R21. (A) Schematic representation of the iPOTD method for chromatin-bound proteome profiling. Protein samples eluted prior and post streptavidin affinity purification were analysed by dot blot assay using an anti-histone H3 (chromatin marker) to evaluate EdU-labelled chromatin capture efficiency and specificity. The intensity of H3 signal was also used to normalize the amount of chromatin-to-be-analysed between replicates and cell lines. (B) The percentage of HSJD-DIPG007 cells at each cell division was assessed by measuring the level of CFSE fluorescence by FACs. (C) Percentage of viable trypan blue-stained cells comparing to control after 48h incubation with 0.1- μ M EdU.

After the evaluation of the doubling time and the toxicity of EdU, we perform our iPOTD strategy using 5 and 3 independent replicates for HSJD-DIPG007 and HSJD-GBM001, respectively. We expanded in parallel cells treated with or without 0.1- μ M EdU during 48h and we perform EdU-labelled chromatin purification using iPOTD for both conditions. We used non-treated samples as a negative control to identify potential protein contaminants among our EdU-labelled chromatin purified fraction. The proteomic composition of the input material (whole proteome) from EdU-treated samples and the chromatin eluates from EdU-treated and non-treated samples were analyzed by quantitative proteomics. The sample preparation, the mass spectrometry run, the identification of proteins, and the initial

quantification for each sample, was performed by the UPF/CRG Proteomics Unit.

We identified 6386 and 6788 proteins present in the whole proteome at least in two replicas for HSJD-DIPG007 and HSJD-GBM001, respectively (Figure R22A). Notably, there was a large overlap between the proteins identified within each cell line (6232). In order to evaluate if both cell lines differ in their proteome composition, we performed a Principal component analysis (PCA) analysis using the absolute abundance of the total proteins identified in any condition. Interestingly, PC-2 segregated our samples based on the cell of origin, which suggests that HSJD-DIPG007 and HSJD-GBM001 harbour specific proteomes (Figure R22B).

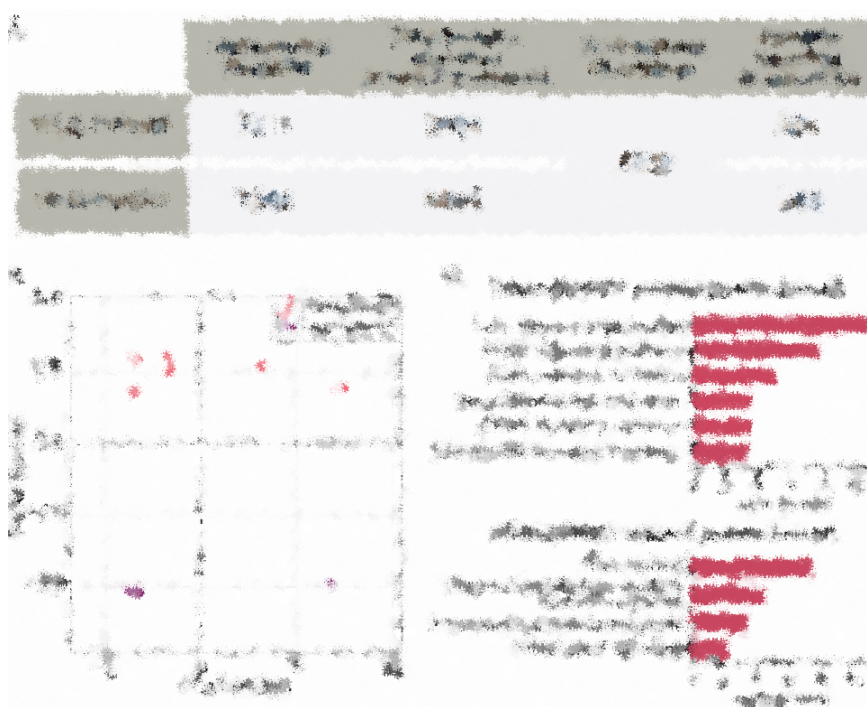


Figure R22. (A) Summary table of the proteome results for HSJD-DIP007 and HSJD-GBM001. (B) PCA analysis on the relative abundance of the total proteins

the highest enriched proteins in HSJD-GBM001, which is consistent with TP53 downregulation in HSJD-DIPG007. Overall, these results suggest that our proteome data capture the main phenotypic differences between HSJD-DIPG007 and HSJD-GBM001, and that could be a valuable source of information to understand the mechanisms promoting DIPG diffusely infiltration, which is one of the main characteristics associated with the aggressiveness of the disease and the main obstacle to DIPG surgical interventions.

Next, in order to identify the chromatin-bound proteins in both cell lines, we first compared the eluates from EdU-treated and non-treated samples in order to select the proteins significantly enriched in EdU (purified chromatin) versus non-EdU (contaminants) conditions. For that, we perform paired-wise comparison analysis for each replicate using DEqMS, a method developed specifically for differential protein expression analysis using quantitative MS-data [329]. This analysis allowed us to identify a total of 1079 and 622 chromatin-bound proteins significantly enriched in EdU samples (p .value<0.05) for HSJD-DIPG007 and HSJD-GBM001, respectively (Figure R23A). PCA analysis using the absolute abundance of EdU treated samples for the chromatome proteins identified in any condition showed that PC-1 and PC-2 segregated our samples based on the cell of origin, which suggests that HSJD-DIPG007 and HSJD-GBM001 also harbour specific chromatomes (Figure R23C). Finally, consistently with enrichment of chromatin-bound proteins using iPOTD, GO terms enrichment analysis showed mRNA binding, DNA repaired, or protein translation, among the top 10 category

terms related to the molecular function of the chromatin-bound proteins identified in HSJD-DIPG007 and HSJD-GBM001 (Figure 27D).

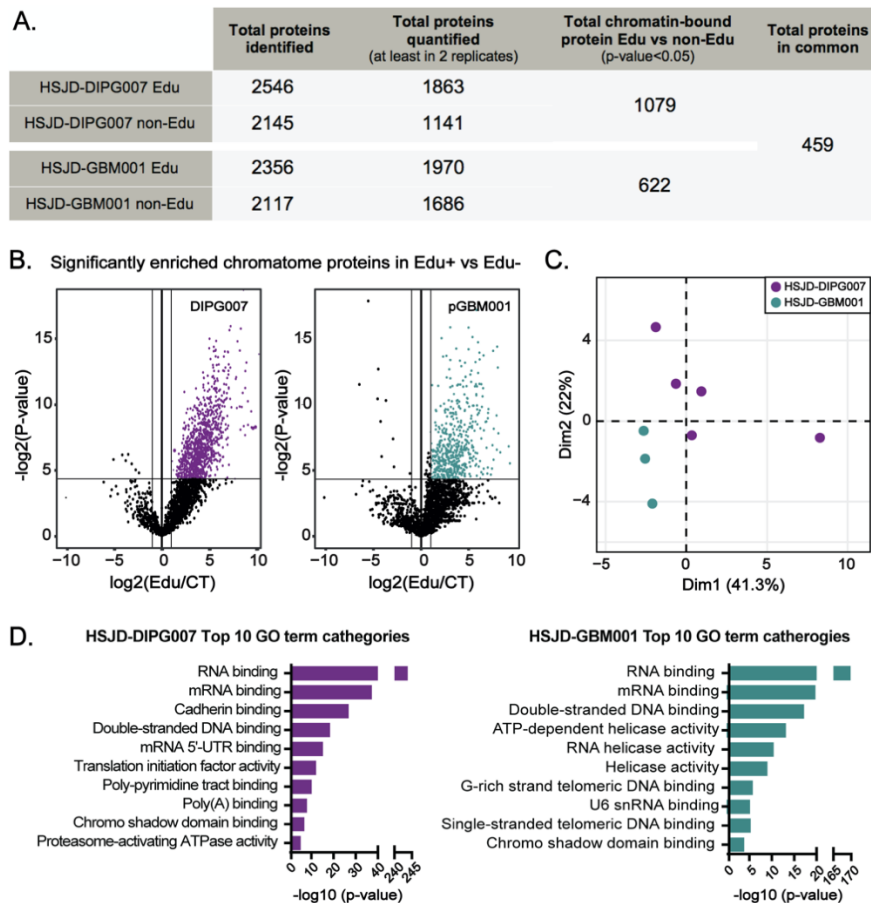


Figure R23. (A) Summary table of the chromatin results for HSJD-DIPG-007 and HSJD-GBM-001. (B) Volcano plot depicting all proteins identified in the iPOTD mass spectrometry analysis in HSJD-DIPG-007 and HSJD-GBM-001. Proteins significantly enriched (p-value < 0.05) in Edu+ over Edu- samples (shown coloured) are identified as chromatin. (B) PCA analysis on the relative abundance of the total chromatin-bound proteins identified in HSJD-DIPG-007 and HSJD-GBM-001. The PCA plot was generated by our laboratory SeqCode webserver (C) GO term enrichment analysis of the chromatin-bound proteins identified in HSJD-DIPG-007 and HSJD-GBM-001 using EnrichR. The top 10 molecular function categories are depicted and ranked by p-value.

The sensitivity to detect low-expressed proteins is reduced when performing biochemical purification methods. To determine iPOTD sensitivity, we compared the absolute abundance for the chromatin proteins with respect to the whole proteome. We found that iPOTD is sensitive enough to detect low expressed proteins both in DIPG and GBM cells (Figure R24A), thus revealing a notable sensitivity. In addition, to evaluate iPOTD specificity on capturing chromatin-bound proteins, we calculate how many of the proteins present in our dataset were classified as nuclear proteins in other cell types. For that, we used a recent identification of the subcellular distribution of proteins in five different cancer cell lines [330]. Remarkably, our chromatome data contains up to 60% of experimentally validated chromatin-bound proteins, which is similar to previously reported iPOTD results in mESCs [331], and almost double compared to total proteome (Figure R24B). These results further supported the sensitivity and the specificity of our iPOTD data and ultimately demonstrated the adequacy of our approach to profile the chromatin composition of glioma cells.

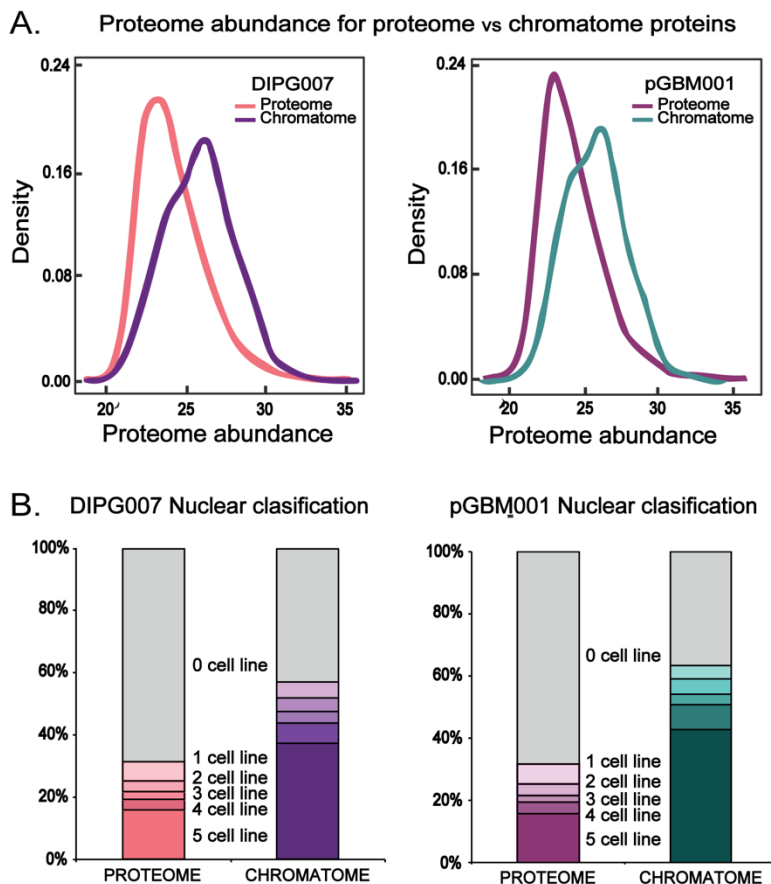


Figure R24 (A) Density plot depicting the abundance of chromatome and proteome identified proteins within the whole protein extract of HSJD-DIPG-007 and HSJD-GBM-001 cells. The x-axis represent the absolute protein abundance values from the total proteome LC-MS/MS analysis, while the y-axis represent the probability density of each value within the list of chromatome or proteome identified proteins. (B) Subcellular distribution of the list of chromatome or proteome identified proteins in HSJD-DIPG-007 and HSJD-GBM-00. The different colours associated with the number of cell lines in which a specific protein have been experimentally localized at nuclear [330].

5. The chromatin-bound proteins enriched in DIPGs represent potential chromatin interactors driving oncogenic outcomes.

To identify proteins enriched in the chromatin of DIPG, we compared the chromatin-bound proteins identified in HSJD-DIPG007 and HSJD-GBM001. For this purpose, we performed two group comparison analysis using the list of chromatome proteins of HSJD-DIPG007 and HSJD-GBM001 and their absolute abundance values in EdU treated samples. We found 164 and 21 proteins significantly enriched in HSJD-DIPG-007 and HSJD-GBM001, respectively (p-value<0.1). To explore the potential function of HSJD-DIPG007 enriched proteins, we used the search tool for retrieval of interacting genes (STRING) to acquire protein-protein interaction (PPI) networks among our 164 DIPG enriched proteins (Figure R25). Then, functional clusters in PPI networks were determined using GO term enrichment analysis.

We found among the top represented proteins networks multiple categories related to

[REDACTED]

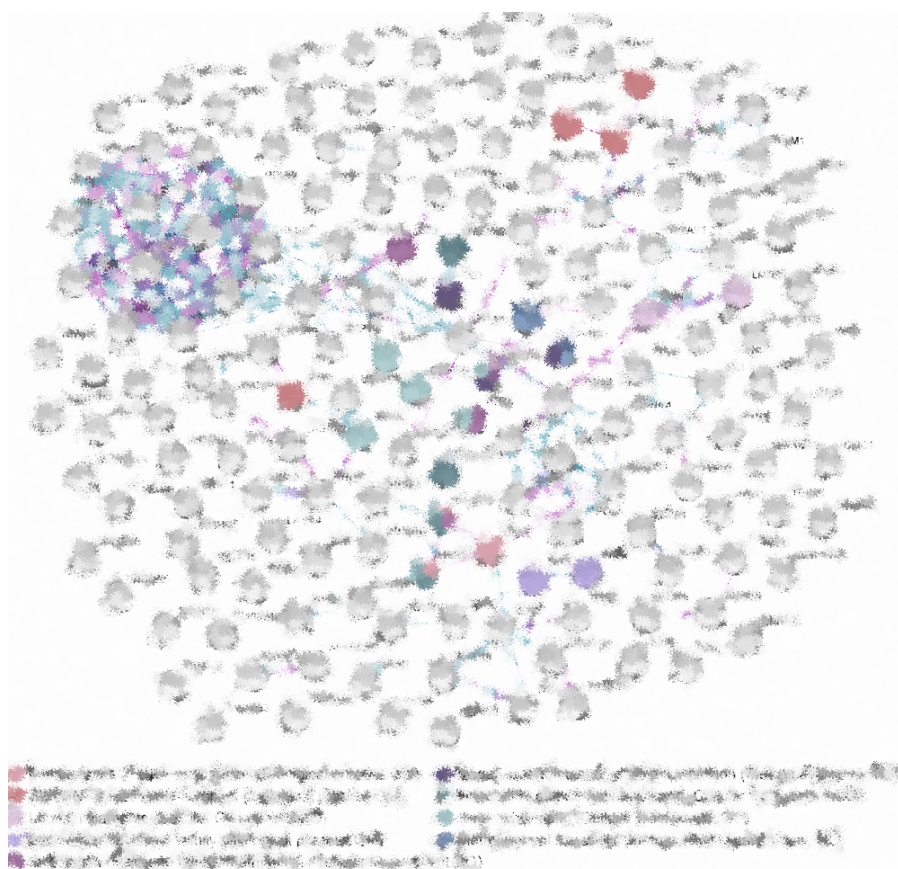




Figure R25. (A) Chromatome characterization from the STRING database. Only experimentally validated protein-protein interactions (blue) and curated databases (pink) were considered to establish associations between proteins networks. GO terms enrichment analysis depicted manually curated categories among the top fifteen Biological process and the top 10 Cellular component captured networks.

In addition to [REDACTED], we also found among the top enriched protein networks some metabolic terms, as spermidine synthase family and methylglyoxal metabolism (Figure R25). Recent studies demonstrated that chromatin regulation via cellular metabolism is much broader than had previously been considered [333]. In this line, methylglyoxal metabolism has been recently associated with control of self-renewal and differentiation of neural precursor cells (NPCs) during development [334]. Notably, during

the time of these analyses, a study associating high levels of spermidine to DIPG malignancy, and suggesting polyamine transport inhibition as a DIPG treatment, was published [335]. In this line, although with significance values below our threshold, we identified enriched in our HSJD-DIPG007 chromatome several proteins previously suggested as therapeutic targets for DIPG, as LSD1, BMI1, EZH2 [278, 332, 335, 336]. These findings support the strength of our strategy and reinforce the reliability of our results.

In order to test our initial hypothesis that the enrichment of specific proteins in the chromatin of DIPG might sustain their tumorigenicity and might be considered potential vulnerabilities for pharmacological targeting, we narrow down our list of HSJD-DIPG007 enriched proteins to focus on the most promising therapeutic targets based on the following criteria: i) to be part of a chromatin remodelling functional category; ii) to be associated with PRC2 misregulation in other cancers, and iii) to be a positive hit in genetic functional screenings performed on DIPG cells [332, 336, 337]. HSJD-DIPG007 enriched proteins having two of the below criteria were manually inspected to select those candidates with a potential relevant function within DIPG biology, that have not been explored previously as DIPG targets, and that was potentially druggable (Figure R26). After this filtering, we end up with 7 candidates for further functional exploration



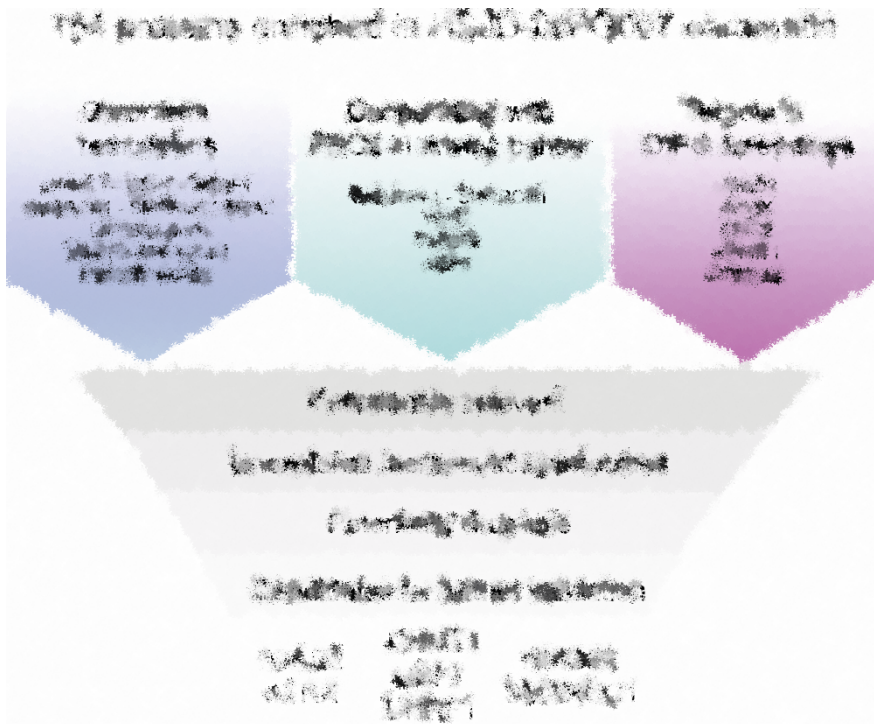


Figure R26. Summary of the main selection criteria to establish the list of chromatin candidates driving DIPG oncogenesis.

All these protein candidates are key epigenetic regulators mediating DNA methylation, nucleosome position, or 3D chromatin structure, and all have been previously reported to exert a function in promoting tumorigenesis, in some cases even depending on PRC2 misregulation. We further speculate about their relevance in the discussion.

6. Validation of our list of potential therapeutic candidates proteins for DIPG treatment.

Before continuing with the functional validation, we wanted to evaluate the chromatin enrichment of our candidates by using classical subcellular fractionation and Western blot analysis. We performed a biochemical fractionation method that combines different ionic and nonionic detergents to separate by centrifugation different cellular fractions, including chromatin [86]. Next, we used these fractions to analyse by western blot some of the DIPG-enriched chromatin proteins. In line with our proteomic approach, we detected an enrichment on HSJD-DIPG007 chromatin for 5 [REDACTED] [REDACTED] among 7 tested proteins using this alternative method (Figure R27). It is important to note that this biochemical fractionation method use chromatin in native conditions and that it is likely to differ on the efficiency of chromatin-bound proteins purification comparing to iPOTD. Thus, we argue that these findings validate our DIPG-enriched chromatin-bound proteins results.

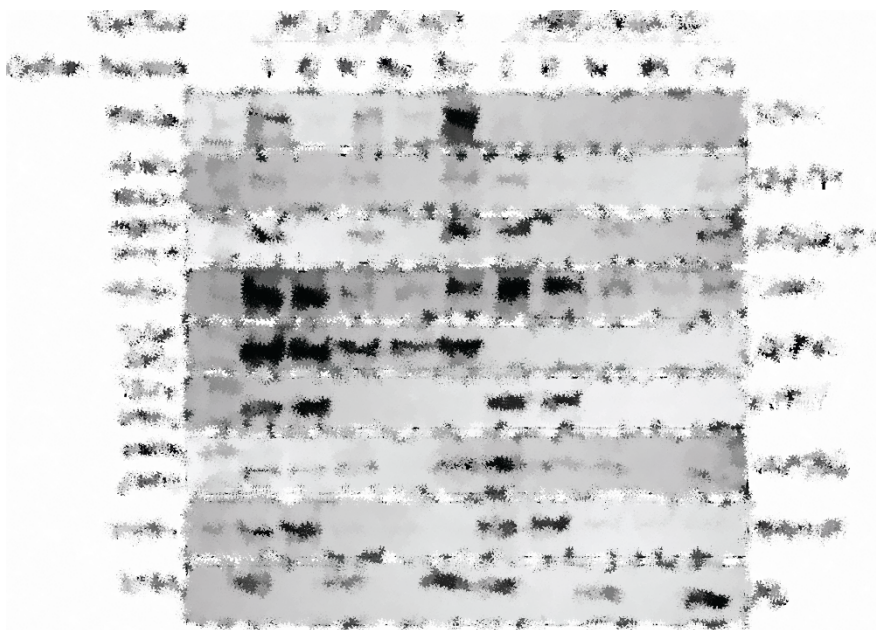


Figure R27. Westernblot validation of iPOTD identified proteins in different subcellular fractions isolated by an alternative biochemical fractionation method [86]. H3 and Vinculin were used as loading controls. (T=Total protein, C=Cytoplasm, N=Nucleus, Np=Nuclear plasma, Chr=Chromatin).

To determine the extent to which our therapeutic candidates could be contributing to sustain DIPG viability, we decided to knockdown our list of 7 candidate proteins in HSJD-DIP007 and HSJD-GBM001. If the candidate protein contribute specifically to sustain DIPG viability, we expect to observe an impact over HSJD-DIPG007 cell fitness, but not over HSJD-GBM001. Thus, we designed two independent shRNAs for targeting each candidate and we inserted each of them on the pLKO.1 lentiviral vector, that contains a puromycin resistance. We succeed to generated pLKO.1 plasmids targeting all our candidates, and two additional ones targeting EZH2 as a positive control of impaired DIPG viability, as EZH2 have been previously described to inhibit DIPG cell growth [278]. A pLKO.1 plasmid with an insert non-targeting randomized sequence was used

as a negative control. Unfortunately, during the process of writing this thesis, we were performing the cell viability assays and, therefore, we could only focus our subsequent studies on the evaluation of the effect of [REDACTED]. The generation of the pLKO.1 plasmids, transduction, Western blot validation, and cell viability analysis were performed in collaboration with Dr. François Le Dily.

We transduce with five independent shRNA construct targeting [REDACTED] and a non-targeting random sequence. Western blot analysis from HSJD-GBM001 cells after puromycin selection show the reduction of [REDACTED] [REDACTED] protein levels upon transduction with shRNA [REDACTED] [REDACTED] (Figure R28A). After several attempts, we were unable to perform Western blot analysis from HSJD-DIPG007 transduced cells. This was because after infection with the specific shRNAs and the addition of puromycin for selection, we rapidly lose the resistant culture. This suggested to us that the depletion of either [REDACTED] was extremely deleterious for HSJD-DIPG007. To accurately quantify this potential deleterious effect in cellular fitness, we plated an equal number of positively selected cells in 96 well plates and we performed MTS assays. The MTS assay protocol is based on the reduction of the MTS tetrazolium compound by NAD(P)H-dependent dehydrogenase enzymes in metabolically active cells, which generate a coloured formazan dye that is quantified by measuring the absorbance. Thus, absorbance values could be correlated to metabolically active cells,

which translate into a number of viable cells. Interestingly, we found that [REDACTED] depletion had a little positive effect on the growth of HSJD-GBM001, whereas HSJD-DIPG007 viability was seriously compromised (Figure R28B). These findings were consistent in two independent experiments, except for [REDACTED] in HSJD-DIPG007 (R1 and R2, Figure R28B).

Given our aim to find therapeutic targets that could be used indistinctly for all DIPG patients, we focused our subsequent studies on evaluating [REDACTED] depletion in other HSJD-DIPG cell lines. To this end, we treated HSJD-GMB001, HSJD-GBM002, HSJD-DIPG007, HSJD-DIPG017 and HSJD-DIPG021 DIPG cells with DMSO or with 5 μ M of [REDACTED] Cat. No. 4485. Initial observations pointed towards a decrease in cell proliferation upon [REDACTED] in DIPG cell lines, whereas differences in GBM were not noticeable (Figure R28C).

Altogether, these results suggested that [REDACTED] are required to sustain DIPG viability, whereas they are dispensable for other high-grade pediatric glioblastomas, and that the characterization of the chromatin-bound proteins could provide essential information about the main proteins driving cell functionalities.

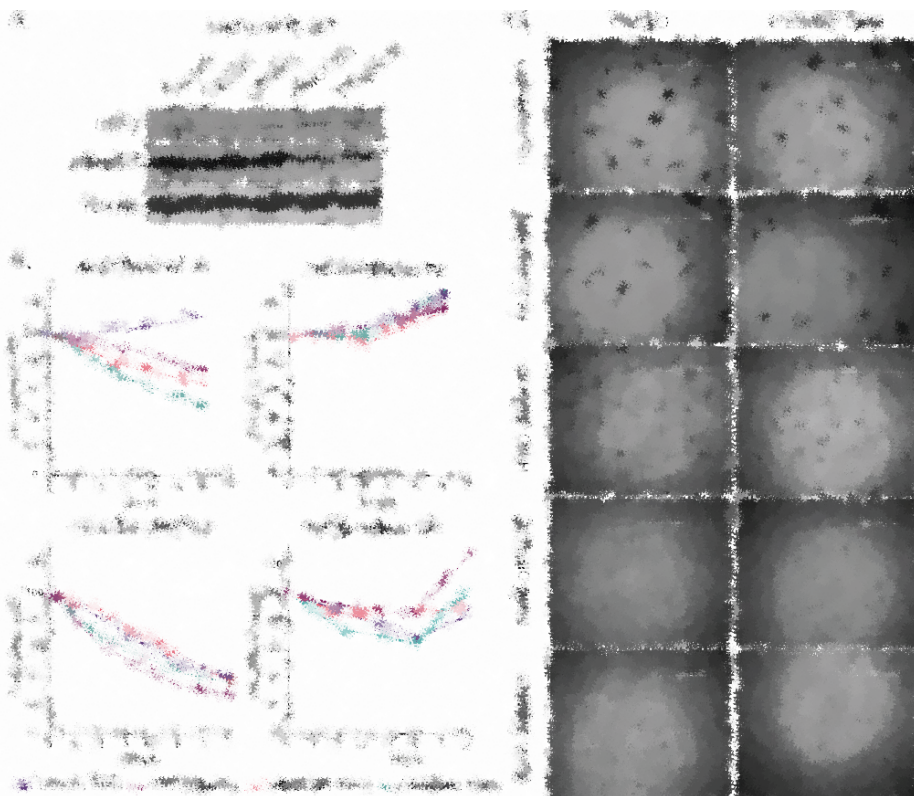


Figure R28. Westernblot of HSJD-GBM001 lysates collected after effective puromycin selection of cells transduced with a non-targeting sequence (Random) or with shRNA targeting [REDACTED]

[REDACTED] To avoid a bias from the potential differences between the initial number of plated cells the data was normalized with respect to day 0. The number of viable cells was depicted as the percentage of viable cells with respect to the number of viable cells transduced with a non-targeting random sequence. (B) Evolution of the percentage of HSJD-DIPG007 and HSJD-GBM001 viable cells after transduction with [REDACTED] in two independent experiments.

(C) Bright-field images of different HSJD cell lines treated with DMSO or 5 μM of [REDACTED].

Discussion

PRC2 involvement in pathogenesis was highlighted with the first PRC2 KO mouse, which unveiled its importance during embryonic development and in adult haematopoiesis [202, 339, 340] Later, the first genome-wide sequencing studies revealed recurrent PRC2 mutations in human hematopoietic malignancies, which reinforced the central function of PRC2 in diseases, especially in cancer [216, 217]. More recently, the identification of the first oncohistone, H3K27M, has brought another landmark, highlighting also the role of the chromatin misregulation in promoting PRC2 oncogenic function [266]. This thesis aims to (1) understand the biological contribution of PHF19, a substoichiometric partner of PRC2.1, during haematopoiesis; and (2), unveil whether PRC2 deregulation might offer therapeutic opportunities for cancer treatment. To this end, in the first part of my thesis, I studied the effect of PHF19 depletion *in vivo*, thus illustrating how an alteration in PRC2 composition could have an impact on adult haematopoiesis that ultimately could lead to diseases. In the second part of my thesis, I aimed to identify potential targets for the treatment of DIPG, by using genetic and proteomic approaches.

The effect of PRC2 depletion in the hematopoietic system.

The homozygous depletion of core PRC2 in the adult hematopoietic compartment leads to HSC exhaustion and post-natal death, whereas heterozygous depletion of core PRC2 subunits increases HSC self-renewal, repopulation capacity, and the incidence of hematopoietic diseases [200, 203, 204, 206, 232, 338]. In addition to the multiple GoF and LoF PRC2 mutations in hematopoietic malignancies, these evidences illustrate that (i) PRC2 function is essential to control the balance between HSC quiescence, self-renewal, and differentiation; (ii) hematopoietic cells are sensitive to and respond differentially to PRC2 dosage; and (iii) any slight deviation in PRC2 function can dramatically alter haematopoiesis. All these suggest that PRC2 is a major regulator of the hematopoietic compartment. However, despite that PRC2 accessory subunits are known to fine regulate the PRC2 function, the role of these proteins in the hematopoietic system has not been thoroughly addressed.

Indeed, in addition to our work about PHF19, only JARID2 has been studied in adult haematopoiesis [205]. Interestingly, *Jarid2* cKO and *Phf19* KO mice hematopoietic phenotypes are less severe than core PRC2 cKO and, importantly, **differ between them**, suggesting that PRC2 accessory subunits in haematopoiesis have a non-overlapping, or even opposing, functions in haematopoiesis.

Comparative transcriptome analysis highlighted a similarity between *Phf19* KO LT-HSC and highly quiescent and HSC-like LT-HSC populations.

Multiple studies have combined the phenotypic and the transcriptomic characterization of HSC *in vivo*, which have established associations between the specific stem cell capacities of different HSC populations and changes in their transcriptome [296, 298, 310]. In this line, our first transcriptomic results suggested that LT-HSC strongly decrease their biosynthetic activity upon depletion of PHF19. We found that overall low biosynthetic activity is the transcriptomic hallmark of a specific subpopulation within HSC known as dHSCs and, indeed, further comparison between our LT-HSC and dHSCs RNA-seq data demonstrated a high transcriptional similarity between both populations in homeostasis [296, 297].

dHSCs are characterized by a high quiescence and multipotent capacity. Further to our purpose to provide explanation to *Phf19* KO mice phenotype, a potential increase in LT-HSC quiescence was consistent with the lower proliferation of *Phf19* KO LT-HSC *in vivo*. However, the high multipotency associated to dHSC was inconsistent with the phenotype of *Phf19* KO mice in response to stress, i.e. an exacerbated bias towards myelopoiesis upon aging, and a decrease in blood production capacity upon serial transplants. In this line, we hypothesized that *Phf19* KO HSC and dHSC must have similar characteristics in homeostasis, particularly high HSC transcriptional identity and quiescence, but not under stress conditions. This was supported by several studies suggesting that the transcriptome of HSC in homeostasis does not correlates with their response upon mobilization [310, 311]. Interestingly, upon ageing, there is an increase of cells with HSC identity in the LT-HSC compartment, that upon stress, displayed a bias toward myelopoiesis and an impaired

contribution to sustain blood production [339, 340]. We then demonstrated a high correlation between *Phf19* KO and aged HSCs transcriptome [298] and, further supporting these similarities, we found a transcriptome bias toward myeloid differentiation in *Phf19* KO downstream haematopoiesis precursors. Moreover, we found an increased correlation between *Phf19* KO LT-HSC transcriptome and LSC, a leukemic cell population with a high quiescence and self-renewal potential that are suggested to emerge upon aging [341-343].

Overall, these results suggested that *Phf19* KO LT-HSC display an increased HSC and quiescent transcriptional identity, and that may induce transcriptional changes in young LT-HSC that recapitulates those occurring during ageing and disease. Indeed, the phenotypes from young *Phf19* KO and old WT mice are very similar except from (i) a lower number of HSCs in young *Phf19* KO mice, in contrast to the progressively increasing number of HSCs during ageing [340], and (ii) a less dramatic defect in *Phf19* KO HSCs long-term repopulation capacity [291]. These differences are not unexpected, as HSC aging is the result of the interconnection between multiple alterations that are not only associated to chromatin structure changes, but also to changes in DNA damage [308], autophagy [344], metabolism [345], or the niche [346], among others. Even so, these differences initially prevented from establishing a correlation between *Phf19* KO and aged mice. In this regard, the cross-study comparative transcriptome analysis demonstrated to be a useful strategy to associated different stem cells capacities to HSCs beyond the classical phenotypic approaches to study HSC function [347].

PHF19 is required to balance HSC chromatin state and its capacity to dynamically change upon differentiation.

One of the main achievements of this thesis was to perform a H3K27me3 ChIP-seq, including spike-in, in as few as 10,000 HSCs. Consistently, one of our most relevant results was to identify for the first time an H3K27me3 increase following the depletion of a PRC2 subunit in HSCs. Our analysis showed that the highest H3K27me3 enrichment occurs at genes that were already repressed in our control HSCs, including essential genes for hematopoietic differentiation. In contrast, HSC identity genes did not gain H3K27me3. This led us to speculate that PHF19 depletion may lock HSC chromatin structure by enhancing facultative heterochromatin at repressed genes. This could represent a disadvantage to trigger transcriptional activation of genes required to exit quiescence and to undergo further differentiation, which may be the origin of the enhanced quiescence and HSC identity of *Phf19* KO LT-HSCs.

Interestingly, one of the most accurate epigenomic studies in murine HSC ageing performed by Sun et al. shows that, as HSCs age, the number and genomic location of H3K27me3 peaks remain largely unchanged, but the intensity of the H3K27me3 peak signal increases up to 50% [310]. Notably, we demonstrated that the H3K27me3 profile of young *Phf19* KO HSCs is more similar to those reported at aged HSCs. Although Sun et al. did not go further and establish an association between these H3K27me3 peaks and genes required to trigger HSCs differentiation, they showed increased DNA methylation at TF binding sites associated with differentiation-

promoting genes combined with a reduction at genes associated with HSC maintenance. In addition, they performed H3K4me3 profile and demonstrated an increase of H3K4me3 across HSC identity and self-renewal genes. Together, Sun et al. suggested that these changes reinforce HSC self-renewal and diminish differentiation in aged HSCs. Similarly, another work associated the increase of DNA methylation at genes required for haematopoiesis lineage commitment with the impaired differentiation and decreased long-term repopulation capacity of aged HSCs [348]. Based on these similarities, we speculate that a decrease in HSCs chromatin dynamic upon PHF19 depletion may be a mechanism involved in HSC aging and that PHF19, or in general an increase in H3K27me3, could contribute to this process. In this line, *Phf19* is among the top genes downregulated upon HSC ageing [310].

PHF19 depleted LT-HSCs displayed increased proliferation and decreased differentiation upon stress induction.

In order to test if *Phf19* KO HSC presented a compromised exit from quiescence and/or differentiation, we cultured single LT-HSCs under conditions that promote cell division. From these experiments we concluded that PHF19 depleted LT-HSC are able to exit quiescence, even with a higher proliferation capacity, but that they have impaired differentiation. This is in line with the reported increase of HSCs in *Phf19* KO BM upon stress induction (as transplantation or ageing) but the decrease contribution of these HSCs to blood production. We speculate that a possible explanation for these results would be that PHF19 KO HSC suffers a bias toward

self-renewing. In this line, it would be interesting to perform transcriptomic and epigenomic assays in these *in vitro* proliferating HSCs, in order to further understand the reason behind the increase proliferation and impaired differentiation in mobilized *Phf19* KO LT-HSCs.

It is also important to note the increased number of single LT-HSCs annotated as proliferative in the absence of PHF19. This suggests a more homogenous response of *Phf19* KO LT-HSCs upon stress, which may be the result of the overrepresentation of one HSC clonal population within the LT-HSC compartment. Interestingly, multiple studies, including the one of Sun et al., have shown that LT-HSC compartment increase in size and homogeneity upon aging as a consequence of the overexpansion of one HSC clonal population with a potent self-renewal activity and a bias toward myeloid differentiation [298, 306, 349]. Moreover, it has been demonstrated that while young HSCs divide mainly asymmetrically, to produce one cell that retains stem cell potential and one that differentiates, aged HSCs divide primarily symmetrically, to produce two stem cells, meaning self-renewal [350]. On the other hand, strong evidence in AML points to the origins of LSCs in pre-leukaemic cells that arise through the sequential accumulation of somatic DNA mutations in highly quiescent HSCs. The consequences of these early mutations are the enhancement or acquisition of self-renewal potential and, often, impairment of differentiation. Overall, our results suggested a more homogenous *Phf19*-KO LT-HSC compartment as a consequence of the overexpansion of a highly proliferative HSC population that displays a decreased differentiation capacity. Once

again, these results resembled those described upon LT-HSC aging and LSCs.

The molecular mechanism of PHF19 in LT-HSC.

Our study leaves some important questions unanswered, such as how PHF19 can limit PRC2 H3K27me3 deposition, and further, how PHF19 depletion leads to the increased deposition of H3K27me3 at HSC repressed genes, but not at HSC fingerprint genes. For that, ChIP-seq analysis of PHF19 and other PRC2 subunits in LT-HSCs would be particularly informative, but they are technically unfeasible due to input limitations [69]. Furthermore, we found that addressing those questions using HPC-7, as an alternative model, is not an option, considering the large transcriptional differences with LT-HSC. In this regard, we believe that the assumption that HPC-7 is a good alternative to LT-HSC should be reconsidered.

Nevertheless, we speculate that PHF19 may be the first example of a constitutive factor need to limit PRC2 activity in LT-HSCs, a mechanism that has been already suggested in mESCs for other PRC2 accessory proteins, such as EPOP [136]. Indeed, as mentioned on the Introduction, PHF19 targeting of PRC2 at H3K36me3 regions have been associated with both, inhibition and increase of PRC2 activity [130-133]. Hence, we hypothesize that upon depletion of PHF19 in LT-HSCs, other PRC2 accessory subunits may increase their interaction with PRC2 and positively regulate PRC2 activity and recruitment, leading to an increase of H3K27me3 deposition at specific regions. Supporting this hypothesis, JARID2 cKO and

MTF2 KO mice have been shown to induce a global decrease of H3K27me3 in adult and embryonic HSCs respectively, suggesting a role as positive regulators of PRC2 function in HSCs [196, 205]. However, this hypothesis may be difficult to demonstrate experimentally in LT-HSCs. To this aim, the generation and characterization of double PRC2 accessory subunits cKO mice might be of particular interest.

Functional model for PHF19 mediated LT-HSC fitness and potential implications in aging and hematopoietic malignancies.

From all the data generated in the first chapter of this thesis, we propose a functional model in which PHF19 is essential to regulate H3K27me3 levels in HSCs, to balance between preservation of cell fate identity and chromatin dynamics (Figure D1A). We believe that this model could explain the main *Phf19* KO phenotypic results obtained in our mice model.

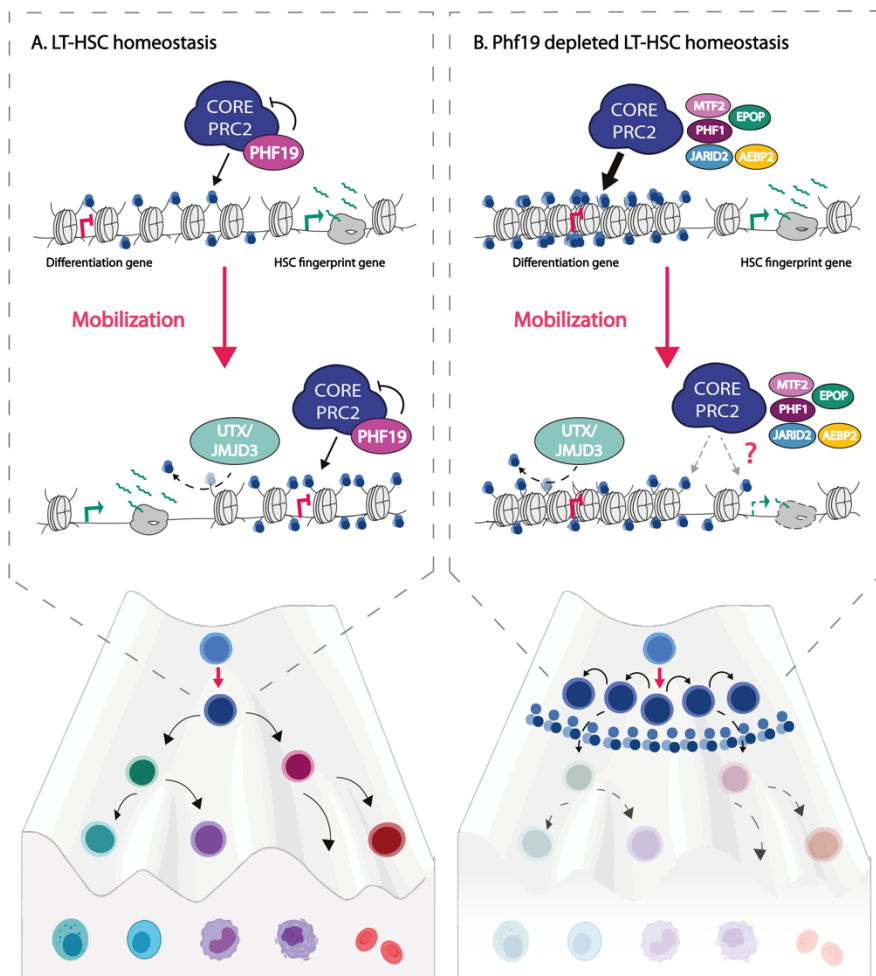


Figure D1. Schematic representation of the proposed model for PHF19 function in LT-HSCs. (A) PHF19 negatively regulates PRC2 activity to maintain proper H3K27me3 levels that allow for an efficient chromatin remodelling upon LT-HSCs mobilization. (B) PHF19 depletion unbalance PRC2 composition and subsequent activity leading to an aberrant increase in H3K27me3. Upon LT-HSCs mobilization, we speculate that this H3K27me3 increase impairs efficient chromatin remodelling and consequent transcriptional changes, which compromise a normal entrance to the hematopoietic differentiation cascade and lead to an accumulation of phenotypically defined LT-HSCs. However, additional experiments must be done to unveil the transcriptional and epigenetic dynamics during mobilization of PHF19 depleted LT-HSCs.

We also speculate that a similar mechanism may occurs in HSC upon aging, and that PHF19 may play a role in this process. However, whether PHF19 is a major regulator in HSC aging requires additional

experimental evidence. We believe that further efforts should be done, as it is particularly relevant considering that (i) aging is the major risk factor for several hematologic malignancies [351] and that (ii) epigenetic changes occurring in aged HSC are reversible, as epigenetic reprogrammed aged HSCs demonstrate similar repopulation capacity as young HSCs [352, 353]. This suggests that understanding PRC2 activity in the hematopoietic system may provide new tools to define pharmacological targets to reduce the negative effects of aging and decrease the incidence of diseases.

On the other hand, HSC quiescence has been connected with the higher accumulation of DNA damage, and it has been suggested as the origin of LSCs and the increased incidence of hematopoietic malignancies in aged mice. We speculate that an increased quiescence in *Phf19* KO LT-HSC could promote a higher accumulation of DNA damage, which could explain the transcriptional similarity with LSCs and the increased incidence of myeloid diseases in *Phf19* KO aged mice. We found high levels of γ HA2X, an early marked of double-stranded breaks of DNA, on LT-HSCs from *Phf19* floxed and *Phf19* KO animals. In this regard, as high levels of γ HA2X in *Phf19* floxed LT-HSCs were unexpected, we believe that a more thorough characterization of DNA damage in *Phf19* KO LT-HSCs is needed to further test our hypothesis, for example using DNA repair reporters to quantify mutagenic Non-Homologous End Joining (NHEJ) [354]. Finally, to conclude the discussion of this first part, I want to underscore that all the results generated in our study about *Phf19*-KO mice (Annex I) have

Discussion

motivated the study of PHF19 function in leukaemia development and progression, a project that is currently being conducted by Gianni Gamarra, a PhD student in our laboratory.

Successful incorporation to our laboratory of a large cohort of pediatric glioblastoma primary cell lines as a valuable tool to study DIPG.

Our understanding of H3K27M function in chromatin and the identification of potential therapeutic targets for DIPG have advanced significantly in recent years. This is largely thanks to the development of the first DIPG cell lines from patients biopsies by a few laboratories, including HSJD among others [355]. The possibility to establish a collaboration with researchers and oncologist from HSJD, bringing these powerful models to our laboratory for the first time, has been a major outcome of this thesis. We have succeeded in growing and expanding 9 different HSJD primary cell lines established from 7 DIPG patients, therefore harbouring diffuse tumours positive for H3K27M mutations, and 2 patients diagnosed with other types of high-grade paediatric glioblastomas, negative for H3K27M mutations. All grew under identical conditions but displayed different particularities that make them more suitable to be manipulated in the laboratory.

Generation of HSJD-DIPG007 and HSJD-GBM001 cultures to perform a CRISPR drop-out screening using ARUNA library.

Since the generation of DIPG cell lines, several studies have reportedly succeeded in introducing DNA and RNA into DIPG cells using transfection and transduction methods, and selecting these cells using FACS sorting and antibiotic resistance [277, 278, 285, 336]. In our attempt to perform a CRISPR Cas9 drop-out screening, we have performed a deep characterization of the different efficiency of these techniques in HSJD-DIPG007 and HSJD-GBM001, which we

Discussion

believe would serve as the basis for future experiments in our laboratory. We first demonstrated that viral transduction is an efficient method in these cells, whereas transfection is highly inefficient. Next, we showed that HSJD-DIPG007 cells viability is compromised by FACS sorting, whereas HSJD-GBM001 sorted cells grow normally. This is extremely important as it is a limitation to select HSJD-DIPG007 cells in order to generate homogenous cultures. Thus, to evaluate the effect of sorting over other HSJD-DIPG cell lines it is within our short-term goals. Finally, we demonstrated that puromycin selection led to highly homogenous cultures, whereas neomycin and blasticidin are a bit less efficient. We speculate that these differences may be driven by the 7-10 days of neomycin and blasticidin selection versus the 3-4 days of puromycin, which may promote the selection and expansion of clones with some natural resistance to these antibiotics. Overall, these findings suggest that lentiviral vectors with puromycin resistance should be always used with HSJD-DIPG007 cells if possible.

Our efforts have resulted in the generation of sufficiently homogenous HSJD-DIPG007 and HSJD-GBM001 Cas9/ARUNA+ cultures to allow us to perform our CRISPR Cas9 screening. We are currently in the final days of our experiment, and during this time we have periodically collected genomic material from our cultures in order to amplify and sequence the sgRNA cassettes. Thus, the expected result of this experiment would be a list of sgRNA depleted in HSJD-DIPG007 but not in HSJD-GBM001 cultures, which correspond to the potential chromatin-bound proteins required to sustain DIPG cell growth. It is worth noting that, to our knowledge,

this would be the first CRISPR Cas9 screening comparing DIPG cells with H3K27WT paediatric glioblastoma cells. Previous studies by Balakrishnan et al. have performed shRNAs screenings comparing DIPG cell lines with and without H3K27M, while Anastas et al. performed a CRISPR Cas9 drop out screening comparing SU-DIPGVI cells treated with and without panabioostat [332, 336]. Their results serve as a reference, but they will not overlap with our data, as any compare DIPG cells with other paediatric glioblastoma cells. Moreover, our custom-built ARUNA library targets 320 chromatin factors more than the shRNA library of Balakrishnan et al. (the entire library from Anastas et al. is not available), further highlighting ARUNA library as an extremely powerful tool, and the potential relevance of our future results.

Finally, it is important to mention that among our imminent goals is to develop isogenic cell lines to better evaluate the scope of our results from HSJD cell lines and to overcome our technical limitations. In this regard, we are considering expressing the H3K27M mutation in HSJD-GBM001 cells, which have been demonstrated to be the most robust cell line of the HSJD cohort and for which we already have generated multiple data. In addition, we are considering expressing H3K27M in mESCs, a model for which our laboratory has a wide knowledge and technical expertise. Notably, these models have already used and proved useful in other DIPG studies, which could also serve us as a future source of information [332, 336].

First characterization of the total proteome and chromatome of DIPG cells.

We have characterised for first time the proteome and the chromatome of a DIPG cell. Interestingly, our results suggest that the comparative analysis of the total proteome of HSJD-DIPG007 and HSJD-GBM001 has the capacity to unveil proteins sustaining the highly diffuse and invasive growth of DIPG cells.

In addition, we managed to performed iPOTD and DEqMS analysis to identify chromatin-bound proteins in HSJD-DIPG007 and HSJD-GBM001. Interestingly, we succeed to identify even more chromatin-bound proteins than those previously identified using iPOTD in mESCs in our laboratory [331]. Notably, LS-MS/MS analysis has several limitations that can lead to variability between samples, which could be mitigated by using statistical analysis that correct for inter-replicate variability, as DEqMS, or by increasing the number of replicates. This could explain the differences in the number of proteins identified between mESC and our cell lines, as we performed different statistical analysis (SAINTs in mESC vs DEqMS in our study) and more replicates (3 vs 5). In addition, as a way of strengthening our chromatome results, we demonstrated that our chromatome data captures some of the lowest abundant proteins among the input material, and that increased by two-fold the percentage of bona fide chromatin-bound proteins identified comparing to our proteome data. Together, our results highlight the reliability of our chromatome results, which we believe that, together with our proteome, will represent a valuable resource for the growing efforts aimed to understand this disease.

Identification of potential therapeutic targets for DIPG based on HSJD-DIPG007 chromatin enriched proteins.

In order to select among our chromatome data the proteins that may be more relevant for DIPG function, we performed multigroup comparison analysis to obtain the proteins significantly enriched on the chromatin of HSJD-DIPG007 as compared to HSJD-GBM001. Notably, we succeeded in validating the enrichment of several of these proteins using an alternative chromatin isolation method. In addition, we identified enriched in our HSJD-DIPG007 chromatome several proteins previously suggested as therapeutic targets for DIPG, as LSD1, BMI1, EZH2, and the spermidine synthase proteins SMS and SMR [278, 332, 335, 336]. These findings are proof of the strength of our strategy and reinforce the reliability of our results.

Next, we narrowed down our list of proteins to select proteins that could present a functional, pathological, and biological relevance within DIPG context and may represent potential therapeutic targets.

We ended up with a list comprising

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Validation of potential therapeutic targets for DIPG treatment.

To determine the extent to which these proteins are necessary to DIPG cells viability, and thus consider them as potential therapeutic candidates, we studied the effect of their knockdown over HSJD-DIPG007 and HSJD-GBM001 viability. Our initial results concerning [REDACTED] showed that decreasing of both proteins compromises HSJD-DIPG007 viability, but not HSJD-GBM001. It is worth noting that the depletion of [REDACTED] impairs HSJD-DIPG007 viability in as much as it is difficult to expand the cultures to obtain enough protein material to evaluate shRNA efficiency. We are thus planning to generate inducible shRNA lentiviral vectors, also harbouring a GFP reporter, in order to avoid potential bias effects due to differences in the efficiency of infection and subsequent knockdown. Finally, our preliminary results using [REDACTED] inhibitors in other HSJD suggest a negative effect over other HSJD-DIPG cell lines. However, further replicates must be performed before drawing any conclusion, and the effect of depleting [REDACTED] in other HSJD cell lines must be also evaluated using shRNAs

and alternative [REDACTED] inhibitors. In addition, other assays such as cell colony formation and invasiveness might be performed.

To conclude, going far beyond the aim of this thesis, we performed ATAC-seq libraries in triplicates for all HSJD cell lines, and we are currently sequencing these samples. Similarly, we are performing an equal batch of Hi-C libraries. We believe that, as a follow-up to the results generated in this thesis, the information gathered from these additional experiments could help us to better understand the molecular mechanism by which a potential therapeutic candidate may compromise DIPG cell survival. Going even further, if our final *in vitro* results provide strong evidences for the suitability of a specific protein as a therapeutic target, in collaboration with Dr Jaume Mora and Dr Ángel Carcaboso from HSJD, we plan to perform xerographs models to evaluate the effect of the inhibition of this protein *in vivo*.

Conclusions

Conclusions from **CHAPTER I**:

From the results obtained in the first chapter of this thesis we can conclude that upon depletion of PHF19:

1. LT-HSCs displayed an increase expression of genes associated with HSCs and with highly quiescent LT-HSCs populations.
2. LT-HSCs transcriptome is similar to aged HSCs.
3. LSK transcriptome is biased toward myeloid lineage differentiation.
4. LT-HSCs transcriptome is more similar to LSCs.
5. LT-HSCs showed a global increase in H3K27me3 levels and a correlated decrease in chromatin accessibility, with PHF19 being the only reported PRC2 subunit displaying a negative effect in PRC2 activity in LT-HSCs.
6. H3K27me3 increase and chromatin accessibility decrease are more significant at genes required for hematopoietic differentiation, while does not affect HSCs fingerprint genes.
7. A higher percentage of single cell LT-HSCs demonstrate high proliferation capacities but impaired differentiation upon *in vitro* stress induction.
8. HPC-7 is not a good alternative *in vitro* model to study LT-HSCs.

Conclusions from **CHAPTER II**:

Conclusions

From the results obtained in the second chapter of this thesis we can conclude that:

1. HSJD-DIPG007 and HSJD-GBM001 can be distinguished based on their proteomes.
2. HSJD-DIPG007 and HSJD-GBM001 proteome comparison is a useful approach to identify proteins sustaining DIPG diffuse and invasive growth.
3. HSJD-DIPG007 and HSJD-GBM001 chromatome characterization using iPOTD is a highly specific and sensitive approach.
4. HSJD-DIPG007 and HSJD-GBM001 can be distinguished based on their chromatome.
5. HSJD-DIPG007 and HSJD-GBM001 chromatome comparison is a useful approach to identify proteins enriched in DIPG chromatin that may play a role in DIPG tumorigenesis.
6. [REDACTED] proteins are required to sustain HSJD-DIPG007 cells viability, but not HSJD-GBM001.

Materials and methods

Material and methods corresponding to CHAPTER I

1. *In vivo* experiments

1.1. Mice

All animals used for this study had a C57BL/6 background. LoxP sites were introduced by homologous recombination around exons 4 and 5, the deletion of which generates a stop codon in exon 6. *Phf19^{fl/fl}* animals were crossed with Sox2Cre (mice were provided by A.R. Nebreda, Institute for Research in Biomedicine, Barcelona). The CRE activity in the Sox2CRE strain has been of maternal inheritance, irrespective of whether they inherit the transgene [349]. Therefore, after backcrossing with new male floxed animals, animals with the knocked-out allele without the SoxCre transgene were selected. Once generated, mice were maintained as heterozygosis (*Phf19^{fl/fl}*) and as homozygosis (*Phf19^{-/-}*). In both cases, offspring were viable and bred normally. Equal numbers of male and female mice were used; no gender biases were noted.

1.2. BM extraction

For pooled mice analyses, clean femurs and tibias were crushed with mortar and pestle in cold dissection medium (1X~ PBS, 2% FCS, and 2 mM EDTA) before cell surface staining (see next section). For sorting lineage depletion was performed using the mouse Lineage Cell Depletion Kit (Miltenyi Biotec) by magnetic-activated cell sorting (MACS) technique according to the manufacturer's instructions.

1.3. Flow cytometry

Before flow cytometry analysis BM cells were lysed with 1X~ red blood cell lysis buffer (eBioscience) and washed with 1X~ PBS. For surface staining, single-cell suspensions in FACS buffer (1X~ PBS, 2% FCS, and 2 mM EDTA) were blocked for 10 min with anti-CD16/CD32 (mouse Fc block, BD Pharmingen). Samples were stained in FACS buffer for 30 min at 4°C with fluorochrome-conjugated antibodies and washed before analysis. BM were analyzed using the lineage marker (Lin) mixture (BD Biosciences) that include the following V450-conjugated antibodies: CD3 ϵ (145-2C11); CD11b (M1/70); CD45R/B220 (RA3-6B2); mouse erythroid cells Ly-76 (Ter119), Ly6G, and Ly-6C (RB6-8C5); and the following fluorochrome-conjugated antibodies for LT-HSC and precursors (LSK): PECy7 anti-mouse Sca I, APC-Fluo780 anti-mouse cKit, APC anti-mouse CD150, and PE anti-mouse CD48. Cell viability was evaluated with 4',6-diamidino-2-phenylindole (DAPI) at a final concentration of 1:10,000.

1.4. Population RNA-seq

For LT-HSCs and LSK cells two RNA replicates were obtained from 5000 to 10,000 HSCs (Lineage–Sca-1+c-Kit+CD150+CD48–) and one RNA replicate was obtained from 5000 to 10,000 LSKs (Lineage–Sca-1+c-Kit+) sorted from flow cytometry as described above. RNA extraction was performed using the RNeasy Micro Kit (Qiagen). SmarSeq2 library preparation was performed by at the CRG Genomic Unit from 3 to 5 ng of total RNA extracted using the SMARTer Ultra Low RNA Kit (Clontech). All samples were

sequenced [50–base pair (bp) single end] on an Illumina HiSeq 3000. The RNA-seq samples were mapped against the mm9 mouse genome assembly using TopHat with the option `-g 1` to discard those reads that could not be uniquely mapped in just one region. Cufflinks was run to quantify the expression in RPKMs of each annotated transcript in RefSeq. We applied the following thresholds to find differentially expressed genes between each pair of *Phf19* floxed versus *Phf19* KO samples [RPKM > 0.5 and KO/Flox FC \geq 1.5 or Flox/KO FC \geq 1.3]. We intersected the up- or down-regulated genes in common between duplicates to end up with the final sets of up-regulated and down-regulated genes, respectively. Reports of functional enrichments of GO and other genomic libraries were generated using the Enrichr tool. For specific GSEA, we considered as expressed genes those with RPKM > 0.1. Next, we rank-ordered them according to fold change and analyzed using the preranked tool of GSEA.

1.5. Low Input Chromatin immunoprecipitation sequencing

10,000 HSC (Lineage⁻Sca-1^{+c}-Kit⁺CD150⁺) cells were sorted by flow cytometry as described above and freshly cross-linked with 1% FA. Low Cell ChIP-seq kit (Active Motif) was used following the manufacturer's instructions for histone H3 lysine-27 trimethylated ChIP grade antibody (39155, Active Motif). For ChIP experiments with spike-in control, a ratio of 1:2 (*Drosophila*:mouse) S2 cells was added prior cell lysis. Total chromatin was sheared with a Covaris sonication system for 20 min at intensity 8 and 20% duty cycle. ChIP-seq libraries were prepared with total recovered chromatin, using the Next Gen DNA Library Kit and the Next Gen Indexing Kit (Active

Materials and methods

Motif) as per the manufacturer's instructions. Libraries were sequenced (50-bp single end) on a HiSeq 3000 platform (Illumina). ChIP-seq samples were mapped against the mm9 mouse genome assembly using Bowtie with the option `-m 1` to discard those reads that could not be uniquely mapped to just one region. ChIP-seq samples normalized by spike-in were mapped against a synthetic genome constituted by the mouse and the fruit fly chromosomes (mm9 and dm3) using Bowtie with the option `-m 1` to discard reads that did not map uniquely to one region. MACS was run with the default parameters but with the shift size adjusted to 100 bp to perform the peak calling against the corresponding control sample. Each set of target genes was retrieved by matching the ChIPseq peaks in the region 2.5 kbp upstream of the transcription start site (TSS) until the end of the transcripts as annotated in RefSeq annotations. The heat maps displaying the density of ChIP-seq reads around the TSS of each target gene set were generated by counting the number of reads in this region for each individual gene and normalizing this value with the total number of mapped reads of the sample. Genes on each ChIP heat map were ranked by the logarithm of the average number of reads in the same genomic region. Box plots showing the ChIP-seq level distribution for H3K27me3 for particular sets of genes were calculated by determining the average value at transcript level on the region ± 2 kbp around each TSS. The values of the samples including spike-in were corrected by the number of fly reads mapped of the sequencing experiment. The University of California Santa Cruz (UCSC) genome browser was used to generate the screenshots depicted [372]. To confirm the general increase of H3K27me3 ChIP-

seq levels from *Phf19* floxed to *Phf19* KO conditions, we segmented the mouse genome into bins of 100 bp. Next, we counted within each bin the number of normalized reads of H3K27me3 corrected by spike-in at each sample. To discard regions of the genome virtually with no signal, we focused on the bins in which, at least in one of the conditions, a minimum amount of ChIP signal was detected (threshold: 5).

1.6. Omni-ATAC-seq

Omni-ATAC-seq libraries were generated as previously described. Briefly, 10,000 Lineage-Sca-1+c-Kit+CD150+CD48- cells per condition were washed in cold PBS and resuspended in 50 μ l of cold lysis buffer 1 [10 mM tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 0.1% (v/v) Igepal CA-630, 0.01% digitonin, and 0.1% Tween 20]. Samples were incubated on ice for 3 min and washed out with 1 ml of cold lysis buffer 2 [10 mM tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, and 0.1% Tween 20]. Samples were centrifuged for 10 min at 4°C, and the nuclei pellet was resuspended in the transposition reaction mix [25 μ l of 2X~ transposition reaction buffer from Nextera kit, 2.5 μ l of Nextera Tn5 transposase from Nextera kit, 16.5 μ l of PBS, 0.5 μ l of digitonin (1%), 0.5 μ l of Tween 20 (10%), and 5 μ l of nuclease-free water] and incubated at 37°C for 1 hour. Samples were purified using the Qiagen MinElute PCR Purification Kit. Transposed DNA was eluted in 10 μ l of elution buffer and subjected to 5 cycles of PCR before amplification using barcoded primers and NEBNext High-Fidelity PCR Master Mix. We determined the optimal total number of PCR cycles for each Omni-ATAC-seq

Materials and methods

library using qPCR. Omni-ATAC-seq libraries were purified using 1.8X~ volumes of AMPure XP beads to remove fragments below 100 bp and sequenced (50-bp single end) on an Illumina HiSeq 3000. The ATAC-seq samples were mapped against the mm9 mouse genome assembly using Bowtie with the option `-m 1` to discard those reads that could not be uniquely mapped to just one region. Mitochondrial reads were removed from each resulting map, and down-sampling was applied to obtain the same number of mapped fragments per sample. Box plots showing the ATAC-seq level distribution particular sets of genes were calculated by determining the average value at transcript level on the region ± 2 kbp around each TSS. The UCSC genome browser was used to generate the screenshots depicted in the manuscript.

1.7. Colony-forming assay

Single-cell HSC culture Lineage-depleted BM cells were stained as previously described for HSC. Single-cell sorting (FACSAria, BD) was performed in 96-well plates containing 100 μ l of filtered HSC expansion medium [StemSpan serum-free expansion medium, murine stem cell factor (300 ng/ml), murine interleukin-11 (20 ng/ml), and human Flt3L (1 ng/ml)]. The plates were incubated at 37°C with 5% CO₂, and medium was replaced every 7 days. Cell divisions were visually scored for 96 hours. At day 9, single wells were stained with PI (1 μ g/ml) and counted using FACSCanto II High Throughput Sampler. At day 14, cells were harvested, pooled, and stained with BM surface markers as previously described.

1.8. γ H2AX immunofluorescence staining.

Cells were directly sorted onto poly-lysine coated slides (100–500 cells per slide), incubated for 10 min, fixed with 4% FA for 10 minutes at room temperature, permeabilized in 0.15% Triton X-100 for 2 minutes at room temperature and blocked in 1% BSA/PBS overnight at 4°C. Slides were then incubated for 1 to 2 h at 37 °C in 1% BSA/PBS with anti-phospho-H2AX (Ser 139) (Millipore, 05636) at a 1/600 dilution. Slides were washed three times in PBS and incubated for 1 h at 37°C in PBS/1% BSA with A488-conjugated donkey anti-rabbit (Molecular Probes, A-21206). Images were acquired on a Leica TCS SPE inverted confocal microscope. The processing and quantification of the images was performed by Raúl Gómez Riera from the CRG Advance Light Microscopy unit. An average of 50 cells were scored from at three independent experiments.

2. *In vitro* experiments

2.1. Cell culture

HPC-7 cells were grown in Iscove's Modified Dulbecco's Medium (IMDM) containing Stem Cell Factor (SCF) (1/1000), 10% Serum (ES), B-Mercaptoethanol (1/100), 1% L-Glutamine&GlutaMAX supplement and 1% Penicillin-Streptomycin (P/S). Cells were cultured in suspension at a density of 0,15-0,8 x 10⁶ cells/ml in T75 flasks. All cultures were incubated at 37°C, 5% CO₂ and 95% humidity.

2.2. Lentiviral production

Materials and methods

To generate lentiviral particles, HEK293T cells were grown to 80% confluence in 10 cm plates and transfected with 10 µg of vector of interest, 1.25 µg of viral envelope vector (containing VSV-G), 8.75 µg of packaging vector p8.91 in a final volume of 500 µl 250 mM CaCl₂. An equal volume of 2XHBS was added dropwise to the mix while bubbling and then rested for 30 minutes. The following morning, fresh media was added and viral supernatants were collected at 48 and 72 hours post-transfection and filtered through a 0.45 µm filter. Lentiviral particles from HEK293T supernatants were precipitated at 4°C by centrifugation at 22000 RCF for 2 hours, re-suspended in appropriate media and stored at -80°C.

2.3. Viral transduction

PLKO.1 lentiviral vector harbouring shRNAs against PHF19 were previously produced in our laboratory and used to generate lentiviral particles as mentioned above. 2×10^6 cells HPC-7 cells were plated per well in 1 ml of medium in a non-treated 6-well plate. Lentiviral particles were added and cells were spininoculated (1000RCF, 90 minutes, 32°C) in presence of 1.5µl protamine sulfate (50mg/ml), followed by 2 hours incubation of 37°C and 5% CO₂. After 2 hours, cells were collected and resuspended in culture media at a density of 5×10^4 cells/ml. 48 hours after infection, HPC-7 cells were selected with 1µg/ml of puromycin for additional 72 hours.

2.4. qPCR shRNA knockdown analysis

To analyse shRNA knockdown efficiency, RNA was extracted from 1×10^6 HPC-7 cells with RNAeasy Mini Kit (Qiagen) according to the manufacture's protocol. RNA concentration was measured at

Nanodrop. Then cDNA was generated from 1µg of RNA with the First Strand cDNA Synthesis Kit (Fermentas) according to the manufacture's protocol. cDNA was diluted to 100µl of water and 2µl of sample were used for each RT-PCR reaction, using SYBR green (Roche), the corresponding primers (primer table) and the Roche LightCycler 480. Expression was normalized to the housekeeping gene RP0. Cells with more than a two-fold decrease of PHF19 expression comparing with control were considered for further experiments.

2.5. Cytospin

Cytofunnels, microscope slide, and filtercard were load on the centrifuge. 100,000-200,000 HPC-7 cells were resuspend in 200ul PBS, dispensed at the cytofunnels, and spin down 3 mins at 350 rpm. Slides were dry for 10 min at RT and stained according to Wright-Giemsa. Wright Stain solution was placed in a coplin glass and slides were introduced for 1 minute. Slides were washed in PBS and stained with Giemsa Stain dilute 1:20 with deionized water. Stain slides for 10 minutes. Rinse in deionized water. Air dry and evaluate at light microscope.

2.6. Cell growth

In order to assess growth rate of HPC-7 cells infected with shRNA empty or against PHF19, 2×10^5 cells were plated in a multi-6-well (M6W) plate and counted every two days. This was repeated three more times (8 days in total). Number of cells was counted using Neubauer chamber slides and trypan blue staining to exclude death cells.

2.7. RNA-seq

For HPC-7 one RNA replicate was obtained from 10^6 HPC-7 cells infected with shRNA empty or against PHF19. RNA extraction was performed using the RNeasy Mini Kit (Qiagen) following manufacturer's instructions. RNA-seq library preparation was performed by at the CRG Genomic Unit from 3 to 5 ng of total RNA. All samples were sequenced and processed as mentioned above.

2.8. Classic chromatin immunoprecipitation sequencing

For PHF19 ChIP HPC-7 cell were double-crosslinked. For the first crosslinking step, 10^8 HPC-7 cells were washed twice with PBS 1X at room temperature before incubating with 10 ml of PBS 1X freshly mixed with 40 μ l of ChIP CrossLink GOLD (Diagenode) for 30 minutes on a shaker a room temperature. Cells were then washed twice before proceeding to the second crosslinking step where HPC-7 cells were incubated in 10 ml of medium with 1% FA for 10 minutes on a shaker at room temperature. For MTF2 and H3K27me3 10^8 HPC-7 were crosslinked only once in 20 ml of medium with 1% FA. After two washes with ice cold PBS 1X, 10 ml of ice cold PBS 1X was added and cells were collected in 15 ml polystyrene Falcon tubes and centrifuged for 5 minutes at 1200 RCF at 4°C. The pellet was then washed with ice cold PBS 1X and either directly used or stored at -80°C. To lyse the cells, the pellet was incubated for 5 minutes on ice with 1 ml of Swelling Buffer (25 mM HEPES pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.1% NP-40, protease and phosphatase inhibitors). Nuclei were isolated using a Dounce homogenizer (10 ml, tight pestle) for 50 strokes and collected by centrifuging at 3000

RCF for 5 minutes at 4°C. Supernatant was discarded and nuclei were resuspended in 1.3 ml of Sonication Buffer (50 mM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Nadeoxycholate, 0.1% SDS, protease and phosphatase inhibitors). Nuclei were sonicated with a bioruptor at max output (30 seconds ON 30 seconds OFF) for around 40 cycles, centrifuged for 20 minutes at full speed at 4°C and supernatant was kept. Sonication quality was checked by reverse crosslinking an aliquot of 30 µl by adding 170 µl of PBS and incubating for 5 hours at 65°C shaking at 300 RCF, purifying the chromatin with a Qiagen PCR Purification Kit according to manufacturer's instructions and then running the eluted DNA on a 1% agarose gel. If DNA fragments were between 125 and 200 bp, we proceeded with the ChIP. ChIP was started from 60 µg of chromatin for PHF19, and 20 µg for MTF2 and H3K27me3, in a final volume of 500 µl in Sonication Buffer. 10% of the volume was kept as input. 20 µl per sample of previously washed Protein A/G dynabeads was added to the 500 µl, as well as 4 µg of antibody (Table M1). The mix was incubated overnight at 4°C on a rotating wheel. The next day, beads were washed for 5 minutes at 4°C with 1 ml of Sonication Buffer, followed by a similar wash with 1 ml of Wash Buffer A (50 mM HEPES pH 7.9, 500 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0,1% SDS, protease and phosphatase inhibitors), 1 ml of Wash Buffer B (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, protease and phosphatase inhibitors) and 1 ml of TE Buffer (10 mM Trish pH 8.0 and 1 mM EDTA). Immunoprecipitated chromatin was eluted by adding 100 µl Elution Buffer (TE Buffer supplemented

Materials and methods

with 1% SDS) and incubated 20 minutes at 65°C shaking at 1000 RCF. The elution step was repeated twice for a final volume of eluted material of 200 μ l. In order to purify the DNA from the eluted chromatin, samples and inputs were incubated 5 hours at 65°C shaking at 300 rpm and then DNA was purified with the Qiagen DNA Purification Kit according to manufacturer's instructions. ChIP-seq libraries were prepared at the CRG Genomic Unit and processed as mentioned above.

Antibodies list

ANTIBODY	QUANTITY	SOURCE	CATALOG N
γH2AX	1/600	Millipore	05636
H3K27ME3	3 μ l	Active Motif	39155
PHF19	5 μ l	Cell signaling	77271
MTF2	5 μ l	Protein Tech	16208-1-AP

Material and methods corresponding to CHAPTER II

1. Cell lines

1.1. HSJD primary cell lines

HSJD-DIPG models were developed from biopsies at diagnosis (Table MM1). As exception to this rule, HSJD-DIPG-007 cells were established from necropsy samples. Histone H3 K27M, ACVR1 and TP53 mutations were assessed by Sanger sequencing at HSJD, in the context of the molecular profiling associated to the clinical diagnosis and prognosis of these tumours. All DIPG cultures harboured the K27M mutation in H3.3 K27M. Mutations in the ACVR1 gene were detected in HSJD-DIPG-007 (R206H), and TP53 mutations were found in 3 of 7 cultures. HSJD paediatric High Grade Gliomas (pHGG) cultures were obtained from biopsies of hemispheric brain tumours diagnosed as astrocytoma's. One of them (HSJD-GBM-002) was an H3.3 G34R mutant tumour. No somatic mutations in H3K27M were found and TP53 mutations were detected in both tumours. Most cell cultures grew as spheroids. As exceptions to this rule, HSJD-DIPG-017, HSJD-DIPG-019 and HSJD-DIPG-021 cells grew attached to the flask as well as a significant proportion of HSJD-GBM-002 cells.

LAB CODE	SEX	AGE (Y)	HISTOLOGY AND ORIGIN OF SAMPLE	HISTONE MUTATION	TP53 MUTATION	ACVR1 MUTATION
HSJD-DIPG-007	M	9.9	Diffuse pons	H3F3A	WT	R206H
HSJD-DIPG-008	M	6.5	Diffuse pons	H3F3A	WT	WT
HSJD-DIPG-012	M	9.7	Diffuse pons	H3F3A	R273C	WT
HSJD-DIPG-014	F	8.2	Diffuse pons	H3F3A	WT	WT

HSJD-DIPG-017	M	8.0	Diffuse pons	H3F3A	R158P	WT
HSJD-DIPG-019	M	6.1	Diffuse pons	H3F3A	WT	WT
HSJD-DIPG-021	F	5.9	Diffuse pons	H3F3A	C135 Y	WT
HSJD-GBM-001	F	10.9	GBM right frontal	WT	G245 S	WT
HSJD-GBM-002	M	15.0	GBM left hemisphere	H3.3 G34R	P278T	WT

Table MMI. HSJD patient-derived *in vitro* models, relevant clinical data and molecular characterization

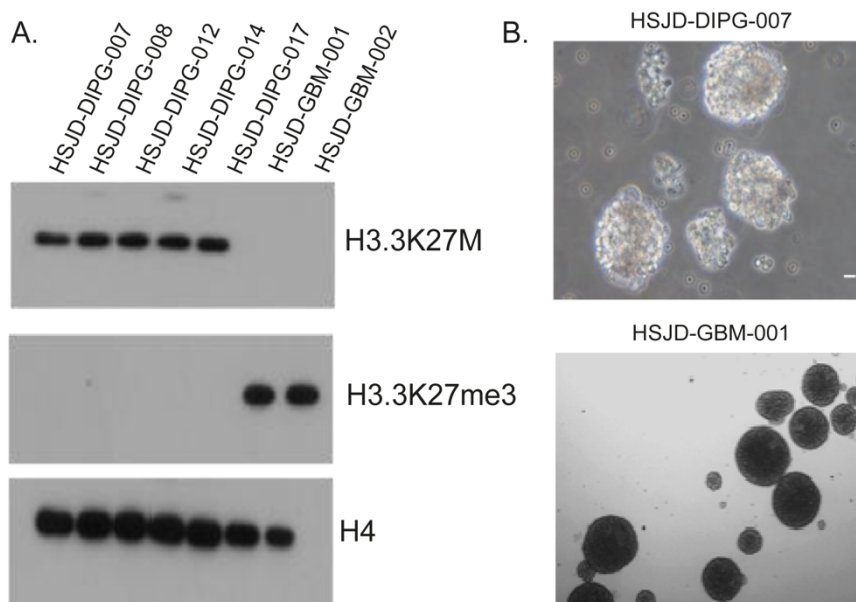


Figure MMI. (A) Westernblot analysis of H3.3K27M, H3.3K27me3 and H4 in different HSJD patient-derived primary cultures. (B) Light microscopy images of HSJD-DIPG-007 and HSJD-GBM-001 growing as tumor neurospheres (x10 magnification) (data produced by Dr Nagore Gené Olaciregui from HSJD [355]).

1.2. HSJD cell lines culture

All HSJD primary cultures were grown in a basal working stem cell media consisting of a 50:50 mixture of Dulbecco's Modified Eagles Medium: Nutrient Mixture F12 and Neurobasal A medium,

containing HEPES buffer solution 1 M, MEM sodium pyruvate solution 100 nM, MEM non-essential amino acids solution 10 mM, Glutamax-I supplement, and antibiotic-antimycotic. This basal working medium was prepared and stored up to one month, and previously to use was supplemented with 20 ng/mL recombinant human-EGF, 20 ng/mL recombinant human-FGF-154, 10 ng/mL recombinant human-PDGFAA, 10 ng/mL recombinant human-PDGFBB, B27 without vitamin A, and 0.02% heparin sulfate.

The cells were monitored for neurosphere formation and growth, with an average addition of fresh supplemented media every 3-5 days for the fastest cultures, HSJD-DIPG007, HSJD-DIPG008, HSJD-DIPG021, and HSJD-GBM001, and every 5-7 days for the others. Fresh media addition was performed after neurosphere decantation and supernatant removal. The fastest cultures were passaged approximately once per week by centrifuging in their original media at 0.4 g for 3 min, removing supernatant, and performing mechanical dissociation of spheroid pellet to obtain a single cell suspension that is resuspended in fresh medium and used to reset new flasks. The slowest growth cell lines were passaged once every two weeks. All cultures were incubated at 37°C, 5% CO₂ and 95% humidity.

2. CRISPR screening

2.1. Cell transduction - Generation of constitutive Cas9:BFP expressing stable cell lines

Lentiviral particles were produced as mentioned above. To generate stable constitutive Cas9:BFP glioblastoma cell lines, HSJD-

Materials and methods

DIPG007 and HSJD-GBM001 cells were first dissociated mechanically, resuspended on complete medium containing Cas9:BFP lentiviral particles and 4 μ g/ml of polybrene, then plated at 3x10⁵ cells/ml in 2ml per well of a 6-well plate, and finally spin infected by centrifuging for 90 minutes at 300 RCF. We infected a total of 2x10⁶ cells. After centrifuging, infected cells were incubated in the presence of lentivirus overnight. The following morning, cells were re-suspended and re-plated in suspension culture flasks with fresh growth medium. The rate of infection was determined 48 hours and 72h post-infection by measuring the levels of BFP in these cell lines using flow cytometry on a FACSDiva (BD Biosciences) followed by analysis using FlowJo (<https://www.flowjo.com/>). After confirming the presence of >40% BFP+ cells, HSJD-DIPG007 and HSJD-GBM001 cells were selected with 6 μ g/mL blasticidin for 7-10 days. To determine the selection efficiency the percentage of BFP+ cells were monitored by flow cytometry until reach more than 70% BFP+ cells.

2.2. CRISPR screening

To perform the CRISPR dropout screening, HSJD-DIPG007 and HSJD-GBM001 cells expressing Cas9:BFP were spin infected with ARUNA sgRNA lentiviral library by first dissociating spheroids mechanically, resuspending on complete medium containing ARUNA library lentiviral particles and 4 μ g/ml of polybrene, and plating 2x10⁵ cells/ml in 2ml per well of a 6-well plate. After titration, we infected a total of 24x10⁶ cells using a MOI of \sim 0.2 (1 μ l ARUNA lentivirus per 10000 μ l medium from our concentrated virus

batch) in order to achieve a final amount of 5×10^6 ARUNA positive cells. This provided an initial representation of 1250X per sgRNA. After centrifuging, infected cells were incubated in the presence of lentivirus overnight. The following morning, cells were re-suspended and re-plated in suspension culture flasks with fresh growth medium. The adequacy of the final MOI was determined 48 hours and 72h post-infection by measuring the levels of GFP in these cell lines using flow cytometry as mentioned above. After confirming the presence of less than 30% GFP+ cells, HSJD-DIPG007 and HSJD-GBM001 cells were selected with 1/10.000 puromycin (stock 10 mg/mL) for 2-4 days until generate stable ARUNA+ cell lines. To determine the selection efficiency the percentage of GFP+ cells were monitored by flow cytometry until reach more than 90% GFP+ cells. After selection, reference samples were collected to quantify the initial sgRNA abundance (Time 0). The remaining infected HSJD-DIPG007 and HSJD-GBM001 cells were then passaged at 1250X representation for each sgRNA (5×10^6 cells) and grown for an additional four weeks, feeding the cells every 3-5 days with fresh media. After outgrowth for one month, surviving cells were harvested, genomic DNA was isolated from all samples and the sgRNA sequences were amplified by PCR before sequencing on a HiSeq2500 sequencer (Illumina).

2.3. PCR amplification of sgRNA cassette and sequencing

Genomic DNA was extracted using Qiagen DNA Purification Kit according to manufacturer's instructions. To maintain >500X sgRNA library representation (2×10^6 cells, 6pg per cell, 1200ng total) , 24

Materials and methods

independent PCR reactions were used to amplify the sgRNA cassette, which were amplified for 20 cycles (with 500 ng of starting gDNA using the Phusion Hot Start II DNA polymerase (Thermo Scientific #F-565). PCR products were pooled and subjected to illumina MiSeq library construction and sequencing, as described above. The sequence data were trimmed to contain only the sgRNA sequence then were mapped to the reference sgRNA library without allowing any mismatches. The read counts were then calculated for each individual sgRNA.

Primer Name	Primer Sequence
ARUNA-seq50bp_FW	acactagctacgtgtagcagtNNNNNNNNNCTTGTGGA AAGGACGAAACACCG
ARUNA-seq50bp_RV	CagacgtgtgctcttccgatctTCTCTAGGCACCGGTTTC AAT

3. iPOTD assay

3.1. Cell growth assay

This assay was performed to study the potential antiproliferative activity of EdU against HSJD-DIPG007 and HSJD-GBM001 cells. Tumorspheres were mechanically dissociated to obtain single-cell suspensions before seeding and incubated with or without 0.1 μ M EdU, 2x10⁵ cells/ml in T75 flask. Every day up to 3 days 2ml aliquot was mechanically dissociated and number of cells was counted using Neubauer chamber slides and trypan blue staining to count the number of non-viable cells.

3.2. CFSE Cell cycle

Dye dilution CFSE CellTrace™ Violet reagent (Thermo Fisher, C34554) was used to follow HSJD-DIPG007 and HSJD-GBM001 cell proliferation. 10^6 cells/ml were seeded on 10 ml on T75 and 10 μ l of CellTrace™ stock solution in DMSO (prepared following manufacturer's instructions) were added. Cells were incubated for 20 minutes at room temperature or 37°C, protected from light. After 20 minutes cells were centrifugated and resuspend in fresh pre-warmed complete culture medium. The dye dilution was determined at time 0 and every 24 hours using flow cytometry on a FACSDiva (BD Biosciences) followed by analysis using FlowJow. The discrete peaks in this histogram represent successive generations of live cells.

3.3. iPOTD and whole protein extract sample processing

iPOTD protocol was performed as described [91]. HSJD-DIPG007 and HSJD-GBM001 cells were pulsed 48 hours with 0.1 μ M of EdU (Invitrogen). Subsequently, the cells were fixed in 1% PFA for 10 min at RT and quenched with 0.125 mM glycine (pH 7) for 5 min at RT. Cells were harvested, pelleted by centrifugation (720RCF, 10 min at 4°C), and lysed for 30 min at 4°C in lysis buffer A [10 mM Hepes (pH 7.9), 10 mM KCl, 1.5 mM MgCl₂, 0.34 M sucrose, 10% (v/v) glycerol, 1 mM DTT, 10 mM β -glycerol phosphate, 1 mM sodium orthovanadate, PIC (Roche), and 0.1% (v/v) Triton X-100]. Nuclei were pelleted by centrifugation (1300RCF, 4 min at 4°C), washed with PBS + PIC, and subjected to Click reaction for 30 min at RT with 0.2 mM biotin-azide (Invitrogen). In Click, an organic azide reacts with a terminal acetylene, and the nucleotide-exposed

Materials and methods

ethynyl residue of EdU is derivatized by a copper-catalyzed cycloaddition reaction, to form a covalent bond between EdU and biotin. Nuclei were pelleted by centrifugation (1300RCF, 4 min at 4°C), washed with PBS + PIC, and suspended in shearing buffer [20 mM tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 0.1% (w/v) SDS, 0.5% (w/v) sodium deoxycolate, 1% (v/v) Triton X-100, 10 mM β -glycerol phosphate, 1 mM sodium orthovanadate, and PIC]. Nuclei suspension was extensively sonicated (Bioruptor Diagenode) for 20 cycles of 10 min at high intensity (30 s on/30 s off pulses). Lysates were centrifuged (15000RCF, 10 min at 4°C), and supernatant was collected as input for further analysis. From this lysates 50 μ g were additionally send to analyse total proteome. To analyze shearing efficiency, 5% of input material was reverse cross-linked by incubation overnight at 65°C with 250 mM NaCl and then digested with proteinase K (0.1 mg/ml) for 1 hour at 55°C. DNA was purified using the PCR purification kit from Qiagen, following the manufacturer's instruction. For chromatin capture, input material was measured using Bradford assay (see below) and equal amounts of protein extract was took and diluted to a same volume with blocking buffer [1% Triton X-100, 2 mM EDTA (pH 8), 150 mM NaCl, 20 mM tris-HCl (pH 8), 20 mM β -glycerol phosphate, 2 mM sodium orthovanadate, PIC, and salmon sperm DNA (10 mg/ml)] and then incubated with preblocked with salmon sperm DNA Dynabead M-280 streptavidin (Invitrogen) for 30 min at 4°C. Beads were washed twice with blocking buffer (without salmon sperm DNA), twice with high-salt blocking buffer (containing 500 mM NaCl), and once with tris-EDTA buffer (pH 8). Beads were suspended in 75 μ l

of RapiGestTM SF (Water, 186001860) and boiled for 20 min at 95°C, and the supernatant was collected for analysis by dot blot and MS. The whole eluate was sent to analysis at the CRG Proteomic unit.

3.4. Bradford assay

Quantification of protein concentration by Bradford assay was performed using Thermo ScientificTM PierceTM BCA Protein Assay Kit (10741395) following manufacture's instructions. Proteins were diluted to 1X, and a standard curve was prepared with bovine serum albumin (BSA). 10 µl of the cell extract lysate was mixed with 200 µl of the freshly-prepared Bradford solution and incubated for 30 min at room temperature. Absorbance was measured in a photo-spectrometer at a wavelength of 595 nm. The protein concentration was calculated based on the standard curve.

3.5. iPOTD samples LCMS analysis

Samples were processed at the CRG Proteomics unit. Samples were analyzed in an Orbitrap Eclipse (Thermo Fisher Scientific) with a chromatographic gradient of 90 minutes. As a quality control BSA controls were digested in parallel and ran between each of the samples to avoid carryover and assess the instrument performance. Eluted proteins were reduced, alkylated, and digested to peptide mixes according to the filter-aided sample preparation method. Acquired data were analyzed using the Proteome Discoverer software suite (v2.3, Thermo Fisher Scientific) and were searched against the SP Human database (Feb.2020), using the search algorithm Mascot v2.6 (<http://www.matrixscience.com/>). The

identified peptides were filtered using false discovery rate (FDR) < 5%.

3.6. DEqMS analysis

Analysis of specific chromatin interactors was carried out using protein abundance values, DEqMS, as previously described [329], and using a threshold corresponding to $p\text{-value} < 0.05$. Paired comparisons was performed between each Edu incubated and non-incubated replicate, and the results were compared among all replicates. NA values were substituted by minimal absorbance value.

3.7. Subcellular fractionation

Subcellular fractionation protocol was performed as described [86]. 30 millions cells were pelleted, resuspended in 10X of packed cell volume (usually 500 μl) of ice-cold Buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl_2 , 0.34 M sucrose, 10% glycerol, 1 mM DTT, 0.1% Triton X-100, protease and phosphatase inhibitors), and incubated for 10 min on ice. An aliquot (50 μl) of the total fraction was taken (fraction T). After incubation, samples were centrifuged at 1300 RCF for 5 minutes at 4°C. An aliquot (50 μl) of the supernatant was taken for the cytoplasmic fraction (fraction C). The pellet was washed with 5 volumes of buffer A and then resuspended in 1X volumen Buffer A (50 μl). An aliquot (5 μl) for the total nuclear fraction was taken (fraction N). The sample was then diluted in 10x volume of Buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, and protease inhibitors), briefly vortexed and incubated for 30 minutes on ice. Samples were centrifuged at 1700 RCF for 5 minutes at 4°C. An aliquot (50 μl) of the supernatant was taken for

the nucleoplasmic fraction (fraction Np). Pellets were washed in 5X volume buffer B and then resuspended in 10X volume (500 μ l) B-SDS 1X Lysis Buffer (50 mM Tris-Cl [pH 7.5], 2 mM EDTA, 2% SDS) to form the chromatin fraction (fraction Chr). All fractions (except the chromatin fraction) were then mixed with an equal amount of B-SDS 2X lysis buffer, and all were boiled for 10 min. The chromatin fraction was finally sonicated for 5 minutes at high potency (30 s ON, 30 s OFF) until clarified (10-40 cycles).

3.8. Western blot

Whole cell extracts for Western blot analysis were prepared in lysis buffer IP300 (buffer). Lysates were incubated for 5 min on ice and then sonicated for 10 cycles (30 sec ON/30 sec OFF) in a Bioruptor (Diagenode). Cell extracts were centrifuged for 25 min at maximum speed at 4°C. Protein extracts (supernatant fractions) were quantified by Bradford assay (see above). Protein (usually around 40 μ g) was diluted with 5x Laemli buffer and heated for 5 min at 98°C. Samples were analysed by SDS-PAGE using 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels, 15-well, 15 μ l (Biorad, #4561086) at 100 V. Proteins were transferred onto nitrocellulose membranes at 250 mA for 70 min at 4°C in Transfer buffer (25 mM Tris-HCl pH 8.3, 200 mM glycine, 20% v/v methanol). Protein transfer was checked by Ponceau S (Sigma) staining. Transferred membranes were blocked with 10% w/v milk in TBS-Tween (10 mM Tris-HCl pH 7.5, 100 mM NaCl, and 0.1% Tween-20) for 30 minutes with rotation at room temperature. Blocked membranes were incubated overnight with the primary antibody (in with 5% w/v milk in TBS-Tween) at 4°C with rotation (see primary antibody table).

Materials and methods

The next day, membranes were washed twice for 5 min with TBS-Tween followed by incubation with the secondary antibody (see secondary antibody table) diluted in TBS-Tween, for 1 h at room temperature. After two washes of 15 min with TBS-Tween (at room temperature), proteins were detected by an enhanced chemiluminescence reagent (Amersham ECL Prime Western Blotting Detection Reagents). Signal was measured used iBright system and analyze by iBright analysis software.

Primary antibodies list

Primary Antibody	Dilution	Source	Catalog N
████████	██████	██████	████████
██████	██████	██████	████████
██████	██████	██████	████████
████████	██████	██████	████████
████████	██████	██████	████████
██████	██████	██████	████████
██████	██████	██████	████████
████████	██████	██████	████████
██████	██████	██████	████████
████████	██████	██████	████████

Secondary antibodies list

Secondary Antibody	Dilution	Source	Catalog N
anti-mouse-horseradish	1/2000	Dako	P0260

anti-rabbit-horseradish	1/2000	Dako	P0448
-------------------------	--------	------	-------

4. Lentiviral short-hairpin RNA (shRNA) knockdown

4.1. shRNA design

PLKO.1 lentiviral vector was used for the expression of shRNAs against the proteins candidates. The following oligos were ordered from SIGMA and used to clone shRNA into pLKO.1 vector specific for:

sgRNA name	sgRNA primary sequence
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]

Materials and methods

Then oligos were annealed by incubating the following reaction at 95°C for 5 minutes then followed to slowly cooling at room temperature: 5ul of 20uM Forward oligo + 5ul of 20uM Reverse oligo + 5ul 10X NEB Buffer 2 + 35ul ddH₂O. The plasmid of insertion was digested with AgeI and EcoRI and purified by Qiagen Gel Extraction kit. 20 ng of digested and purified plasmid vector was ligated with 2ul of annealed oligos using T4 DNA ligase from New England Biolabs (NEB).

Then 2µl of ligations was transformed into competent bacterial cells and plated in LB agar plates with 100ug/ml of ampicillin resistance. Colonies grew overnight in the plates. Then colonies were picked and grown overnight in 5ml of LB medium containing 100ug/ml of ampicillin. The plasmid was purified using GeneAll MiniPrep Plasmid purification kit. Then 50ng of purified plasmid was sent for sequencing and positive colonies were grown in large bacterial cultures and plasmids were purified using Invitrogen Pureink HiPure Plasmid Filter DNA purification kit.

4.2. Protein extracts preparation

Whole cell extracts for Western blot analysis were prepared in lysis buffer IP300 (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 10% glycerol, 0.2% NP-40). Lysates were incubated for 5 min on ice and then sonicated 5 cycles (30 sec ON/30 sec OFF) in a Bioruptor (Diagenode). Cell extracts were centrifuged for 25 min at maximum speed at 4°C. Protein concentration was quantified by Bradford assay (Bio-Rad) according to manufacturer instructions. 40 µg of protein were loaded into SDS-PAGE gel for further analysis.

4.3. MTS assay cell viability

HSJD cells were plated on 96-well plates in triplicates at a density of 5,000 cells per well on 200 μ l of complete medium. The MTS assay (CellTiter 96® AQueous One Solution Cell Proliferation Assay (MTS), Promega, G3582) was used to determine the percentage of viable cells after each infection or inhibitor treatment and compared to control wells. Assays were performed by adding 40 μ l of the CellTiter 96® AQueous One Solution Reagent directly to culture wells, incubating for 1–4 hours and then recording absorbance at 490nm with a 96-well plate reader. The quantity of formazan product as measured by the amount of 490nm absorbance is directly proportional to the number of living cells in culture.

4.3.1. [REDACTED] inhibition

HSJD cells were plated on 96-well plates as mentioned above. 0.3 μ l of DMSO of 5 μ M of [REDACTED] Cat. No. 4485 were added to the 200 μ l of medium.

INDEX

Abbreviation	Explanation
5mC	5-carbon of the cytosine pyrimidine ring
aHSC	active hematopoietic stem cells
AML	Acute myeloid leukaemia
ATAC	Assay for Transposase-Accessible Chromatin
ATRT	atypical teratoid rhabdoid tumours
BFP	Blue Fluorescent Protein
BM	Bone Marrow
C-BERST	Biotinylation at genomic elements by restricted spatial tagging
CGI	CpG islands
CHD	chromodomain helicase DNA-binding
ChEP	Chromatin Enrichment for Proteomics
CHESS	chromatin enriching salt preparations
ChIP	Chromatine immunoprecipitation
cKO	Conditiona knockout
CLP	Common lymphoid progenitors
CMP	Common myeloid progenitors
CRC	chromatin remodeling complexes
CTCF	CCCTC-binding factor
CUT&RUN	Cleavage Under Targets and Release Using Nuclease
CUT&TAG	Cleavage Under Targets & Tagmentation
D-CAP	differential chromatin-associated proteins
DC	Dendritic Cell
DE	Differential express
dHSC	Dormant Hematopoietic Stem Cells
DIPG	Diffuse Intrinsic Pontine Gliomas
DKO	Double knockout
DLBCL	Diffuse large B-cell lymphomas
Dm-ChP	DNA-mediated chromatin pulldown
Edu	thymidine analogue 5-ethynyl-2'-deoxyuridine

FACS	Fluorescence activated cells sorting
FAIRE	Formaldehyde Assisted Isolation of Regulatory Elements
FL	follicular lymphomas
GBM	glioblastoma multiforme
GFP	Green Fluorescent Protein
GMP	Granulocyte-monocyte Progenitors
GoF	Gain of function
GSEA	gene set enrichment analysis
hESC	Human Embryonic stem cell
HOX	homeobox genes
HSC	Hematopoietic Stem Cell
HSJD	Hospital Sant Joan de Deu
INO80	inositol-requiring 80
iPODT	identification of proteins on Total DNA
iPOND	isolation of proteins on nascent DNA
ISWI	imitation switch
LAD	lamin-associated domains
LMPP	Lymphoid-Primed Multipotent Progenitors
lncRNA	Long non coding RNA
LoF	Loss of function
LSC	Leukemic Stem Cell
LSK	Lin+ Sca1- c-Kit+ Cells
MDS	Myeloid dyspla
MEP	Megakaryocyte/Erythrocyte Progenitor
mESC	Mouse Embryonic Stem Cells
MOI	multiplicity of infection
MPNST	malignant peripheral nerve sheath tumours
MPP	Multipotent Progenitors
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3 carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H tetrazolium, inner salt
NB	Neuroblastoma
NK	Natural Killer

NPC	neural precursor cells
NSC	neuronal stem cells
OPC	oligodendrocyte precursors
PcG	Polycomb Group of proteins
PCR	Polymerase chain reaction
PI	propidium iodide
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PTM	Post translational modification
SCC	oesophageal squamous cell carcinomas
ST-HSC	Short term hematopoietic stem cells
████████	██
T-ALL	T-cell lymphoblastic leukaemia
TAD	Topological associated domains
TDG	thymine DNA glycosylase
TF	Transcription Factor
TSS	Transcription Start Site
WT	Wild type

References

References

1. Ummethum, H. and S. Hamperl, *Proximity Labeling Techniques to Study Chromatin*. Front Genet, 2020. **11**: p. 450.
2. Espejo, I., L. Di Croce, and S. Aranda, *The changing chromatinome as a driver of disease: A panoramic view from different methodologies*. Bioessays, 2020. **42**(12): p. e2000203.
3. Cramer, P., *A tale of chromatin and transcription in 100 structures*. Cell, 2014. **159**(5): p. 985-994.
4. Aranda, S., G. Mas, and L. Di Croce, *Regulation of gene transcription by Polycomb proteins*. Sci Adv, 2015. **1**(11): p. e1500737.
5. Kornberg, R.D., *Chromatin structure: a repeating unit of histones and DNA*. Science, 1974. **184**(4139): p. 868-71.
6. Bonnet, J., et al., *Quantification of Proteins and Histone Marks in Drosophila Embryos Reveals Stoichiometric Relationships Impacting Chromatin Regulation*. Dev Cell, 2019. **51**(5): p. 632-644 e6.
7. Sun, H.B., J. Shen, and H. Yokota, *Size-dependent positioning of human chromosomes in interphase nuclei*. Biophys J, 2000. **79**(1): p. 184-90.
8. Klemm, S.L., Z. Shipony, and W.J. Greenleaf, *Chromatin accessibility and the regulatory epigenome*. Nat Rev Genet, 2019. **20**(4): p. 207-220.
9. Allshire, R.C. and H.D. Madhani, *Ten principles of heterochromatin formation and function*. Nat Rev Mol Cell Biol, 2018. **19**(4): p. 229-244.
10. Saksouk, N., E. Simboeck, and J. Dejardin, *Constitutive heterochromatin formation and transcription in mammals*. Epigenetics Chromatin, 2015. **8**: p. 3.
11. Stewart-Morgan, K.R., N. Reveron-Gomez, and A. Groth, *Transcription Restart Establishes Chromatin Accessibility after DNA Replication*. Mol Cell, 2019. **75**(2): p. 284-297 e6.
12. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome*. Nature, 2012. **489**(7414): p. 75-82.
13. Gloss, B.S. and M.E. Dinger, *Realizing the significance of noncoding functionality in clinical genomics*. Exp Mol Med, 2018. **50**(8): p. 1-8.
14. Gasperini, M., J.M. Tome, and J. Shendure, *Towards a comprehensive catalogue of validated and target-linked human enhancers*. Nat Rev Genet, 2020. **21**(5): p. 292-310.

15. Hnisz, D., et al., *Super-enhancers in the control of cell identity and disease*. Cell, 2013. **155**(4): p. 934-47.
16. Lambert, S.A., et al., *The Human Transcription Factors*. Cell, 2018. **172**(4): p. 650-665.
17. Kagey, M.H., et al., *Mediator and cohesin connect gene expression and chromatin architecture*. Nature, 2010. **467**(7314): p. 430-5.
18. Lambert, S.A., et al., *The Human Transcription Factors*. Cell, 2018. **175**(2): p. 598-599.
19. Lee, T.I. and R.A. Young, *Transcriptional regulation and its misregulation in disease*. Cell, 2013. **152**(6): p. 1237-51.
20. Alarcon, C., et al., *Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways*. Cell, 2009. **139**(4): p. 757-69.
21. Hamperl, S. and K.A. Cimprich, *Conflict Resolution in the Genome: How Transcription and Replication Make It Work*. Cell, 2016. **167**(6): p. 1455-1467.
22. Iwafuchi-Doi, M., et al., *The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation*. Mol Cell, 2016. **62**(1): p. 79-91.
23. Brahma, S. and S. Henikoff, *Epigenome Regulation by Dynamic Nucleosome Unwrapping*. Trends Biochem Sci, 2020. **45**(1): p. 13-26.
24. Narlikar, G.J., R. Sundaramoorthy, and T. Owen-Hughes, *Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes*. Cell, 2013. **154**(3): p. 490-503.
25. Hota, S.K. and B.G. Bruneau, *ATP-dependent chromatin remodeling during mammalian development*. Development, 2016. **143**(16): p. 2882-97.
26. Clapier, C.R., et al., *Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes*. Nat Rev Mol Cell Biol, 2017. **18**(7): p. 407-422.
27. Piquet, S., et al., *The Histone Chaperone FACT Coordinates H2A.X-Dependent Signaling and Repair of DNA Damage*. Mol Cell, 2018. **72**(5): p. 888-901 e7.
28. Clouaire, T. and G. Legube, *A Snapshot on the Cis Chromatin Response to DNA Double-Strand Breaks*. Trends Genet, 2019. **35**(5): p. 330-345.

References

29. Khare, S.P., et al., *HlStome--a relational knowledgebase of human histone proteins and histone modifying enzymes*. *Nucleic Acids Res*, 2012. **40**(Database issue): p. D337-42.
30. Huang, H., et al., *SnapShot: histone modifications*. *Cell*, 2014. **159**(2): p. 458-458 e1.
31. Chan, J.C. and I. Maze, *Nothing Is Yet Set in (Hi)stone: Novel Post-Translational Modifications Regulating Chromatin Function*. *Trends Biochem Sci*, 2020. **45**(10): p. 829-844.
32. Zhao, Y. and B.A. Garcia, *Comprehensive Catalog of Currently Documented Histone Modifications*. *Cold Spring Harb Perspect Biol*, 2015. **7**(9): p. a025064.
33. Sidoli, S., et al., *A mass spectrometry-based assay using metabolic labeling to rapidly monitor chromatin accessibility of modified histone proteins*. *Sci Rep*, 2019. **9**(1): p. 13613.
34. Simithy, J., S. Sidoli, and B.A. Garcia, *Integrating Proteomics and Targeted Metabolomics to Understand Global Changes in Histone Modifications*. *Proteomics*, 2018. **18**(18): p. e1700309.
35. Ji, X., et al., *Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions*. *Proc Natl Acad Sci U S A*, 2015. **112**(12): p. 3841-6.
36. Jenuwein, T. and C.D. Allis, *Translating the histone code*. *Science*, 2001. **293**(5532): p. 1074-80.
37. Barth, T.K. and A. Imhof, *Fast signals and slow marks: the dynamics of histone modifications*. *Trends Biochem Sci*, 2010. **35**(11): p. 618-26.
38. Gates, L.A., C.E. Foulds, and B.W. O'Malley, *Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle*. *Trends Biochem Sci*, 2017. **42**(12): p. 977-989.
39. Zhang, Y., et al., *Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail*. *Nucleic Acids Res*, 2010. **38**(13): p. 4246-53.
40. Yin, Y., et al., *Impact of cytosine methylation on DNA binding specificities of human transcription factors*. *Science*, 2017. **356**(6337).
41. Luo, C., P. Hajkova, and J.R. Ecker, *Dynamic DNA methylation: In the right place at the right time*. *Science*, 2018. **361**(6409): p. 1336-1340.


42. Wu, X. and Y. Zhang, *TET-mediated active DNA demethylation: mechanism, function and beyond*. Nat Rev Genet, 2017. **18**(9): p. 517-534.
43. Du, Q., et al., *Methyl-CpG-binding domain proteins: readers of the epigenome*. Epigenomics, 2015. **7**(6): p. 1051-73.
44. Cheng, T.L., et al., *Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain*. Sci Rep, 2017. **7**: p. 42790.
45. Sharifi-Zarchi, A., et al., *DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism*. BMC Genomics, 2017. **18**(1): p. 964.
46. Nott, A., et al., *Histone deacetylase 3 associates with MeCP2 to regulate FOXO and social behavior*. Nat Neurosci, 2016. **19**(11): p. 1497-1505.
47. Kavalali, E.T., E.D. Nelson, and L.M. Monteggia, *Role of MeCP2, DNA methylation, and HDACs in regulating synapse function*. J Neurodev Disord, 2011. **3**(3): p. 250-6.
48. Zheng, H. and W. Xie, *The role of 3D genome organization in development and cell differentiation*. Nat Rev Mol Cell Biol, 2019. **20**(9): p. 535-550.
49. Szabo, Q., F. Bantignies, and G. Cavalli, *Principles of genome folding into topologically associating domains*. Sci Adv, 2019. **5**(4): p. eaaw1668.
50. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome*. Science, 2009. **326**(5950): p. 289-93.
51. Wang, S., et al., *Spatial organization of chromatin domains and compartments in single chromosomes*. Science, 2016. **353**(6299): p. 598-602.
52. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions*. Nature, 2012. **485**(7398): p. 376-80.
53. Dixon, J.R., D.U. Gorkin, and B. Ren, *Chromatin Domains: The Unit of Chromosome Organization*. Mol Cell, 2016. **62**(5): p. 668-80.
54. Ganji, M., et al., *Real-time imaging of DNA loop extrusion by condensin*. Science, 2018. **360**(6384): p. 102-105.
55. Reuter, J.A., D.V. Spacek, and M.P. Snyder, *High-throughput sequencing technologies*. Mol Cell, 2015. **58**(4): p. 586-97.

References

56. Hawkins, R.D., G.C. Hon, and B. Ren, *Next-generation genomics: an integrative approach*. Nat Rev Genet, 2010. **11**(7): p. 476-86.
57. Tsompana, M. and M.J. Buck, *Chromatin accessibility: a window into the genome*. Epigenetics Chromatin, 2014. **7**(1): p. 33.
58. Rizzo, J.M. and S. Sinha, *Analyzing the global chromatin structure of keratinocytes by MNase-seq*. Methods Mol Biol, 2014. **1195**: p. 49-59.
59. Song, L. and G.E. Crawford, *DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells*. Cold Spring Harb Protoc, 2010. **2010**(2): p. pdb prot5384.
60. Buenrostro, J.D., et al., *ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide*. Curr Protoc Mol Biol, 2015. **109**: p. 21 29 1-21 29 9.
61. Corces, M.R., et al., *An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues*. Nat Methods, 2017. **14**(10): p. 959-962.
62. Park, P.J., *ChIP-seq: advantages and challenges of a maturing technology*. Nat Rev Genet, 2009. **10**(10): p. 669-80.
63. Orlando, V., *Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation*. Trends Biochem Sci, 2000. **25**(3): p. 99-104.
64. Schmidt, D., et al., *ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions*. Methods, 2009. **48**(3): p. 240-8.
65. Thomas, R., et al., *Features that define the best ChIP-seq peak calling algorithms*. Brief Bioinform, 2017. **18**(3): p. 441-450.
66. Skene, P.J. and S. Henikoff, *An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites*. Elife, 2017. **6**.
67. Kaya-Okur, H.S., et al., *CUT&Tag for efficient epigenomic profiling of small samples and single cells*. Nat Commun, 2019. **10**(1): p. 1930.
68. Kaya-Okur, H.S., et al., *Efficient low-cost chromatin profiling with CUT&Tag*. Nat Protoc, 2020. **15**(10): p. 3264-3283.

69. Hainer, S.J., et al., *Profiling of Pluripotency Factors in Single Cells and Early Embryos*. Cell, 2019. **177**(5): p. 1319-1329 e11.
70. Cokus, S.J., et al., *Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning*. Nature, 2008. **452**(7184): p. 215-9.
71. Gao, F., et al., *Integrated detection of both 5-mC and 5-hmC by high-throughput tag sequencing technology highlights methylation reprogramming of bivalent genes during cellular differentiation*. Epigenetics, 2013. **8**(4): p. 421-30.
72. Gouil, Q. and A. Keniry, *Latest techniques to study DNA methylation*. Essays Biochem, 2019. **63**(6): p. 639-648.
73. Gavrilov, A., et al., *Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification*. Methods Mol Biol, 2009. **567**: p. 171-88.
74. Jia, R., et al., *Novel insights into chromosomal conformations in cancer*. Mol Cancer, 2017. **16**(1): p. 173.
75. Splinter, E., et al., *Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation*. Methods, 2012. **58**(3): p. 221-30.
76. van Berkum, N.L. and J. Dekker, *Determining spatial chromatin organization of large genomic regions using 5C technology*. Methods Mol Biol, 2009. **567**: p. 189-213.
77. van Berkum, N.L., et al., *Hi-C: a method to study the three-dimensional architecture of genomes*. J Vis Exp, 2010(39).
78. Krietenstein, N., et al., *Ultrastructural Details of Mammalian Chromosome Architecture*. Mol Cell, 2020. **78**(3): p. 554-565 e7.
79. Buenrostro, J.D., et al., *Single-cell chromatin accessibility reveals principles of regulatory variation*. Nature, 2015. **523**(7561): p. 486-90.
80. Grosselin, K., et al., *High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer*. Nat Genet, 2019. **51**(6): p. 1060-1066.
81. Farlik, M., et al., *Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics*. Cell Rep, 2015. **10**(8): p. 1386-97.
82. Nagano, T., et al., *Single-cell Hi-C reveals cell-to-cell variability in chromosome structure*. Nature, 2013. **502**(7469): p. 59-64.

References

83. Gauchier, M., et al., *Purification and enrichment of specific chromatin loci*. Nat Methods, 2020. **17**(4): p. 380-389.
84. Gao, X.D., et al., *C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2*. Nat Methods, 2018. **15**(6): p. 433-436.
85. van Mierlo, G. and M. Vermeulen, *Chromatin Proteomics to Study Epigenetics - Challenges and Opportunities*. Mol Cell Proteomics, 2021. **20**: p. 100056.
86. Mendez, J. and B. Stillman, *Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis*. Mol Cell Biol, 2000. **20**(22): p. 8602-12.
87. Alajem, A., et al., *Differential association of chromatin proteins identifies BAF60a/SMARCD1 as a regulator of embryonic stem cell differentiation*. Cell Rep, 2015. **10**(12): p. 2019-31.
88. Federation, A.J., et al., *Highly Parallel Quantification and Compartment Localization of Transcription Factors and Nuclear Proteins*. Cell Rep, 2020. **30**(8): p. 2463-2471 e5.
89. Kustatscher, G., et al., *Chromatin enrichment for proteomics*. Nat Protoc, 2014. **9**(9): p. 2090-9.
90. Sirbu, B.M., et al., *Analysis of protein dynamics at active, stalled, and collapsed replication forks*. Genes Dev, 2011. **25**(12): p. 1320-7.
91. Aranda, S., et al., *Chromatin-Bound Proteome Profiling by Genome Capture*. STAR Protoc, 2020. **1**(1): p. 100014.
92. Palozola, K.C., J. Lerner, and K.S. Zaret, *A changing paradigm of transcriptional memory propagation through mitosis*. Nat Rev Mol Cell Biol, 2019. **20**(1): p. 55-64.
93. 
94. Schuettengruber, B., et al., *Genome Regulation by Polycomb and Trithorax: 70 Years and Counting*. Cell, 2017. **171**(1): p. 34-57.
95. Piunti, A. and A. Shilatifard, *Epigenetic balance of gene expression by Polycomb and COMPASS families*. Science, 2016. **352**(6290): p. aad9780.

96. Hojfeldt, J.W., et al., *Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity*. Mol Cell, 2019. **76**(3): p. 423-436 e3.
97. Healy, E., et al., *PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation*. Mol Cell, 2019. **76**(3): p. 437-452 e6.
98. Loubiere, V., A.M. Martinez, and G. Cavalli, *Cell Fate and Developmental Regulation Dynamics by Polycomb Proteins and 3D Genome Architecture*. Bioessays, 2019. **41**(3): p. e1800222.
99. Chammas, P., I. Mocavini, and L. Di Croce, *Engaging chromatin: PRC2 structure meets function*. Br J Cancer, 2020. **122**(3): p. 315-328.
100. Holoch, D. and R. Margueron, *Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity*. Trends Biochem Sci, 2017. **42**(7): p. 531-542.
101. Klose, R.J., et al., *Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins*. PLoS Genet, 2013. **9**(8): p. e1003717.
102. Laugesen, A., J.W. Hojfeldt, and K. Helin, *Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation*. Mol Cell, 2019. **74**(1): p. 8-18.
103. Lee, C.H., et al., *Distinct Stimulatory Mechanisms Regulate the Catalytic Activity of Polycomb Repressive Complex 2*. Mol Cell, 2018. **70**(3): p. 435-448 e5.
104. Mu, Z., et al., *EZH2 knockdown suppresses the growth and invasion of human inflammatory breast cancer cells*. J Exp Clin Cancer Res, 2013. **32**: p. 70.
105. Jiao, L. and X. Liu, *Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2*. Science, 2015. **350**(6258): p. aac4383.
106. Margueron, R., et al., *Role of the polycomb protein EED in the propagation of repressive histone marks*. Nature, 2009. **461**(7265): p. 762-7.
107. Poepsel, S., V. Kasinath, and E. Nogales, *Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes*. Nat Struct Mol Biol, 2018. **25**(2): p. 154-162.
108. Justin, N., et al., *Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2*. Nat Commun, 2016. **7**: p. 11316.

References

109. Gonzalez, M.E., et al., *Downregulation of EZH2 decreases growth of estrogen receptor-negative invasive breast carcinoma and requires BRCA1*. *Oncogene*, 2009. **28**(6): p. 843-53.
110. Hansen, K.H., et al., *A model for transmission of the H3K27me3 epigenetic mark*. *Nat Cell Biol*, 2008. **10**(11): p. 1291-300.
111. Oksuz, O., et al., *Capturing the Onset of PRC2-Mediated Repressive Domain Formation*. *Mol Cell*, 2018. **70**(6): p. 1149-1162 e5.
112. Reveron-Gomez, N., et al., *Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication*. *Mol Cell*, 2018. **72**(2): p. 239-249 e5.
113. Shen, X., et al., *EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency*. *Mol Cell*, 2008. **32**(4): p. 491-502.
114. Hubbard, K.S., et al., *Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs*. *F1000Res*, 2013. **2**: p. 35.
115. van de Leemput, J., et al., *CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells*. *Neuron*, 2014. **83**(1): p. 51-68.
116. McCabe, M.T., et al., *Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27)*. *Proc Natl Acad Sci U S A*, 2012. **109**(8): p. 2989-94.
117. Lee, C.H., et al., *Allosteric Activation Dictates PRC2 Activity Independent of Its Recruitment to Chromatin*. *Mol Cell*, 2018. **70**(3): p. 422-434 e6.
118. Kasinath, V., et al., *Structures of human PRC2 with its cofactors AEBP2 and JARID2*. *Science*, 2018. **359**(6378): p. 940-944.
119. Chen, S., et al., *A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin*. *Mol Cell*, 2020. **77**(6): p. 1265-1278 e7.

120. Forsbach, G., et al., *Pituitary function in human pseudocyesis*. J Endocrinol Invest, 1987. **10**(1): p. 39-43.
121. Youmans, D.T., J.C. Schmidt, and T.R. Cech, *Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits*. Genes Dev, 2018. **32**(11-12): p. 794-805.
122. Choi, J., et al., *DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation*. Nat Struct Mol Biol, 2017. **24**(12): p. 1039-1047.
123. Son, J., et al., *Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin*. Genes Dev, 2013. **27**(24): p. 2663-77.
124. Chen, S., et al., *Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding*. Mol Cell, 2018. **69**(5): p. 840-852 e5.
125. Hubner, J.M., et al., *EZH1/CXorf67 mimics K27M mutated oncohistones and functions as an intrinsic inhibitor of PRC2 function in aggressive posterior fossa ependymoma*. Neuro Oncol, 2019. **21**(7): p. 878-889.
126. Jain, S.U., et al., *PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism*. Nat Commun, 2019. **10**(1): p. 2146.
127. Piunti, A., et al., *CATACOMB: An endogenous inducible gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism*. Sci Adv, 2019. **5**(7): p. eaax2887.
128. Ragazzini, R., et al., *EZH1 constrains Polycomb Repressive Complex 2 activity in germ cells*. Nat Commun, 2019. **10**(1): p. 3858.
129. Oliviero, G., et al., *Dynamic Protein Interactions of the Polycomb Repressive Complex 2 during Differentiation of Pluripotent Cells*. Mol Cell Proteomics, 2016. **15**(11): p. 3450-3460.
130. Li, H., et al., *Polycomb-like proteins link the PRC2 complex to CpG islands*. Nature, 2017. **549**(7671): p. 287-291.
131. Musselman, C.A., et al., *Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1*. Nat Struct Mol Biol, 2012. **19**(12): p. 1266-72.

References

132. Yuan, W., et al., *H3K36 methylation antagonizes PRC2-mediated H3K27 methylation*. J Biol Chem, 2011. **286**(10): p. 7983-7989.
133. Ballare, C., et al., *Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity*. Nat Struct Mol Biol, 2012. **19**(12): p. 1257-65.
134. Smith, M.M. and A.G. Dawson, *Effect of triiodothyronine on alcohol dehydrogenase and aldehyde dehydrogenase activities in rat liver. Implications for the control of ethanol metabolism*. Biochem Pharmacol, 1985. **34**(13): p. 2291-6.
135. Cai, L., et al., *An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting*. Mol Cell, 2013. **49**(3): p. 571-82.
136. Beringer, M., et al., *EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells*. Mol Cell, 2016. **64**(4): p. 645-658.
137. Liefke, R., V. Karwacki-Neisius, and Y. Shi, *EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape*. Mol Cell, 2016. **64**(4): p. 659-672.
138. Zhang, Z., et al., *PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming*. Stem Cells, 2011. **29**(2): p. 229-40.
139. Fernandes, I., et al., *Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and -independent mechanisms*. Mol Cell, 2003. **11**(1): p. 139-50.
140. Hildebrand, J.D. and P. Soriano, *Overlapping and unique roles for C-terminal binding protein 1 (CtBP1) and CtBP2 during mouse development*. Mol Cell Biol, 2002. **22**(15): p. 5296-307.
141. Mozzetta, C., et al., *The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing*. Mol Cell, 2014. **53**(2): p. 277-89.
142. Yang, Y. and G. Li, *Post-translational modifications of PRC2: signals directing its activity*. Epigenetics Chromatin, 2020. **13**(1): p. 47.
143. Wang, X., et al., *Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA*. Nat Struct Mol Biol, 2017. **24**(12): p. 1028-1038.

144. Kalb, R., et al., *Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression*. *Nat Struct Mol Biol*, 2014. **21**(6): p. 569-71.
145. Cooper, S., et al., *Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2*. *Nat Commun*, 2016. **7**: p. 13661.
146. Sanulli, S., et al., *Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation*. *Mol Cell*, 2015. **57**(5): p. 769-783.
147. Yuan, W., et al., *Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation*. *Science*, 2012. **337**(6097): p. 971-5.
148. Perino, M., et al., *MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding*. *Nat Genet*, 2018. **50**(7): p. 1002-1010.
149. Brinkman, A.B., et al., *Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk*. *Genome Res*, 2012. **22**(6): p. 1128-38.
150. Jermann, P., et al., *Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation*. *Proc Natl Acad Sci U S A*, 2014. **111**(33): p. E3415-21.
151. Reddington, J.P., et al., *Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes*. *Genome Biol*, 2013. **14**(3): p. R25.
152. Reddington, J.P., D. Sproul, and R.R. Meehan, *DNA methylation reprogramming in cancer: does it act by reconfiguring the binding landscape of Polycomb repressive complexes?* *Bioessays*, 2014. **36**(2): p. 134-40.
153. Bonev, B., et al., *Multiscale 3D Genome Rewiring during Mouse Neural Development*. *Cell*, 2017. **171**(3): p. 557-572 e24.
154. Denholtz, M., et al., *Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization*. *Cell Stem Cell*, 2013. **13**(5): p. 602-16.
155. Joshi, O., et al., *Dynamic Reorganization of Extremely Long-Range Promoter-Promoter Interactions between Two States of Pluripotency*. *Cell Stem Cell*, 2015. **17**(6): p. 748-757.

References

156. Kundu, S., et al., *Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation*. *Mol Cell*, 2017. **65**(3): p. 432-446 e5.
157. Schoenfelder, S., et al., *Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome*. *Nat Genet*, 2015. **47**(10): p. 1179-1186.
158. Vieux-Rochas, M., et al., *Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain*. *Proc Natl Acad Sci U S A*, 2015. **112**(15): p. 4672-7.
159. McLaughlin, K., et al., *DNA Methylation Directs Polycomb-Dependent 3D Genome Re-organization in Naive Pluripotency*. *Cell Rep*, 2019. **29**(7): p. 1974-1985 e6.
160. Cerase, A., et al., *Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy*. *Proc Natl Acad Sci U S A*, 2014. **111**(6): p. 2235-40.
161. Portoso, M., et al., *PRC2 is dispensable for HOTAIR-mediated transcriptional repression*. *EMBO J*, 2017. **36**(8): p. 981-994.
162. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs*. *Cell*, 2007. **129**(7): p. 1311-23.
163. Colognori, D., et al., *Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X*. *Mol Cell*, 2019. **74**(1): p. 101-117 e10.
164. Davidovich, C., et al., *Promiscuous RNA binding by Polycomb repressive complex 2*. *Nat Struct Mol Biol*, 2013. **20**(11): p. 1250-7.
165. Kaneko, S., et al., *Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin*. *Mol Cell*, 2014. **53**(2): p. 290-300.
166. Kaneko, S., et al., *PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells*. *Nat Struct Mol Biol*, 2013. **20**(11): p. 1258-64.
167. Zhao, J., et al., *Genome-wide identification of polycomb-associated RNAs by RIP-seq*. *Mol Cell*, 2010. **40**(6): p. 939-53.
168. Beltran, M., et al., *The interaction of PRC2 with RNA or chromatin is mutually antagonistic*. *Genome Res*, 2016. **26**(7): p. 896-907.

169. Cifuentes-Rojas, C., et al., *Regulatory interactions between RNA and polycomb repressive complex 2*. Mol Cell, 2014. **55**(2): p. 171-85.
170. Herzog, V.A., et al., *A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element*. Nat Genet, 2014. **46**(9): p. 973-981.
171. Davidovich, C., et al., *Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA*. Mol Cell, 2015. **57**(3): p. 552-8.
172. Wang, X., et al., *Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines*. Mol Cell, 2017. **65**(6): p. 1056-1067 e5.
173. Reynolds, N., et al., *NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression*. EMBO J, 2012. **31**(3): p. 593-605.
174. Morey, L., et al., *MBD3, a component of the NuRD complex, facilitates chromatin alteration and deposition of epigenetic marks*. Mol Cell Biol, 2008. **28**(19): p. 5912-23.
175. Egan, C.M., et al., *CHD5 is required for neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression*. Dev Cell, 2013. **26**(3): p. 223-36.
176. Sparmann, A., et al., *The chromodomain helicase Chd4 is required for Polycomb-mediated inhibition of astroglial differentiation*. EMBO J, 2013. **32**(11): p. 1598-612.
177. Mohd-Sarip, A., et al., *DOC1-Dependent Recruitment of NURD Reveals Antagonism with SWI/SNF during Epithelial-Mesenchymal Transition in Oral Cancer Cells*. Cell Rep, 2017. **20**(1): p. 61-75.
178. Pasini, D., et al., *Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes*. Nucleic Acids Res, 2010. **38**(15): p. 4958-69.
179. Lavarone, E., C.M. Barbieri, and D. Pasini, *Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity*. Nat Commun, 2019. **10**(1): p. 1679.
180. Brien, G.L., et al., *Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation*. Nat Struct Mol Biol, 2012. **19**(12): p. 1273-81.

References

181. Mysliwiec, M.R., et al., *Jarid2 (Jumonji, AT rich interactive domain 2) regulates NOTCH1 expression via histone modification in the developing heart*. J Biol Chem, 2012. **287**(2): p. 1235-41.
182. Francis, N.J., et al., *Reconstitution of a functional core polycomb repressive complex*. Mol Cell, 2001. **8**(3): p. 545-56.
183. Sarma, K., et al., *ATRX directs binding of PRC2 to Xist RNA and Polycomb targets*. Cell, 2014. **159**(4): p. 869-83.
184. Ren, W., et al., *Disruption of ATRX-RNA interactions uncovers roles in ATRX localization and PRC2 function*. Nat Commun, 2020. **11**(1): p. 2219.
185. Wang, Z.X., et al., *The transcription factor Zfp281 controls embryonic stem cell pluripotency by direct activation and repression of target genes*. Stem Cells, 2008. **26**(11): p. 2791-9.
186. Guo, X., et al., *LincRNA-1614 coordinates Sox2/PRC2-mediated repression of developmental genes in pluripotency maintenance*. J Mol Cell Biol, 2018. **10**(2): p. 118-129.
187. Abboud, N., et al., *A cohesin-OCT4 complex mediates Sox enhancers to prime an early embryonic lineage*. Nat Commun, 2015. **6**: p. 6749.
188. Ezhkova, E., et al., *EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair*. Genes Dev, 2011. **25**(5): p. 485-98.
189. O'Carroll, D., et al., *The polycomb-group gene Ezh2 is required for early mouse development*. Mol Cell Biol, 2001. **21**(13): p. 4330-6.
190. Pasini, D., et al., *Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity*. EMBO J, 2004. **23**(20): p. 4061-71.
191. Shumacher, A., C. Faust, and T. Magnuson, *Positional cloning of a global regulator of anterior-posterior patterning in mice*. Nature, 1996. **383**(6597): p. 250-3.
192. Conway, E., et al., *A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities*. Mol Cell, 2018. **70**(3): p. 408-421 e8.
193. Dickinson, M.E., et al., *High-throughput discovery of novel developmental phenotypes*. Nature, 2016. **537**(7621): p. 508-514.

194. Vizan, P., et al., *The Polycomb-associated factor PHF19 controls hematopoietic stem cell state and differentiation*. Sci Adv, 2020. **6**(32): p. eabb2745.
195. Landeira, D. and A.G. Fisher, *Inactive yet indispensable: the tale of Jarid2*. Trends Cell Biol, 2011. **21**(2): p. 74-80.
196. Rothberg, J.L.M., et al., *Mtf2-PRC2 control of canonical Wnt signaling is required for definitive erythropoiesis*. Cell Discov, 2018. **4**: p. 21.
197. Takeuchi, T., et al., *Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation*. Genes Dev, 1995. **9**(10): p. 1211-22.
198. Kim, H., et al., *Aebp2 as an epigenetic regulator for neural crest cells*. PLoS One, 2011. **6**(9): p. e25174.
199. Grijzenhout, A., et al., *Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs*. Development, 2016. **143**(15): p. 2716-23.
200. Ikeda, K., et al., *Maintenance of the functional integrity of mouse hematopoiesis by EED and promotion of leukemogenesis by EED haploinsufficiency*. Sci Rep, 2016. **6**: p. 29454.
201. Herrera-Merchan, A., et al., *Ectopic expression of the histone methyltransferase Ezh2 in haematopoietic stem cells causes myeloproliferative disease*. Nat Commun, 2012. **3**: p. 623.
202. Lessard, J., et al., *Functional antagonism of the Polycomb-Group genes eed and Bmi1 in hemopoietic cell proliferation*. Genes Dev, 1999. **13**(20): p. 2691-703.
203. Mochizuki-Kashio, M., et al., *Ezh2 loss in hematopoietic stem cells predisposes mice to develop heterogeneous malignancies in an Ezh1-dependent manner*. Blood, 2015. **126**(10): p. 1172-83.
204. Xie, H., et al., *Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner*. Cell Stem Cell, 2014. **14**(1): p. 68-80.
205. Celik, H., et al., *JARID2 Functions as a Tumor Suppressor in Myeloid Neoplasms by Repressing Self-Renewal in Hematopoietic Progenitor Cells*. Cancer Cell, 2018. **34**(5): p. 741-756 e8.
206. Hidalgo, I., et al., *Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest*. Cell Stem Cell, 2012. **11**(5): p. 649-62.

References



207. Liu, P.P., et al., *Polycomb Protein EED Regulates Neuronal Differentiation through Targeting SOX11 in Hippocampal Dentate Gyrus*. Stem Cell Reports, 2019. **13**(1): p. 115-131.
208. Sun, B., et al., *Polycomb Protein Eed is Required for Neurogenesis and Cortical Injury Activation in the Subventricular Zone*. Cereb Cortex, 2018. **28**(4): p. 1369-1382.
209. Chiacchiera, F., et al., *PRC2 preserves intestinal progenitors and restricts secretory lineage commitment*. EMBO J, 2016. **35**(21): p. 2301-2314.
210. Ezhkova, E., et al., *Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells*. Cell, 2009. **136**(6): p. 1122-35.
211. Cervantes, S., et al., *Late-stage differentiation of embryonic pancreatic beta-cells requires Jarid2*. Sci Rep, 2017. **7**(1): p. 11643.
212. Cho, E., et al., *Cardiac-specific developmental and epigenetic functions of Jarid2 during embryonic development*. J Biol Chem, 2018. **293**(30): p. 11659-11673.
213. Griffiths, S., et al., *EED and EZH2 constitutive variants: A study to expand the Cohen-Gibson syndrome phenotype and contrast it with Weaver syndrome*. Am J Med Genet A, 2019. **179**(4): p. 588-594.
214. Li, J., et al., *EZH2-mediated H3K27 trimethylation mediates neurodegeneration in ataxia-telangiectasia*. Nat Neurosci, 2013. **16**(12): p. 1745-53.
215. von Schimmelmann, M., et al., *Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration*. Nat Neurosci, 2016. **19**(10): p. 1321-30.
216. Nikoloski, G., et al., *Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes*. Nat Genet, 2010. **42**(8): p. 665-7.
217. Ernst, T., et al., *Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders*. Nat Genet, 2010. **42**(8): p. 722-6.
218. Kim, K.H. and C.W. Roberts, *Targeting EZH2 in cancer*. Nat Med, 2016. **22**(2): p. 128-34.
219. Morin, R.D., et al., *Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin*. Nat Genet, 2010. **42**(2): p. 181-5.

220. Sneeringer, C.J., et al., *Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas*. Proc Natl Acad Sci U S A, 2010. **107**(49): p. 20980-5.
221. McCabe, M.T., et al., *EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations*. Nature, 2012. **492**(7427): p. 108-12.
222. Knutson, S.K., et al., *A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells*. Nat Chem Biol, 2012. **8**(11): p. 890-6.
223. Ntziachristos, P., et al., *Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia*. Nature, 2014. **514**(7523): p. 513-7.
224. Simon, C., et al., *A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia*. Genes Dev, 2012. **26**(7): p. 651-6.
225. Marchione, D.M., et al., *Histone H3K27 dimethyl loss is highly specific for malignant peripheral nerve sheath tumor and distinguishes true PRC2 loss from isolated H3K27 trimethyl loss*. Mod Pathol, 2019. **32**(10): p. 1434-1446.
226. Tan, J., et al., *Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells*. Genes Dev, 2007. **21**(9): p. 1050-63.
227. Beguelin, W., et al., *EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation*. Cancer Cell, 2013. **23**(5): p. 677-92.
228. Cooney, E., et al., *Novel EED mutation in patient with Weaver syndrome*. Am J Med Genet A, 2017. **173**(2): p. 541-545.
229. Lee, W., et al., *PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors*. Nat Genet, 2014. **46**(11): p. 1227-32.
230. He, Y., et al., *The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex*. Nat Chem Biol, 2017. **13**(4): p. 389-395.
231. Qi, W., et al., *An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED*. Nat Chem Biol, 2017. **13**(4): p. 381-388.

References

232. Lee, S.C., et al., *Polycomb repressive complex 2 component Suz12 is required for hematopoietic stem cell function and lymphopoiesis*. *Blood*, 2015. **126**(2): p. 167-75.
233. Yeh, H.H., et al., *Ras induces experimental lung metastasis through up-regulation of RbAp46 to suppress RECK promoter activity*. *BMC Cancer*, 2015. **15**: p. 172.
234. Ntziachristos, P., et al., *Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia*. *Nat Med*, 2012. **18**(2): p. 298-301.
235. Koontz, J.I., et al., *Frequent fusion of the JAZF1 and JJAZ1 genes in endometrial stromal tumors*. *Proc Natl Acad Sci U S A*, 2001. **98**(11): p. 6348-53.
236. Zhang, M., et al., *Somatic mutations of SUZ12 in malignant peripheral nerve sheath tumors*. *Nat Genet*, 2014. **46**(11): p. 1170-2.
237. Zhang, J., et al., *The genetic basis of early T-cell precursor acute lymphoblastic leukaemia*. *Nature*, 2012. **481**(7380): p. 157-63.
238. Wang, J., et al., *HNF1B-mediated repression of SLUG is suppressed by EZH2 in aggressive prostate cancer*. *Oncogene*, 2020. **39**(6): p. 1335-1346.
239. Verberne, E.A., et al., *JARID2 haploinsufficiency is associated with a clinically distinct neurodevelopmental syndrome*. *Genet Med*, 2021. **23**(2): p. 374-383.
240. Gebre-Medhin, S., et al., *Recurrent rearrangement of the PHF1 gene in ossifying fibromyxoid tumors*. *Am J Pathol*, 2012. **181**(3): p. 1069-77.
241. Chiang, S., et al., *Frequency of known gene rearrangements in endometrial stromal tumors*. *Am J Surg Pathol*, 2011. **35**(9): p. 1364-72.
242. Maganti, H.B., et al., *Targeting the MTF2-MDM2 Axis Sensitizes Refractory Acute Myeloid Leukemia to Chemotherapy*. *Cancer Discov*, 2018. **8**(11): p. 1376-1389.
243. Jain, P., et al., *PHF19 mediated regulation of proliferation and invasiveness in prostate cancer cells*. *Elife*, 2020. **9**.
244. Ren, Z., et al., *PHF19 promotes multiple myeloma tumorigenicity through PRC2 activation and broad H3K27me3 domain formation*. *Blood*, 2019. **134**(14): p. 1176-1189.

245. Pajtler, K.W., et al., *Molecular heterogeneity and CXorf67 alterations in posterior fossa group A (PFA) ependymomas*. Acta Neuropathol, 2018. **136**(2): p. 211-226.
246. Varambally, S., et al., *The polycomb group protein EZH2 is involved in progression of prostate cancer*. Nature, 2002. **419**(6907): p. 624-9.
247. Bracken, A.P., et al., *EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer*. EMBO J, 2003. **22**(20): p. 5323-35.
248. Kleer, C.G., et al., *EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells*. Proc Natl Acad Sci U S A, 2003. **100**(20): p. 11606-11.
249. Di Carlo, V., I. Mocavini, and L. Di Croce, *Polycomb complexes in normal and malignant hematopoiesis*. J Cell Biol, 2019. **218**(1): p. 55-69.
250. Conway, E., E. Healy, and A.P. Bracken, *PRC2 mediated H3K27 methylations in cellular identity and cancer*. Curr Opin Cell Biol, 2015. **37**: p. 42-8.
251. Souroullas, G.P., et al., *An oncogenic Ezh2 mutation induces tumors through global redistribution of histone 3 lysine 27 trimethylation*. Nat Med, 2016. **22**(6): p. 632-40.
252. Robbez-Masson, L.J., et al., *Functional analysis of a breast cancer-associated FGFR2 single nucleotide polymorphism using zinc finger mediated genome editing*. PLoS One, 2013. **8**(11): p. e78839.
253. Cromer, M.K., et al., *Identification of somatic mutations in parathyroid tumors using whole-exome sequencing*. J Clin Endocrinol Metab, 2012. **97**(9): p. E1774-81.
254. Hodis, E., et al., *A landscape of driver mutations in melanoma*. Cell, 2012. **150**(2): p. 251-63.
255. Milosevic, J.D., et al., *Clinical significance of genetic aberrations in secondary acute myeloid leukemia*. Am J Hematol, 2012. **87**(11): p. 1010-6.
256. Puda, A., et al., *Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies*. Am J Hematol, 2012. **87**(3): p. 245-50.
257. Ghislin, S., et al., *PHF19 and Akt control the switch between proliferative and invasive states in melanoma*. Cell Cycle, 2012. **11**(8): p. 1634-45.

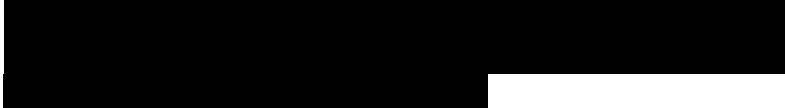
270. Bender, S., et al., *Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas*. *Cancer Cell*, 2013. **24**(5): p. 660-72.
271. Buczkowicz, P., et al., *Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations*. *Nat Genet*, 2014. **46**(5): p. 451-6.
272. Buczkowicz, P., et al., *Histopathological spectrum of paediatric diffuse intrinsic pontine glioma: diagnostic and therapeutic implications*. *Acta Neuropathol*, 2014. **128**(4): p. 573-81.
273. Taylor, K.R., et al., *ACVR1 mutations in DIPG: lessons learned from FOP*. *Cancer Res*, 2014. **74**(17): p. 4565-70.
274. Wu, G., et al., *The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma*. *Nat Genet*, 2014. **46**(5): p. 444-450.
275. Mackay, A., et al., *Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma*. *Cancer Cell*, 2017. **32**(4): p. 520-537 e5.
276. Fang, D., et al., *H3.3K27M mutant proteins reprogram epigenome by sequestering the PRC2 complex to poised enhancers*. *Elife*, 2018. **7**.
277. 
278. Mohammad, F., et al., *EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas*. *Nat Med*, 2017. **23**(4): p. 483-492.
279. Piunti, A., et al., *Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas*. *Nat Med*, 2017. **23**(4): p. 493-500.
280. Stafford, J.M., et al., *Multiple modes of PRC2 inhibition elicit global chromatin alterations in H3K27M pediatric glioma*. *Sci Adv*, 2018. **4**(10): p. eaau5935.
281. 
282. Nagaraja, S., et al., *Histone Variant and Cell Context Determine H3K27M Reprogramming of the Enhancer*

References

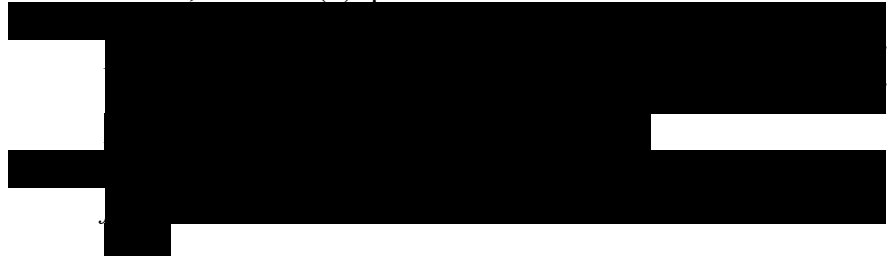
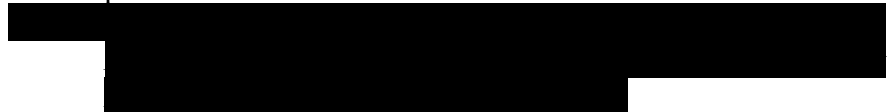
- Landscape and Oncogenic State*. Mol Cell, 2019. **76**(6): p. 965-980 e12.
283. Larson, J.D., et al., *Histone H3.3 K27M Accelerates Spontaneous Brainstem Glioma and Drives Restricted Changes in Bivalent Gene Expression*. Cancer Cell, 2019. **35**(1): p. 140-155 e7.
284. Haag, D., et al., *H3.3-K27M drives neural stem cell-specific gliomagenesis in a human iPSC-derived model*. Cancer Cell, 2021. **39**(3): p. 407-422 e13.
285. Harutyunyan, A.S., et al., *H3K27M in Gliomas Causes a One-Step Decrease in H3K27 Methylation and Reduced Spreading within the Constraints of H3K36 Methylation*. Cell Rep, 2020. **33**(7): p. 108390.
286. Krug, B., et al., *Pervasive H3K27 Acetylation Leads to ERV Expression and a Therapeutic Vulnerability in H3K27M Gliomas*. Cancer Cell, 2019. **35**(5): p. 782-797 e8.
287. Nagaraja, S., et al., *Transcriptional Dependencies in Diffuse Intrinsic Pontine Glioma*. Cancer Cell, 2017. **31**(5): p. 635-652 e6.
288. Rodrigues, C.P.e.a., *Epigenetic Regulators as the Gatekeepers of Hematopoiesis Trends in Genetics*. Trends in Genetics, 2021. **Volume 37, Issue 2, 125 - 142**.
289. Orkin, S.H. and L.I. Zon, *Hematopoiesis: an evolving paradigm for stem cell biology*. Cell, 2008. **132**(4): p. 631-44.
290. Ceredig, R., A.G. Rolink, and G. Brown, *Models of haematopoiesis: seeing the wood for the trees*. Nat Rev Immunol, 2009. **9**(4): p. 293-300.
291. Ramkumar, C., R.M. Gerstein, and H. Zhang, *Serial transplantation of bone marrow to test self-renewal capacity of hematopoietic stem cells in vivo*. Methods Mol Biol, 2013. **976**: p. 17-24.
292. Kiel, M.J., et al., *SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells*. Cell, 2005. **121**(7): p. 1109-21.
293. Ali, M.A.E., et al., *Functional dissection of hematopoietic stem cell populations with a stemness-monitoring system based on NS-GFP transgene expression*. Sci Rep, 2017. **7**(1): p. 11442.
294. Fukushima, T., et al., *Discrimination of Dormant and Active Hematopoietic Stem Cells by G0 Marker Reveals Dormancy*

- Regulation by Cytoplasmic Calcium*. Cell Rep, 2019. **29**(12): p. 4144-4158 e7.
295. Bernitz, J.M., et al., *Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions*. Cell, 2016. **167**(5): p. 1296-1309 e10.
296. Cabezas-Wallscheid, N., et al., *Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy*. Cell, 2017. **169**(5): p. 807-823 e19.
297. Lauridsen, F.K.B., et al., *Differences in Cell Cycle Status Underlie Transcriptional Heterogeneity in the HSC Compartment*. Cell Rep, 2018. **24**(3): p. 766-780.
298. Sun, D., et al., *Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal*. Cell Stem Cell, 2014. **14**(5): p. 673-88.
299. Nazaraliyev, A., E. Richard, and C.M. Sawai, *In-vivo differentiation of adult hematopoietic stem cells from a single-cell point of view*. Curr Opin Hematol, 2020. **27**(4): p. 241-247.
300. Rossi, L., et al., *Hematopoietic stem cell characterization and isolation*. Methods Mol Biol, 2011. **750**: p. 47-59.
301. Picelli, S., et al., *Full-length RNA-seq from single cells using Smart-seq2*. Nat Protoc, 2014. **9**(1): p. 171-81.
302. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool*. BMC Bioinformatics, 2013. **14**: p. 128.
303. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update*. Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
304. Gazit, R., et al., *Fgd5 identifies hematopoietic stem cells in the murine bone marrow*. J Exp Med, 2014. **211**(7): p. 1315-31.
305. Chambers, S.M., et al., *Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation*. PLoS Biol, 2007. **5**(8): p. e201.
306. Yamamoto, R., et al., *Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment*. Cell Stem Cell, 2018. **22**(4): p. 600-607 e4.
307. Benz, C., et al., *Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs*. Cell Stem Cell, 2012. **10**(3): p. 273-83.


References

308. Beerman, I., et al., *Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle*. *Cell Stem Cell*, 2014. **15**(1): p. 37-50.
309. Rossi, D.J., et al., *Hematopoietic stem cell quiescence attenuates DNA damage response and permits DNA damage accumulation during aging*. *Cell Cycle*, 2007. **6**(19): p. 2371-6.
310. Yu, V.W.C., et al., *Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells*. *Cell*, 2016. **167**(5): p. 1310-1322 e17.
311. Corces, M.R., et al., *Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution*. *Nat Genet*, 2016. **48**(10): p. 1193-203.
312. Schmidl, C., et al., *ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors*. *Nat Methods*, 2015. **12**(10): p. 963-965.
313. Blanco, E., L.D. Croce, and S. Aranda, *Comparative ChIP-seq (Comp-ChIP-seq): a practical guideline for experimental design and a novel computational methodology*. *bioRxiv*, 2019: p. 532622.
314. Yang, J., et al., *Single cell transcriptomics reveals unanticipated features of early hematopoietic precursors*. *Nucleic Acids Res*, 2017. **45**(3): p. 1281-1296.
315. Rube, C.E., et al., *Accumulation of DNA damage in hematopoietic stem and progenitor cells during human aging*. *PLoS One*, 2011. **6**(3): p. e17487.
316. Flach, J., et al., *Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells*. *Nature*, 2014. **512**(7513): p. 198-202.
317. Lu, Z., et al., *Polycomb Group Protein YY1 Is an Essential Regulator of Hematopoietic Stem Cell Quiescence*. *Cell Rep*, 2018. **22**(6): p. 1545-1559.
318. Chen, J., et al., *Identification of new therapeutic targets and natural compounds against diffuse intrinsic pontine glioma (DIPG)*. *Bioorg Chem*, 2020. **99**: p. 103847.
319. 
320. Jones, C. and S.J. Baker, *Unique genetic and epigenetic mechanisms driving paediatric diffuse high-grade glioma*. *Nat Rev Cancer*, 2014. **14**(10).


321. Filbin, M.G., et al., *Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq*. Science, 2018. **360**(6386): p. 331-335.
322. Vinci, M., et al., *Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells*. Nat Med, 2018. **24**(8): p. 1204-1215.
323. Nikbakht, H., et al., *Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma*. Nat Commun, 2016. **7**: p. 11185.
324. Pathania, M., et al., *H3.3(K27M) Cooperates with Trp53 Loss and PDGFRA Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas*. Cancer Cell, 2017. **32**(5): p. 684-700 e9.
325. Sanson, K.R., et al., *Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities*. Nat Commun, 2018. **9**(1): p. 5416.
326. Joung, J., et al., *Author Correction: Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening*. Nat Protoc, 2019. **14**(7): p. 2259.

- 
329. Zhu, Y., et al., *DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis*. Mol Cell Proteomics, 2020. **19**(6): p. 1047-1057.
 330. Orre, L.M., et al., *SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization*. Mol Cell, 2019. **73**(1): p. 166-182 e7.
 331. Aranda, S., et al., *Chromatin capture links the metabolic enzyme AHCY to stem cell proliferation*. Sci Adv, 2019. **5**(3): p. eaav2448.
- 

References

333. Dai, Z., V. Ramesh, and J.W. Locasale, *The evolving metabolic landscape of chromatin biology and epigenetics*. Nat Rev Genet, 2020. **21**(12): p. 737-753.
334. Rodrigues, D.C., et al., *Methylglyoxal couples metabolic and translational control of Notch signalling in mammalian neural stem cells*. Nat Commun, 2020. **11**(1): p. 2018.
335. Khan, A., et al., *Dual targeting of polyamine synthesis and uptake in diffuse intrinsic pontine gliomas*. Nat Commun, 2021. **12**(1): p. 971.
- 
337. Lin, G.L., et al., *Therapeutic strategies for diffuse midline glioma from high-throughput combination drug screening*. Sci Transl Med, 2019. **11**(519).
338. Aoyama, K., et al., *Ezh1 Targets Bivalent Genes to Maintain Self-Renewing Stem Cells in Ezh2-Insufficient Myelodysplastic Syndrome*. iScience, 2018. **9**: p. 161-174.
339. Rossi, D.J., et al., *Cell intrinsic alterations underlie hematopoietic stem cell aging*. Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9194-9.
340. Pang, W.W., et al., *Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age*. Proc Natl Acad Sci U S A, 2011. **108**(50): p. 20012-7.
341. Ng, S.W., et al., *A 17-gene stemness score for rapid determination of risk in acute leukaemia*. Nature, 2016. **540**(7633): p. 433-437.
342. Baquero, P., et al., *Targeting quiescent leukemic stem cells using second generation autophagy inhibitors*. Leukemia, 2019. **33**(4): p. 981-994.
343. Vetrie, D., G.V. Helgason, and M. Copland, *The leukaemia stem cell: similarities, differences and clinical prospects in CML and AML*. Nat Rev Cancer, 2020. **20**(3): p. 158-173.
344. Auberger, P. and A. Puissant, *Autophagy, a key mechanism of oncogenesis and resistance in leukemia*. Blood, 2017. **129**(5): p. 547-552.
345. Chandel, N.S., et al., *Metabolic regulation of stem cell function in tissue homeostasis and organismal ageing*. Nat Cell Biol, 2016. **18**(8): p. 823-32.

346. Fleming, H.E., et al., *Wnt signaling in the niche enforces hematopoietic stem cell quiescence and is necessary to preserve self-renewal in vivo*. *Cell Stem Cell*, 2008. **2**(3): p. 274-83.
347. Rossi, L., et al., *Less is more: unveiling the functional core of hematopoietic stem cells through knockout mice*. *Cell Stem Cell*, 2012. **11**(3): p. 302-17.
348. Beerman, I., et al., *Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging*. *Cell Stem Cell*, 2013. **12**(4): p. 413-25.
349. Yamamoto, R., et al., *Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells*. *Cell*, 2013. **154**(5): p. 1112-1126.
350. Florian, M.C., et al., *Aging alters the epigenetic asymmetry of HSC division*. *PLoS Biol*, 2018. **16**(9): p. e2003389.
351. Mejia-Ramirez, E. and M.C. Florian, *Understanding intrinsic hematopoietic stem cell aging*. *Haematologica*, 2020. **105**(1): p. 22-37.
352. Wahlestedt, M., et al., *Clonal reversal of ageing-associated stem cell lineage bias via a pluripotent intermediate*. *Nat Commun*, 2017. **8**: p. 14533.
353. Wahlestedt, M., et al., *An epigenetic component of hematopoietic stem cell aging amenable to reprogramming into a young state*. *Blood*, 2013. **121**(21): p. 4257-64.
354. Biechonski, S., et al., *Attenuated DNA damage responses and increased apoptosis characterize human hematopoietic stem cells exposed to irradiation*. *Sci Rep*, 2018. **8**(1): p. 6071.
355. Olaciregui, N.G., *New therapeutic approaches in diffuse intrinsic pontine glioma (dipg)*, in *Department of Pediatric Hematology and Oncology, Hospital Sant Joan de Déu*. 2019, University of Barcelona.

- 
372. Kent, W.J., et al., *The human genome browser at UCSC*.
Genome Res, 2002. **12**(6): p. 996-1006.

References

Annex

Papers published during this thesis

Vizán P*, Gutiérrez A*, **Espejo I**, García-Montolio M, Lange M, Carretero A, Lafzi A, de Andrés-Aguayo L, Blanco E, Thambyrajah R, Graf T, Heyn H, Bigas A, Di Croce L. The Polycomb-associated factor PHF19 controls hematopoietic stem cell state and differentiation. *Sci Adv.* 2020; 6(32):eabb2745. DOI: 10.1126/sciadv.abb2745

Espejo I, Di Croce L, Aranda S. The changing chromatinome as a driver of disease: a panoramic view from different methodologies. *Bioessays.* 2020 Dec; 42(12):e 2000203. DOI: 10.1002/bies.202000203

Annex I

SCIENCE ADVANCES | RESEARCH ARTICLE

EPIGENETICS

The Polycomb-associated factor PHF19 controls hematopoietic stem cell state and differentiation

Pedro Vizán^{1*}, Arantxa Gutiérrez^{1*}, Isabel Espejo¹, Marc García-Montolio¹, Martin Lange^{1†}, Ana Carretero², Atefeh Lafzi³, Luisa de Andrés-Aguayo¹, Enrique Blanco¹, Roshana Thamyrajah⁴, Thomas Graf¹, Holger Heyn³, Anna Bigas^{4,5}, Luciano Di Croce^{1,6,7‡}

Adult hematopoietic stem cells (HSCs) are rare multipotent cells in bone marrow that are responsible for generating all blood cell types. HSCs are a heterogeneous group of cells with high plasticity, in part, conferred by epigenetic mechanisms. PHF19, a subunit of the Polycomb repressive complex 2 (PRC2), is preferentially expressed in mouse hematopoietic precursors. Here, we now show that, in stark contrast to results published for other PRC2 subunits, genetic depletion of *Phf19* increases HSC identity and quiescence. While proliferation of HSCs is normally triggered by forced mobilization, defects in differentiation impede long-term correct blood production, eventually leading to aberrant hematopoiesis. At molecular level, PHF19 deletion triggers a redistribution of the histone repressive mark H3K27me3, which notably accumulates at blood lineage-specific genes. Our results provide novel insights into how epigenetic mechanisms determine HSC identity, control differentiation, and are key for proper hematopoiesis.

INTRODUCTION

Blood regenerates at a high level, providing an ideal platform for studying stem cell function. Hematopoietic stem cells (HSCs) in bone marrow (BM) are rare cells at the top of a hierarchical hematopoietic system; in this model, HSCs are defined by having long-term repopulation capacity and produce the more committed progenitors, which eventually produce all the differentiated cell types. Although well established in general terms, this model has been recently challenged in two ways: (i) HSCs can be already biased toward specific lineages, and (ii) new technologies (particularly single-cell transcriptomics) have demonstrated a high level of heterogeneity among HSCs, although it has been impossible to define discrete categories. Thus, it is becoming more accepted that boundaries between cell compartments, probably not only within HSCs but also in all undifferentiated progenitor compartments, are less strictly defined (1, 2). In addition to being functionally defined as a group of cells capable of long-term repopulation capacity, HSCs can also be molecularly defined as cells that are able to maintain the highest level of plasticity. Strong gene regulation programs confer irreversibility during the differentiation process, and cell types/compartments can be functionally defined by their potency capacity after isolation using associated surface markers. In this context, epigenetic mechanisms are likely to play a central role in controlling and perpetuating the transcriptional programs important for the hematopoietic system. A number of epigenetic alterations have been linked to multiple hematopoietic diseases (3, 4).

The Polycomb group (PcG) of proteins form multiprotein complexes that are major epigenetic regulators of gene expression. The Polycomb repressive complex 2 (PRC2) catalyzes the trimethylation of lysine-27 in the histone H3 N-terminal tail (H3K27me3), which is associated with chromatin compaction and gene repression (5, 6). PRC2 comprises three core components [Enhancer of zeste homolog 2 (EZH2), suppressor of zeste 12 (SUZ12), and embryonic ectoderm development (EED)] and several substoichiometrical accessory factors that regulate its function (7, 8). In the past two decades, PRC2 proteins have been studied in the context of hematopoiesis and their mutations have been linked to hematological disorders (9). Depletion of PRC2 core components in murine models lead to marked effects on hematopoiesis at either neonatal or fetal stage, demonstrating the crucial role of PRC2 in hematopoietic homeostasis (10–12). Three homologs of the *Drosophila melanogaster* Polycomb-like (*Pcl*) gene have been described as PRC2-associated factors: *Phf1/Pcl1*, *Mlf2/Pcl2*, and *Phf19/Pcl3*. PHF19 has been mainly characterized in mouse embryonic stem cells (mESCs) (13–15), where it plays a role in self-renewal and differentiation. *Phf19* expression declines upon mESC differentiation, although adult stem cell compartments within specific tissues still retain substantially high expression levels (16, 17). Here, we report a murine model devoid of PHF19, which allowed us to investigate its role in hematopoiesis. We have uncovered how PHF19 controls HSCs identity and long-term contribution to blood production. Furthermore, we identified chromatin alterations in HSCs in the absence of PHF19 and unveiled the key role of H3K27me3 localization in HSCs differentiation and in preventing malignant hematopoiesis.

RESULTS

PHF19 plays a role in long-term hematopoietic reconstitution

In the hematopoietic system, *Phf19* expression is relatively elevated in undifferentiated progenitors and progressively decreases during differentiation (fig. S1A) (18, 19). To study its role in adult homeostasis, we generated a mouse genetic model by deleting exons 4 and 5, which produces a stop codon in exon 6 (fig. S1, B to E). *Phf19*-KO (knockout) mice breed normally, and histopathological analysis of

¹Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Dr. Aiguader 88, Barcelona 08003, Spain. ²Department of Animal Health and Anatomy, School of Veterinary Medicine, Universitat Autònoma de Barcelona, Bellaterra, Spain. ³CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Baldri Reixac 4, Barcelona 08028, Spain. ⁴Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona 08003, Spain. ⁵CIBERONC, Barcelona, Spain. ⁶Universitat Pompeu Fabra (UPF), Barcelona, Spain. ⁷ICREA, Pg. Lluís Companys 23, Barcelona 08003, Spain.

*These authors contributed equally to this work.

†Present address: Bayer AG, Pharmaceuticals Division, Research and Development, Berlin, Germany.

‡Corresponding author. Email: luciano.dicroce@crg.eu

young adult mice (8 to 12 weeks) did not exhibit any major alterations in the brain, lungs, heart, stomach, intestine, kidney, testis, ovary, or uterus. As reported for other Polycomb members (20, 21), *Phf19*-KO mice presented anterior-to-posterior homeotic transformations, such as the presence of rib anlage or extra ribs in ~40% (6 of 14) of animals (fig. S1F; note that no alterations in *Phf19*-Flox mice were detected). Aged *Phf19*-KO mice (>60 weeks) presented a high penetrance of splenomegaly (fig. S1G), suggestive of a defect in the hematopoietic system.

We analyzed the hematopoietic hierarchy of fetal liver [embryonic day 14.5 (E14.5)] and of 8-week-old mouse BM. Lack of PHF19 led to an increase in embryonic phenotypically defined HSCs (Lineage⁻ Sca-1^{c-Kit}⁺CD150⁺CD48⁻), but in adults, we observed a reduction in HSCs, as well as in LSK (Lineage⁻ Sca-1^{c-Kit}⁺) progenitors (Fig. 1A). We also documented a significant reduction in cellularity (raw cell count) in BM (Fig. 1B). Furthermore, we investigated the adult HSC compartment by assessing the proportion of proliferative Ki67⁺ positive cells (Fig. 1C) and of bromodeoxyuridine (BrdU) incorporation during 24 hours (Fig. 1D). Our data indicated a significant increase in quiescent cells in *Phf19*-KO-derived HSCs, which may account for the observed decrease in LSK, HSC cell populations, and total cellularity in adults.

To determine the role of *Phf19* in BM performance, we performed transplantation assays. Initially, we transplanted 0.5 million cells of whole BM (WBM) obtained from 8-week-old *Phf19*-Flox or *Phf19*-KO mice in competition with 0.5 million cells of CD45 congenic wild-type WBM cells. No differences were observed in the engraftment capacity or in blood contribution (fig. S1H). We then performed serial noncompetitive repopulation assays by transplanting 1 million cells of *Phf19*-Flox or *Phf19*-KO WBM and monitoring blood contribution after 3 months of each transplant (Fig. 1E and fig. S1I). Our analysis indicated a slight bias toward myeloid lineage in *Phf19*-KO transplanted mice, which was only statistically significant after the third transplant (data not shown). We observed a clear decrease in the contribution of *Phf19*-KO transplanted donors in serial transplants, pointing to a malfunction of HSCs. Colony-forming assays from lineage-specific depleted BM showed a similar behavior: No differences were found in the type or the total number of colonies after the first plating (fig. S1J), but after replating, *Phf19*-KO-derived cells were significantly impaired in their capacity to produce new colonies (Fig. 1F). Parallel experiments using progenitor LSK population led to similar results (fig. S1K).

***Phf19*-KO HSCs are mobilized upon stress but are not fully functional**

To determine whether the decrease in BM HSC levels under homeostatic conditions (Fig. 1A) is the cause of the long-term lower efficiency in blood production (Fig. 1E), we analyzed the total proportion of LSK progenitors and HSCs upon different stress conditions that have been described to force HSC mobilization: aging, transplantation, and 5-fluorouracil (5-FU) injection. The hematopoietic compartment analysis showed that the reduced number of HSCs in *Phf19*-KO animals was compensated upon stress (Fig. 2A and fig. S2A), indicating that *Phf19*-KO HSCs are able to proliferate upon stress. Therefore, the reduced long-term capacity for replenishing the blood system might be related to a differentiation defect.

To assess cell division kinetics and differentiation capacity of PHF19-depleted HSCs, we next performed single-cell culture of

phenotypically defined HSCs (22, 23). Single Lineage⁻ Sca-1^{c-Kit}⁺ CD150⁺CD48⁻ HSCs were individually sorted on 96-well plates, and the number of cells was monitored for 4 days (Fig. 2B). Each individual well was scored according to its division dynamics: proliferating (i.e., sustained cell growth over the 4 days, rendering colony sizes from five to hundreds of cells), nonproliferating (i.e., not able to divide more than twice during 4 days), and proliferating and stop (i.e., division that eventually stopped or reduced the number of cells). Under culture conditions, most wells were scored as proliferating, both in *Phf19*-Flox and *Phf19*-KO plates, but a significant increase in the percentage of proliferating colonies in *Phf19*-KO wells was observed, which correlates with a reduction in nonproliferating wells (Fig. 2C). Notably, not only *Phf19*-KO wells had more proliferating colonies, but also these colonies individually contained a higher number of cells (fig. S2B). We kept colonies in culture for additional 5 days and then calculated colony size by fluorescence cytometer after propidium iodide (PI) staining. According to the division status during the first 4 days, the number of big colonies (>5000 cells) was higher in the *Phf19*-KO plates than in the *Phf19*-Flox plates (fig. S2C). Last, after 2 weeks, we measured the expression of lineage and HSC markers. We observed an impairment of differentiation in *Phf19*-KO condition (Fig. 2D). Moreover, the remaining CD150⁺CD48⁻ population was higher in the *Phf19*-KO than in the *Phf19*-Flox cells (fig. S2D). Together, these data demonstrate that HSCs from *Phf19*-KO cells have higher proliferation but lower differentiation capacity compared to control cells. This is in agreement with our *in vivo* data, which indicate that *Phf19*-KO HSC cells are not able to fully differentiate in serial transplant experiments despite its increased capacity of generating progenitors upon stress.

PHF19 regulates HSC gene expression through epigenetic control of PRC2 activity

To gain insight into how PHF19 is able to control the quiescence state in homeostasis, we first performed RNA sequencing (RNA-seq) transcriptomic analysis on HSCs (Lineage⁻ Sca-1^{c-Kit}⁺CD150⁺CD48⁻) isolated from young adult mice (table S1). Term enrichment analysis of common up-regulated genes of two independent replicates showed that nuclear receptors and, in particular, retinoic acid category were overrepresented. On the other hand, down-regulated genes were enriched for categories related to biosynthesis and energy production (fig. S3A). This is reminiscent of a recently characterized, highly unbiased, and quiescent HSC population, the so-called dormant HSCs (24).

We next did a gene set enrichment analysis (GSEA) for specific gene sets. HSC-specific signatures, a merge of differentially expressed genes provided by Gazit *et al.* (25) and Chambers *et al.* (26), were enriched in *Phf19*-KO (Fig. 3, A and C), indicating that HSC identity is enhanced upon PHF19 depletion. Furthermore, the Myc network, which is required for HSC differentiation (24, 27), was down-regulated in *Phf19*-KO HSCs (Fig. 3, B and C). In addition, a decrease of enrichment in biosynthetic pathways (such as ribosome or mRNA processing) and an increase in positively regulated retinoic acid targets (28) were observed (Fig. 3C and fig. S3B), further supporting an intensification in HSC quiescence and dormancy (24, 29). We also compared our data with an expression signature derived from HSCs depending on their divisional history (30). Up-regulated genes in HSCs that have undergone several divisions inversely correlate with gene expression of *Phf19*-KO HSCs and vice versa for down-regulated

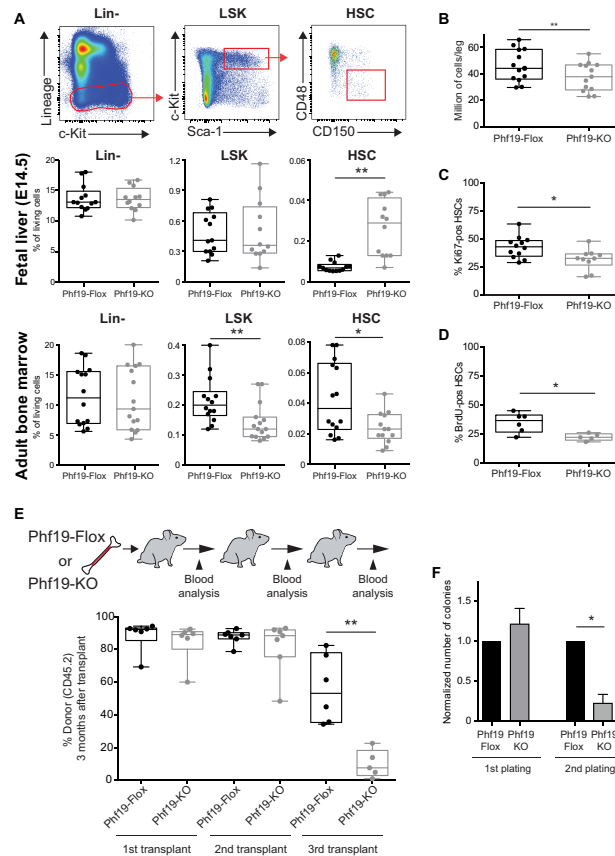


Fig. 1. Characterization of the hematopoietic system in a *Phf19*-depleted mouse model. (A) Top: BM gating strategy for phenotypically defining lineage negative (Lin⁻), LSK, and HSCs (Lineage⁻Sca-1⁻c-Kit⁺CD150⁺CD48⁺). Middle: Quantification from 13 control floxed (*Phf19*-Flox) and 12 *Phf19*-KO E14.5 embryos, with percentages shown for the live cells measured as 4',6'-diamidino-2-phenylindole (DAPI) negative (box plot). Bottom: Quantification from 14 control floxed (*Phf19*-Flox) and 13 to 15 *Phf19*-KO mice, with percentages shown for the live cells measured as DAPI negative (box plot). (B) Raw number of WBM cells counted per leg in 13 independent experiments (with at least three animals per experiment) (box plot). (C) Percentage of Ki67⁺ HSC cells in 12 mice for *Phf19*-Flox and 10 mice for *Phf19*-KO (box plot). (D) Percentage of BrdU⁺ HSCs in six *Phf19*-Flox mice or five *Phf19*-KO mice at 24 hours after BrdU injection (box plot). (E) Percentage of donor-derived (CD45.2) cells in peripheral blood in serially transplanted recipients 3 months after each transplant (box plot). (F) Normalized number of colonies in the first and second lineage-negative plated cells for colony-forming assay, performed in three independent replicates (means + SEM). **P* < 0.05 and ***P* < 0.01. Unpaired *t* test (A and C to E). Paired *t* test (B and F).

genes (fig. S3B). Moreover, since *Phf19*-KO HSCs are proliferative but defective in differentiation and long-term repopulation, we wondered whether they resembled aged HSCs (31). GSEA demonstrated that up-regulated genes in old animals are enriched in *Phf19*-KO

condition; meanwhile, down-regulated genes are enriched in *Phf19*-Flox transcriptome (fig. S3B). Together, transcriptomics data corroborated that *Phf19*-depleted cells in homeostasis are in a more multipotent and quiescent state and have a decreased capacity upon

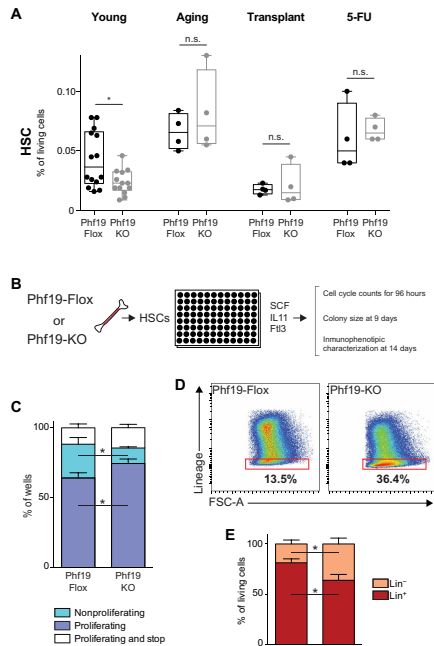


Fig. 2. Characterization of *Phf19*-KO HSCs after mobilization. (A) Quantification of phenotypically defined HSCs in young or aged (>60-week-old) mice after transplantation (at 6 months after transplant) and at 14 days after 5-FU injection. The percentage of the live cells measured as DAPI negative is shown (box plot). (B) Schematic representation of a single-cell HSC in vitro functional assay. (C) Proportion of wells of proliferating, nonproliferating, and proliferating and then stopped HSCs in three independent experiments (means + SEM). (D) Gating of lineage-negative marker versus FSC-A (forward scatter area) of pooled wells after 14 days in culture. (E) Percentage of lineage-negative and lineage-positive cells from three independent replicates (means + SEM). * $P < 0.05$; n.s., not significant. Paired t test (C and E).

induction of differentiation. Last, the role of Jumonji and AT-rich interaction domain containing 2 (JARID2) in the hematopoietic system has been recently investigated (32). A comparative analysis of transcriptomic data revealed a similar effect for JARID2 and PHF19 on HSCs gene regulation (fig. S3B).

PHF19 is an associated PRC2 factor (7) and controls the trimethylation on lysine-27 of histone H3 (H3K27me3) on chromatin (8). We thus performed chromatin immunoprecipitation sequencing (ChIP-seq) experiments on Lineage⁺ Sca-1⁺ Kit⁺ CD150⁺ population, and unexpectedly, we found a global increase in H3K27me3 levels, even when spike-in control for quantitative measurements between conditions was used (fig. S3C). We also corroborated that such in-

crease is global and not only detectable inside the peaks (fig. S3D). This acquired H3K27me3 deposition pattern resembled the one observed in aged HSCs (31) since genes marked with H3K27me3 in aged HSCs intensively accumulate H3K27me3 upon depletion of PHF19, and conversely, the H3K27me3 genes in young HSCs are also more intensively marked with H3K27me3 in *Phf19*-Flox HSCs (fig. S3E). At molecular level, Polycomb complexes regulate differentiation by restricting lineage-determining gene expression (5). We thus reasoned that in *Phf19*-KO cells, this mechanism could be implicated in the increased HSC identity observed at a functional and transcriptomic level. As expected, the genomic accumulation of H3K27me3 levels inversely correlated with expression data (fig. S3F). However, despite the global increase in H3K27me3, the HSC gene sets did not show a significant increase in H3K27me3 levels (Fig. 3D, left). Significant increase was observed for genes that are characteristic of differentiated lineages (Fig. 3D, middle). Moreover, transcription factors implicated in HSC differentiation into erythroid, granulocyte, lymphoid, or megakaryocyte (20) also showed a significant increase in H3K27me3 (Fig. 3D, right). Increased H3K27me3 deposition in lineage-specific transcription factors could be observed for essential genes—such as *Cebpa* (for myeloid) (33); *Pax5* and *Irf5* (for B cells) (34, 35); *Tcf7* (for T cells) (36); several members of Notch pathway [*Dll1* (depicted), *Dll4*, *Jag2*, or *Hes1*], which are well-known regulators of both T and B cells differentiation (37); and *Epo* (for erythroid cells) (38)—and could not be observed in HSC fingerprint genes (*Vdr* depicted as an example; fig. S3G). To further validate how the depletion of PHF19 influences chromatin organization, we performed assay for transposase-accessible chromatin and sequencing (ATAC-seq) and verified that the increase in H3K27me3-containing regions was accompanied by a loss of chromatin accessibility (fig. S3H). This decrease could be visualized in most of the master differentiation transcription factors mentioned (Fig. 3E).

Differentiation program genes are poorly expressed in HSCs. Therefore, to evaluate the impact of chromatin changes on gene expression, we designed two strategies. First, we performed single-cell RNA-seq. Our first observation was that *Phf19*-KO HSCs were transcriptionally more similar between each other than *Phf19*-Flox HSCs, both using all expressed genes (fig. S3I) and using a HSC signature specifically defined from single-cell data (Fig. 3F) (23). Phenotypically defined HSC comprises a heterogeneous population, which contains cells primed for differentiation. We reasoned that we could detect transcripts of differentiation genes in individual HSC cells. We thus computed the number of cells expressing differentiation gene set (Fig. 3D, middle) (25, 26). As expected, some, but not all, of the cells expressed differentiation genes. Interestingly, we observed we observed a significant decrease in the number of cells that expressed these genes in the *Phf19*-KO condition (Fig. 3G). Second, upon forced in vitro differentiation (as depicted in Fig. 2B), we monitored by quantitative polymerase chain reaction (qPCR) the expression of transcription factors that presented an accumulation of H3K27me3 mark within their promoter regions. As represented in Fig. 3H, all genes showed a decrease in mRNA expression, with the only exception of *Cebpa*. Since differentiation is not completely abolished in *Phf19*-KO cultured HSCs, we hypothesized that myeloid lineage could be enhanced (see next section), which is in agreement with the increased expression levels of *Cebpa*. Together, our data indicate that the absence of PHF19 produces an epigenetic rewiring of HSCs, leading to an enhancement of stem identity likely instructed by an aberrant repression of differentiation programs.

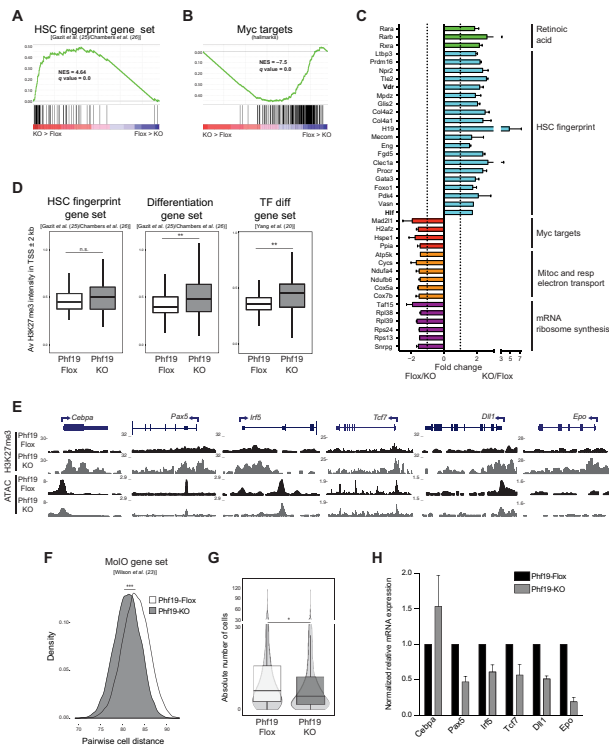


Fig. 3. HSC gene expression control and epigenetic changes associated with *Phf19* depletion. (A) GSEA showing positive enrichment in *Phf19*-KO transcriptome for HSC gene set [Gazit *et al.* (25) and Chambers *et al.* (26)]. (B) GSEA showing negative enrichment in *Phf19*-KO transcriptome for Myc target gene set. (C) Normalized fold change expression from two independent RNA sequencing (RNA-seq) experiments in relevant genes (means \pm SD). (D) ChIP-seq levels of H3K27me3 in the transcription start site (TSS) \pm 2 kb of specific gene sets: the HSC gene set, the differentiation gene set, and for transcription factors associated with hematopoietic differentiation processes. (E) University of California Santa Cruz (UCSC) genome browser screenshots for H3K27me3 and ATAC of hematopoiesis master regulators in *Phf19*-FloX and *Phf19*-KO. (F) Density plot of pairwise distances of cells based on single cell–based HSC (MoIO) signature (23) genes for *Phf19*-FloX and *Phf19*-KO cells. (G) Distribution shown by violin plot and box plot of absolute number of cells that expressed differentiation genes for 174 *Phf19*-FloX and *Phf19*-KO cells. (H) mRNA expression levels of cultured HSCs after 9 days under growth/differentiation conditions depicted in Fig. 2B, relative to *Rplp0* and normalized for *Phf19*-FloX of two replicates (means \pm SEM). * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$. Paired *t* test (D). Wilcoxon rank-sum test (F and G). NES, normalized enrichment score.

Lack of PHF19 causes long-term aberrant hematopoiesis

To study the consequences of the increased proliferation and impairment of differentiation observed in *Phf19*-KO HSCs, we performed WBM transplant from aged mice, which already presented defects in hematopoiesis (as revealed by splenomegaly; fig. S1G), to irradiated young recipient mice. We again detected a peripheral blood lineage myeloid bias after serial transplant (fig. S4A), and impairment of blood contribution by *Phf19*-KO donors was accelerated and already evident in the first and second rounds of trans-

plant (fig. S4B). Moreover, a high incidence of splenomegaly was observed in mice euthanized for WBM extraction for the second transplant (Fig. 4A). Moreover, survival of *Phf19*-KO versus *Phf19*-FloX transplanted mice was significantly reduced (Fig. 4B).

Flow cytometry characterization of BM from survivor mice showed the presence of an uncharacterized population (fig. S4C), suggesting anomalous hematopoiesis. We next analyzed spleens of the same young transplant survivor mice (four of six with splenomegaly in *Phf19*-KO and one of six with splenomegaly in *Phf19*-FloX)

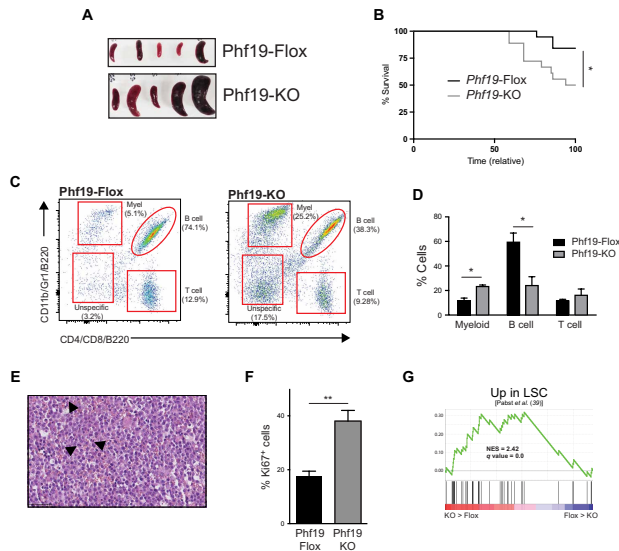


Fig. 4. Lack of Phf19 causes aberrant hematopoiesis over the long term. (A) Picture depicting splenomegaly at 6 months after transplantation with WBM from aged mice. Photo credit: Pedro Vizán, CRG. (B) Survival curve of all mice (*Phf19*-Flox versus *Phf19*-KO) after their last transplant. Time normalized for each experiment to final sacrifice time of all animals. (C) Representative gating of whole spleen analyzed for myeloid, B cell, and T cell proportions (donor CD45.2-positive cells present in the spleen). (D) Quantification of donor CD45.2-positive myeloid, B cell, and T cell proportions in five *Phf19*-Flox and three *Phf19*-KO mice. In *Phf19*-KO conditions, the sum does not reach 100% because of the unspecific population observed in the panel for hematoxylin and eosin staining (means + SEM). (E) Representative hematoxylin and eosin staining of *Phf19*-KO spleens. Arrowheads indicate mitotic figures. (F) Percentage of Ki67-positive cells in two spleens per condition (more than four white pulp areas per spleen) (means + SEM). (G) GSEA showing positive enrichment in *Phf19*-KO transcriptome for leukemic stem cell (LSC) differentially up-regulated genes (37). * $P < 0.05$ and ** $P < 0.01$. Log-rank (Mantel-Cox) test (B). Unpaired t test (D and F).

by flow cytometry. When we evaluated donor CD45.2 cell contributions, we observed a decrease in B cell content together with an increase in myeloid infiltration (Fig. 4, C and D), as well as the presence of an unusual, unspecific splenic cell population (Fig. 4C). Further characterization of the spleens by histological analysis revealed that the tissue architecture was compromised in *Phf19*-KO transplanted mice, with almost indistinguishable white and red pulps (fig. S4D), together with the presence of a population of large pleomorphic cells with irregularly shaped nuclei (Fig. 4E). The cytological features of these cells are compatible with myeloid dysplasia. Many mitotic figures were present, and immunohistochemistry for the proliferation marker Ki67 corroborated a higher rate of proliferation in *Phf19*-KO spleens than in *Phf19*-Flox spleens (Fig. 4F). Moreover, the macrophage marker CD68, which is found mostly in the red pulp (fig. S4E, left), was diffusely spread in all areas of the *Phf19*-KO samples (fig. S4E, right), in accordance with our flow cytometry observation of unspecific and increased myeloid cells.

In light of these results, we reevaluated our transcriptomic data. Term enrichment of up-regulated genes in *Phf19*-KO condition also showed the pathways in cancer category [Kyoto Encyclopedia of Genes and Genomes (KEGG) mouse 2019, $P = 9.16 \times 10^{-7}$]. Fur-

thermore, a set of differentially up-regulated genes found in a human leukemic stem cell (LSC) population (39) was enriched in *Phf19*-KO versus *Phf19*-Flox (Fig. 4G). It has also been observed that LSC transcriptional programs are partially shared with stem cell transcription programs, and a reduced gene catalog of 17 genes was reported to correlate with poor prognosis and relapse for acute myeloid leukemia (40). We manually examined these genes in our transcriptomic data. Thirteen of the original 17 genes were expressed [reads per kilobase million (RPKM) > 0.1] in our dataset, and 10 (~77%) were up-regulated. These observations are quite notable since these specific alterations in the transcriptome are already evident in young mice. Together, our results indicate that PHF19 is necessary to maintain a correct hematopoietic balance and that its depletion induces an aberrant state, which primes HSCs for malignant progression.

DISCUSSION

Here, we have shown that PHF19 contributes to maintain the correct levels of H3K27me3 within the HSC compartment, influencing their long-term function and differentiation predisposition. Blood production has been classically explained by a hierarchical model in

which HSCs have the capacity to self-renew and differentiate into a population of more committed progenitors, which are still able to self-renew and produce all blood types of the organism. In the past few years, this model is being reinterpreted as a continuous change between cellular states rather than a stepwise process. Within the HSC compartment, variability is given by characteristics that are not initially related to their differentiation potential but rather with their cellular state, such as biosynthetic rate or energy metabolism. Thus, HSCs have been further characterized in dormant and active with a continuum among the two states (24). Our results indicate that PHF19 controls the balance between these two conditions.

Epigenetics is one of the main factors that govern cell plasticity, as well as the transition towards new cell states. In particular, the PcG of proteins, which main function is transcriptional repression, have been reported to play a role in HSC performance (9). Loss-of-function studies of PRC2 core components demonstrated their essential role in hematopoiesis (10–12), but the associated lethality has hampered an in-depth characterization of transcriptional pathways and H3K27me3 (re)distribution implicated in HSC identity. Similarly, metal response element binding transcription factor 2 (MTF2) deletion impairs fetal erythropoiesis (41), causing a pre-birth death that has also impeded further studying of the role of H3K27me3 in adult HSCs. Several studies have unveiled that PCL proteins and JARID2 give rise to two biochemically distinct versions of the PRC2 complex: PRC2.1 and PRC2.2 (6). It has been recently reported that JARID2 depletion in hematopoietic precursors (32) affects the expression of self-renewal genes in HSCs and multiple progenitor precursors. Transcriptional comparison shows that JARID2-depleted HSCs resemble *Phf19*-KO HSCs (fig. S3B) despite a reduction in H3K27me3 levels. We believe that PRC2 variants control self-renewal and differentiation by complementary mechanisms. Notably, PRC2.2 may be more active than PRC2.1, as previously suggested (42). Thus, PRC2.2 reduction might cause direct derepression of a set of HSC genes, and conversely, PRC2.1 reduction might allow PRC2.2 to increment H3K27me3 deposition at differentiation genes, producing a more homogeneous HSC-like transcriptome. These two distinct molecular mechanisms would paradoxically produce a similar phenotype: increase of HSC/dormancy features and impairment of proper differentiation. Together, this highlights the importance of the proper balance between PRC2 variations and their specific role with respect to the control of adult stem cell functionality. Unfortunately, nowadays, there are no available data of PRC2 component localization at genome-wide level, which impedes a thorough analysis of the impact of distinct PRC2 versions in HSC gene regulation.

As we have demonstrated in this study, *Phf19* deletion does not affect breeding or life span. Although we observed changes in the number of HSCs (higher proportion in fetal liver and lower proportion in young adults), these alterations do not have fatal consequences during development or in adult homeostasis under animal facility conditions. However, under challenging conditions, such as aging or serial transplants, we have unveiled the role for PHF19 in controlling HSC functions. In *Phf19*-KO animals, dormant HSC characteristics are enhanced under resting conditions, such as high quiescence, or low metabolic and biosynthetic activities. Upon induced mobilization, proliferation is triggered, but differentiation is impaired. This double-edged behavior could be linked to epigenetic states. The main role of PcG is to maintain cell identity by restricting lineage-specific gene transcription. In *Phf19*-KO HSCs, we observed an increase in H3K27me3 mark and chromatin compaction at

genes necessary for differentiation. At the same time, once differentiation is triggered, this compaction would reduce the accessibility of specific transcription factors.

In summary, we have demonstrated that PHF19 depletion increases the dormancy-like HSC state by a mechanism that, at the same time, impairs differentiation and reduces durability upon serial transplants. Several studies have used label dilution models to quantify HSC divisional history, which has allowed a link between low mitotic rate and greatest repopulation ability (30, 43). The transcriptional program associated with low divisional history resembles the *Phf19*-KO expression data (fig. S3B). Thus, although we observed a higher number of HSCs in fetal liver, which could be indicative of an increased proliferation, adult HSCs are more quiescent. This could be explained by an early exhaustion of highly proliferative HSCs during fetal and early life stages or by the fact that label dilution models tend to misinterpret label retention and are simply labeling HSCs with high repopulation potential, irrespective of mitotic history (44). Beyond these strong functional and transcriptional similarities, the quiescent features in *Phf19*-KO do not yield in fully functional dormant HSCs. In this regard, it is remarkable that *Phf19*-KO HSCs resemble aged HSCs (fig. S3B), which have been observed to accumulate during aging, but are less functional at generating new blood cells (45). Notably, the H3K27me3 increase observed upon PHF19 depletion is in line with the same mark distribution reported for aged HSCs (fig. S3E) (31). As for the *Phf19*-KO mice, the consequences in aged HSCs are a myeloid bias and a higher propensity to hematological disorders. Moreover, quiescence is the most significant feature in cancer stem cells and the leading cause of relapse. This study underlines the complex balance between the silencing of differentiation programs and the maintenance of elevated levels of plasticity to sustain stem cell identity.

METHODS

Mice

All animal procedures were approved by the ethical committee of the Barcelona Biomedical Research Park and by the animal experimentation commission of the Catalan government (Generalitat de Catalunya). All animals used for this study had a C57BL/6 background. LoxP sites were introduced by homologous recombination around exons 4 and 5, the deletion of which generates a stop codon in exon 6. *Phf19^{fl/fl}* animals were crossed with Sox2Cre (mice were provided by A.R. Nebreda, Institute for Research in Biomedicine, Barcelona). The CRE activity in the Sox2CRE strain has been of maternal inheritance, irrespective of whether they inherit the transgene (46). Therefore, after backcrossing with new male floxed animals, animals with the knocked-out allele without the Sox2Cre transgene were selected. Once generated, mice were maintained as heterozygosis (*Phf19^{fl/fl}*) and as homozygosis (*Phf19^{-/-}*). In both cases, offspring were viable and bred normally. Equal numbers of male and female mice were used; no gender biases were noted. For BM transplants experiments, 8- to 12-week-old congenic CD45.1 and CD45.1/CD45.2 mice (bred in-house) were used.

BM extraction and splenocytes isolation

For fetal liver analysis, the tissue was harvested from E14.5 embryos and the dissected liver was crushed with the end of a 1-ml syringe through a 40- μ m cell strainer into Iscove's modified Dulbecco's medium and 10% fetal bovine serum. For cell BM analysis of individual

mice, clean femurs and tibias were flushed using a 23-gauge needle and a 10-cm³ syringe with cold dissection medium [1× phosphate-buffered saline (PBS), 2% fetal calf serum (FCS), and 2 mM EDTA] before cell surface staining (see next section). For pooled mice analyses, the clean bones were crushed with mortar and pestle in cold dissection medium (1× PBS, 2% FCS, and 2 mM EDTA). For sorting and colony-forming unit assays (see below), lineage depletion was performed using the mouse Lineage Cell Depletion Kit (Miltenyi Biotec) by magnetic-activated cell sorting (MACS) technique according to the manufacturer's instructions. Splenocytes were freshly dissociated by mincing the tissue to obtain a single-cell suspension.

Flow cytometry

Before flow cytometry analysis, fetal liver cell suspension, blood, BM, and splenocytes, erythrocytes were lysed with 1× red blood cell lysis buffer (eBioscience) and washed with 1× PBS. For surface staining, single-cell suspensions in fluorescence-activated cell sorting (FACS) buffer (1× PBS, 2% FCS, and 2 mM EDTA) were blocked for 10 min with anti-CD16/CD32 (mouse Fc block, BD Pharmingen). Samples were stained in FACS buffer for 30 min at 4°C with fluorochrome-conjugated antibodies and washed before analysis. Blood and splenocytes were incubated with the following fluorochrome-conjugated monoclonal antibodies for lineage analysis: phycoerythrin (PE) anti-mouse CD4, PE anti-mouse CD8, PE anti-mouse B220, PECy7 anti-mouse B220, allophycocyanin (APC) anti-mouse CD45.1, fluorescein isothiocyanate (FITC) anti-mouse CD45.2, PECy7 anti-mouse CD11b, and PECy7 anti-mouse GR-1. Fetal liver and BM were analyzed using the lineage marker (Lin) mixture (BD Biosciences) that include the following V450-conjugated antibodies: CD3e (145-2C11); CD11b (M1/70) (this antibody was excluded in fetal liver analyses); CD45R/B220 (RA3-6B2); mouse erythroid cells Ly-76 (Ter119), Ly6G, and Ly-6C (RB6-8C5); and the following fluorochrome-conjugated antibodies for HSC and precursors (LSK): PECy7 anti-mouse Sca 1, APC-Fluo780 anti-mouse cKit, APC anti-mouse CD150, and PE anti-mouse CD48. Cell viability was evaluated with 4',6-diamidino-2-phenylindole (DAPI) at a final concentration of 1:10,000.

For intracellular staining (Ki67 or BrdU), BM cells were resuspended in FACS buffer and stained for 10 min at 4°C with anti-CD16/CD32 (mouse Fc block, BD Pharmingen) for 30 min at 4°C with a cocktail of lineage marker biotin-conjugated monoclonal antibodies against CD2, CD3, CD11b, CD14, CD15, CD16, CD19, CD56, CD123, and CD235a (Miltenyi Biotec) and, lastly, with fluorochrome-conjugated antibodies specific for HSCs (as detailed above). Cells were then washed and incubated with the secondary antibody PE-CF594 streptavidin for 10 min at 4°C. For Ki67 immunostaining, surface-stained BM cells were washed and fixed with reagent A (Invitrogen) for 15 min at room temperature and then washed and permeabilized in reagent B in the presence of FITC anti-mouse Ki67 for 20 min at room temperature. After two washes, cells were resuspended in 500 μ l of 1× PBS containing ribonuclease (RNase; 5 μ g/ml) and DAPI (2 μ g/ml). For BrdU (see next section for BrdU in vivo injection), surface-stained BM cells were washed, and the BrdU incorporation assay was performed using the BrdU Kit from BD Biosciences.

BrdU incorporation assay

Subject mice were intraperitoneally injected twice (at 12-hour intervals) with BrdU (150 μ l of 10 mg/ml). At 12 hours after the second injection,

mice were euthanized, and BM cells were isolated and stained for cytometry as described above.

BM transplantation

For competitive transplantation assay, 0.5×10^6 million nucleated BM cells from femurs and tibias (see next section) of 8- to 12-week-old CD45.2 mice were mixed with 0.5×10^6 million CD45.1 BM cells and transplanted by retro-orbital injection into 8- to 12-week-old CD45.1/CD45.2 recipient mice that were lethally irradiated using a split dose of 8 gray (Gy) (two doses of 4 Gy administered at least 4 hours apart). CD45.2 contribution to peripheral blood was assessed 3 months after transplantation. For both young and old serial transplantation assays, 1×10^6 nucleated BM cells from 8- to 12-week-old (young) or 80- to 85-week-old (old) CD45.2 mice were transplanted by retro-orbital injection into 8- to 12-week-old lethally irradiated (two split doses of 4 Gy, 4 hours apart) CD45.1/CD45.2 recipient mice. The CD45.2 contribution to peripheral blood was assessed at 3 months after transplantation. For serial transplants, animals were euthanized after 6 months, and pooled 1×10^6 nucleated BM cells were used to transplant new 8- to 12-week-old lethally irradiated CD45.1/CD45.2 recipient mice. Recipient mice were treated for 4 weeks after transplantation with Enrovet (enrofloxacin) at 600 μ g/ml in their drinking water.

Colony-forming assay

Lineage-depleted or LSK-sorted BM cells were plated in duplicate in 35-mm tissue culture dishes (STEMCELL Technologies) in methylcellulose-based medium MethoCult™ GF M3434 (STEMCELL Technologies) at 2000 cells per plate. Plates were incubated 10 to 14 days at 37°C, 100% humidity, and 5% CO₂. Colonies were scored using an inverted microscope and resuspended and counted in 1× PBS. Cells at 100,000 per plate were used for replating.

Single-cell HSC culture

Lineage-depleted BM cells were stained as previously described for HSC. Single-cell sorting (FACSARIA, BD) was performed in 96-well plates containing 100 μ l of filtered HSC expansion medium [StemSpan serum-free expansion medium, murine stem cell factor (300 ng/ml), murine interleukin-11 (20 ng/ml), and human Flt3L (1 ng/ml)]. The plates were incubated at 37°C with 5% CO₂, and medium was replaced every 7 days. Cell divisions were visually scored for 96 hours. At day 9, single wells were stained with PI (1 μ g/ml) and counted using FACSCanto II High Throughput Sampler. In addition, an independent plate was harvested and pooled for RNA extraction. At day 14, cells were harvested, pooled, and stained with BM surface markers as previously described.

Real-time PCR

RNA was isolated from cultured HSCs with the RNeasy Plus Mini Kit (Qiagen), and cDNA was generated from 1 μ g of RNA with the qScript cDNA Synthesis Kit (Quantabio), both following the manufacturer's instructions. Real-time PCR reactions were performed using SYBR Green I PCR Master Mix (Roche) and the Roche LightCycler 480.

Population RNA-seq

Two RNA replicates were obtained from 5000 to 10,000 HSCs (Lineage⁻Sca-1⁺c-Kit⁺CD150⁺CD48⁻) sorted from flow cytometry as described above. RNA extraction was performed using the RNeasy Micro Kit (Qiagen). Library preparation was performed with

SCIENCE ADVANCES | RESEARCH ARTICLE

3 to 5 ng of total RNA using the SMARTer Ultra Low RNA Kit (Clontech) and sequenced [50-base pair (bp) single end] on an Illumina HiSeq 3000. The RNA-seq samples were mapped against the mm9 mouse genome assembly using TopHat (47) with the option $-g$ 1 to discard those reads that could not be uniquely mapped in just one region. Cufflinks (48) was run to quantify the expression in RPKMs of each annotated transcript in RefSeq (49). We applied the following thresholds to find differentially expressed genes between each pair of *Phf19*-Flox versus *Phf19*-KO samples [RPKM > 0.5 and KO/Flox FC \geq 1.5 or Flox/KO FC \geq 1.3]. We intersected the up- or down-regulated genes in common between duplicates to end up with the final sets of up-regulated (901) and down-regulated (556) genes, respectively. Reports of functional enrichments of Gene Ontology and other genomic libraries were generated using the Enrichr tool (44). For specific GSEA, we considered as expressed genes those with RPKM > 0.1. Next, we rank-ordered them according to fold change and analyzed using the preranked tool of GSEA (50).

Chromatin immunoprecipitation sequencing

Lineage⁺ Sca-1⁺ c-Kit⁺ CD150⁺ cells (10,000) were sorted by flow cytometry as described above and freshly cross-linked with 1% formaldehyde. Low Cell ChIP-seq kit (Active Motif) was used following the manufacturer's instructions for histone H3 lysine-27 trimethylated ChIP grade antibody (39155, Active Motif). For ChIP experiments with spike-in control, a ratio of 1:2 (*Drosophila*:mouse) S2 cells was added prior cell lysis. Total chromatin was sheared with a Covaris sonication system for 20 min at intensity 8 and 20% duty cycle. ChIP-seq libraries were prepared with total recovered chromatin, using the Next Gen DNA Library Kit and the Next Gen Indexing Kit (Active Motif) as per the manufacturer's instructions. Libraries were sequenced (50-bp single end) on a HiSeq 3000 platform (Illumina). ChIP-seq samples were mapped against the mm9 mouse genome assembly using Bowtie with the option $-m$ 1 to discard those reads that could not be uniquely mapped to just one region (51). ChIP-seq samples normalized by spike-in were mapped against a synthetic genome constituted by the mouse and the fruit fly chromosomes (mm9 and dm3) using Bowtie with the option $-m$ 1 to discard reads that did not map uniquely to one region. MACS was run with the default parameters but with the shift size adjusted to 100 bp to perform the peak calling against the corresponding control sample (52). Each set of target genes was retrieved by matching the ChIP-seq peaks in the region 2.5 kbp upstream of the transcription start site (TSS) until the end of the transcripts as annotated in RefSeq annotations (49). The heat maps displaying the density of ChIP-seq reads around the TSS of each target gene set were generated by counting the number of reads in this region for each individual gene and normalizing this value with the total number of mapped reads of the sample. Genes on each ChIP heat map were ranked by the logarithm of the average number of reads in the same genomic region. Box plots showing the ChIP-seq level distribution for H3K27me3 for particular sets of genes were calculated by determining the average value at transcript level on the region \pm 2 kbp around each TSS. The values of the samples including spike-in were corrected by the number of fly reads mapped of the sequencing experiment. The University of California Santa Cruz (UCSC) genome browser was used to generate the screenshots depicted in the manuscript (53). To confirm the general increase of H3K27me3 ChIP-seq levels from *Phf19*-Flox to *Phf19*-KO conditions, we segmented the mouse genome into bins of 100 bp. Next, we counted within each bin the

number of normalized reads of H3K27me3 corrected by spike-in at each sample. To discard regions of the genome virtually with no signal, we focused on the bins in which, at least in one of the conditions, a minimum amount of ChIP signal was detected (threshold: 5).

Omni-ATAC-seq

Omni-ATAC-seq libraries were generated as previously described (54). Briefly, 10,000 Lineage⁺ Sca-1⁺ c-Kit⁺ CD150⁺ CD48⁺ cells per condition were washed in cold PBS and resuspended in 50 μ l of cold lysis buffer 1 [10 mM tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 0.1% (v/v) Igepal CA-630, 0.01% digitonin, and 0.1% Tween 20]. Samples were incubated on ice for 3 min and washed out with 1 ml of cold lysis buffer 2 [10 mM tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, and 0.1% Tween 20]. Samples were centrifuged for 10 min at 4°C, and the nuclei pellet was resuspended in the transposition reaction mix [25 μ l of 2 \times transposition reaction buffer from Nextera kit, 2.5 μ l of Nextera Tn5 transposase from Nextera kit, 16.5 μ l of PBS, 0.5 μ l of digitonin (1%), 0.5 μ l of Tween 20 (10%), and 5 μ l of nuclease-free water] and incubated at 37°C for 1 hour. Samples were purified using the Qiagen MinElute PCR Purification Kit. Transposed DNA was eluted in 10 μ l of elution buffer and subjected to 5 cycles of PCR before amplification using barcoded primers and NEBNext High-Fidelity PCR Master Mix. We determined the optimal total number of PCR cycles for each Omni-ATAC-seq library using qPCR. Omni-ATAC-seq libraries were purified using 1.8 \times volumes of AMPure XP beads to remove fragments below 100 bp and sequenced (50-bp single end) on an Illumina HiSeq 3000. The ATAC-seq samples were mapped against the mm9 mouse genome assembly using Bowtie with the option $-m$ 1 to discard those reads that could not be uniquely mapped to just one region (51). Mitochondrial reads were removed from each resulting map, and down-sampling was applied to obtain the same number of mapped fragments per sample. Box plots showing the ATAC-seq level distribution particular sets of genes were calculated by determining the average value at transcript level on the region \pm 2 kbp around each TSS. The UCSC genome browser was used to generate the screenshots depicted in the manuscript (53).

Single-cell RNA-seq

Library preparation (MARSeq)

Single-cell libraries from polyadenylate-tailed RNA were constructed, applying massively parallel single-cell RNA-seq (MARSeq) (55, 56). Briefly, single HSC (Lineage⁺ Sca-1⁺ c-Kit⁺ CD150⁺ CD48⁺) cells were FACS-sorted into 384-well plates, containing lysis buffer [0.2% Triton (Sigma-Aldrich) and RNase inhibitor (Invitrogen)] and reverse transcription (RT) primers. The RT primers contained the single-cell barcodes and unique molecular identifiers (UMIs) for subsequent demultiplexing and correction for amplification biases, respectively. Single-cell lysates were denatured and immediately placed on ice. The RT reaction mix containing SuperScript III reverse transcriptase (Invitrogen) was added to each sample. In the RT reaction, spike-in artificial transcripts (External RNA Controls Consortium, Ambion) were included at a dilution of 1:16 \times 10⁶ per cell. After RT, the cDNA was pooled using an automated pipeline (epMotion, Eppendorf). Unbound primers were eliminated by incubating the cDNA with exonuclease I [New England Biolabs (NEB)]. A second pooling was performed through cleanup with solid-phase reversible immobilization (SPRI) magnetic beads (Beckman Coulter). Subsequently, pooled cDNAs were converted into double-stranded DNA

with the Second Strand Synthesis enzyme (NEB), followed by cleanup and linear amplification by T7 *in vitro* transcription overnight. Afterward, the DNA template was removed by Turbo deoxyribonuclease I (Ambion), and the RNA was purified with SPRI beads. Amplified RNA was chemically fragmented with Zn²⁺ (Ambion) and then purified with SPRI beads. The fragmented RNA was ligated with ligation primers containing a pool barcode and partial Illumina Read1 sequencing adapter using T4 RNA ligase I (NEB). Ligated products were reverse transcribed using the AffinityScript RT enzyme (Agilent Technologies) and a primer complementary to the ligated adapter, partial Read1. The cDNA was purified with SPRI beads. Libraries were completed through a PCR step using the KAPA HiFi HotStart ReadyMix (Kapa Biosystems) and a forward primer that contains Illumina P5-Read1 sequence and a reverse primer containing the P7-Read2 sequence. The final library was purified with SPRI beads to remove excess primers. Library concentration and molecular size were determined with High Sensitivity DNA Chip (Agilent Technologies). The libraries consist of 192 single-cell pools. Multiplexed pools (2) were run in one Illumina HiSeq 2500 Rapid two-lane flow cell following the manufacturer's protocol. Primary data analysis was carried out with the standard Illumina pipeline.

Data preprocessing

Sequencing was carried out as paired-end reads, wherein the first read contains the transcript sequence and the second read contains the cell barcode and UMIs. Data analyses were performed as described previously (56). Briefly, quality check of the generated reads was performed with the FastQC quality control suite. Samples that reached the quality standards were then processed to deconvolute the reads to single-cell level by demultiplexing according to the cell and pool barcodes. Reads were filtered to remove polyT sequences. Sequencing reads were mapped to the mouse reference genome (Gencode release M11, assembly GRCh38) with the RNA pipeline of the GEMTools 1.7.0 suite using default parameters (6% of mismatches, minimum of 80% matched bases, and minimum quality threshold of 26). Cells with less than 60% of reads mapping to the reference genome or more than 2×10^6 total reads were discarded. Gene quantification was performed using UMI-corrected transcript information to correct for amplification biases, collapsing read counts for reads mapping to a gene with the same UMI (allowing an edit distance up to two nucleotides in UMI comparisons). Only unambiguously mapped reads were considered. Thresholds were set not only to reduce technical noise but also to conserve the sensitivity to identify low-frequency outlier cell populations and to capture differences between fresh and cryopreserved cells.

Skeletal staining

Skeletal whole mounts were performed as previously described (57). Briefly, the completely eviscerated and skinned animals were fixed in 96% ethanol for 5 days and transferred to acetone for 2 days. Staining was performed in 0.005% Alizarin Red S, 0.015% Alcian blue 8GS in 5% acetic acid, 5% H₂O, and 90% ethanol for 5 days at room temperature. Samples were washed in H₂O and cleared for 2 days in 1% KOH, followed by clearing steps in 0.8% KOH and 20% glycerol, in 0.5% KOH and 50% glycerol, and in 0.2% KOH and 80% glycerol for 1 week each. Cleared skeletons were stored in 100% glycerol.

Spleen histopathology

Paraffin-embedded tissue sections (3 μ m thick) were air dried and further dried at 60°C overnight and then dewaxed. Hematoxylin/

eosin standard protocols were done using a CoverStainer (Dako-Agilent) following the manufacturer's procedures. Immunohistochemistry was performed using an Autostainer Plus (Dako-Agilent). Before immunohistochemistry, sections were dewaxed for Ki67 as part of the antigen retrieval process using the low-pH EnVision FLEX Target Retrieval Solutions (Dako, Burlington). For CD68, samples were dewaxed and rehydrated, and thereafter, antigen retrieval was performed using an autoclave for 20 min at 121°C with citrate buffer (pH 6). Samples were blocked with Peroxidase-Blocking Solution (S2023, Dako-Agilent), 5% normal goat serum, and 2.5% bovine serum albumin. Primary antibodies rabbit polyclonal anti-Ki67 (Abcam) and rabbit polyclonal anti-CD68 (orb47985, Biorbyt) were diluted at 1:2000 and 1:4000, respectively, with EnVision FLEX Antibody Diluent (Dako, Agilent) and incubated for 60 min. The secondary antibody used was a BrightVision poly-horseradish peroxidase anti-rabbit immunoglobulin G (IgG) biotin-free, ready to use (Immunologic, DPVR-110HRP). Antigen-antibody complexes were visualized with 3-3'-diaminobenzidine (Dako). Sections were counterstained with hematoxylin (Dako, S202084) and mounted with Toluene-Free Mounting Medium (CS705, Dako) using the Dako CoverStainer. Specificity of staining was confirmed with a rabbit IgG polyclonal isotype control (Abcam). Bright-field images were acquired with a NanoZoomer-2.0 HT C9600 digital scanner (Hamamatsu) equipped with a 20 \times objective. All images were visualized with a gamma correction set to 1.8 in the image control panel of the NDP.view2 U123888-01 software (Hamamatsu Photonics, France).

Quantification and statistical analysis

Animal data are presented as box plots produced using Prism 6 software (GraphPad, San Diego, CA). Quantified data are presented as means \pm SEM. Statistical tests between groups are detailed in the figure legends. Student's paired or unpaired *t* test, depending on the experimental design, was broadly used in all figures. In addition, Wilcoxon rank-sum test for single-cell data (Fig. 3, F and G, and fig. S31) and log-rank (Mantel-Cox) test (Fig. 4B) were used. Significance is depicted throughout the figures as **P* < 0.05, ***P* < 0.01, and ****P* < 0.001.

Data and software availability

Raw data and processed information of the ChIP-seq, RNA-seq, and ATAC-seq experiments generated here were deposited in the National Center for Biotechnology Information Gene Expression Omnibus (58) repository under the accession number GSE135283.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/32/eabb2745/DC1>

[View/request a protocol for this paper from Bio-protocols.](#)

REFERENCES AND NOTES

1. S. Haas, A. Trumpp, M. D. Milsom, Causes and consequences of hematopoietic stem cell heterogeneity. *Cell Stem Cell* **22**, 627–638 (2018).
2. E. Laurenti, B. Gottgens, From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).
3. P. Ntziachristos, O. Abdel-Wahab, I. Aifantis, Emerging concepts of epigenetic dysregulation in hematological malignancies. *Nat. Immunol.* **17**, 1016–1024 (2016).
4. C. Plass, S. M. Pfister, A. M. Lindroth, O. Bogatyrova, R. Claus, P. Lichter, Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat. Rev. Genet.* **14**, 765–780 (2013).
5. L. Di Croce, K. Helin, Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* **20**, 1147–1155 (2013).

SCIENCE ADVANCES | RESEARCH ARTICLE

6. B. Schuettengruber, H. M. Bourbon, L. Di Croce, G. Cavalli, Genome regulation by Polycomb and Trithorax: 70 years and counting. *Cell* **171**, 34–57 (2017).
7. P. Vizan, M. Beringer, C. Ballaré, L. Di Croce, Role of PRC2-associated factors in stem cells and disease. *FEBS J* **282**, 1723–1735 (2015).
8. A. Laugesen, J. W. Hoffeldt, K. Helin, Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. *Mol. Cell* **74**, 8–18 (2019).
9. V. Di Carlo, I. Mocarini, L. Di Croce, Polycomb complexes in normal and malignant hematopoiesis. *J. Cell Biol* **218**, 55–69 (2019).
10. H. Xie, J. Xu, J. H. Hsu, M. Nguyen, Y. Fujiwara, C. Peng, S. H. Orkin, Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* **14**, 68–80 (2014).
11. S. C. W. Lee, S. Miller, C. Hyland, M. Kauppi, M. Lebois, L. D. Rago, D. Metcalf, S. A. Kinkel, E. C. Josefsson, M. E. Blewitt, I. J. Majewski, W. S. Alexander, Polycomb repressive complex 2 component Suz12 is required for hematopoietic stem cell function and lymphopoiesis. *Blood* **126**, 167–175 (2015).
12. W. Yu, F. Zhang, S. Wang, Y. Fu, J. Chen, X. Liang, H. Le, W. T. Pu, B. Zhang, Depletion of polycomb repressive complex 2 core component EED impairs fetal hematopoiesis. *Cell Death Dis.* **8**, e2744 (2017).
13. C. Ballaré, M. Lange, A. Lapinaite, M. G. Martin, L. Morey, G. Pascual, R. Liefke, B. Simon, Y. Shi, O. Gozani, T. Carlomagno, S. A. Benitah, L. D. Croce, PHF19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* **19**, 1257–1265 (2012).
14. J. Hunkapiller, Y. Shen, A. Diaz, G. Cagney, D. M. Cleary, M. Ramalho-Santos, N. Krogan, B. Ren, J. S. Song, J. F. Reiter, Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. *PLoS Genet.* **8**, e1002576 (2012).
15. G. L. Brien, G. Gambero, D. J. O'Connell, E. Jerman, S. A. Turner, C. M. Egan, E. J. Dunne, M. C. Jurgens, K. Wynne, L. Piao, A. J. Lohan, N. Ferguson, X. Shi, K. M. Sinha, B. J. Loftus, G. Cagney, A. P. Bracken, Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase N066 to embryonic stem cell genes during differentiation. *Nat. Struct. Mol. Biol.* **19**, 1273–1281 (2012).
16. M. D. Doynova, J. F. Markworth, D. Cameron-Smith, M. H. Vickers, J. M. O'Sullivan, Linkages between changes in the 3D organization of the genome and transcription during myotube differentiation in vitro. *Skelet. Muscle* **7**, 5 (2017).
17. A. K. San Roman, A. Touglietti, D. T. Breault, R. A. Shivdasani, Distinct processes and transcriptional targets underlie CDX2 requirements in intestinal stem cells and differentiated villus cells. *Stem Cell Rep.* **5**, 673–681 (2015).
18. F. O. Bagger, S. Kinalis, N. Rapin, BloodSpot: A database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Res.* **47**, D881–D885 (2019).
19. D. Lara-Alecio, A. Weiner, E. Lorenzo-Vivas, I. Zaretsky, D. A. Jaitin, E. David, H. Kerem-Shaul, A. Mitsudo, D. Winter, S. Jung, N. Friedman, I. Amit, Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
20. J. Yang, Y. Tanaka, M. Seay, Z. Li, J. Jin, L. X. Garmire, X. Zhu, A. Taylor, W. Li, G. Euskirchen, S. Halene, Y. Kluger, M. P. Snyder, I.-H. Park, X. Pan, S. M. Weissman, Single cell transcriptomics reveals unanticipated features of early hematopoietic precursors. *Nucleic Acids Res.* **45**, 1281–1296 (2017).
21. S. Wang, F. He, W. Xiong, S. Gu, H. Liu, T. Zhang, X. Yu, Y. Chen, S. Wang, F. He, W. Xiong, S. Gu, H. Liu, T. Zhang, X. Yu, Y. Chen, Polycomblike-2-deficient mice exhibit normal left-right asymmetry. *Dev. Dyn.* **236**, 853–861 (2007).
22. I. Beerman, J. Seita, M. A. Inlay, I. L. Weissman, D. J. Rossi, Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. *Cell Stem Cell* **15**, 37–50 (2014).
23. N. K. Wilson, D. G. Kent, F. Buettner, M. Shehata, I. C. Macaulay, F. J. Calero-Nieto, M. S. Castillo, C. A. Oedekoven, E. Diamanti, R. Schulte, C. P. Ponting, T. Voet, C. Caldas, J. Stingl, A. R. Green, F. J. Theis, B. Göttgens, Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell* **16**, 712–724 (2015).
24. N. Cabezas-Wallscheid, F. Buettner, P. Sommerkamp, D. Klimmeck, L. Ladell, F. B. Thalheimer, D. Pastor-Flores, L. P. Roma, S. Renders, P. Zeisberger, A. Przybylla, K. Schönberger, R. Scognamiglio, S. Altamura, C. M. Florian, M. Fawaz, D. Vonficht, M. Tesio, P. Collier, D. Pavlinic, H. Geiger, T. Schroeder, V. Benes, T. P. Dick, M. A. Rieger, O. Stegle, A. Trumpp, Vitamin A-retinoic acid signaling regulates hematopoietic stem cell dormancy. *Cell* **169**, 807–823.e19 (2017).
25. R. Gazit, P. K. Mandal, W. Ebina, A. Ben-Zvi, C. Nombela-Arieta, L. E. Silberstein, D. J. Rossi, Fgd5 identifies hematopoietic stem cells in the murine bone marrow. *J. Exp. Med.* **211**, 1315–1331 (2014).
26. S. M. Chambers, N. C. Boles, K.-Y. K. Lin, M. P. Tierney, T. V. Bowman, S. B. Bradfute, A. J. Chen, A. A. Merchant, O. Sirin, D. C. Weksberg, M. G. Merchant, C. J. Fisk, C. A. Shaw, M. A. Goodell, Hematopoietic fingerprints: An expression database of stem cells and their progeny. *Cell Stem Cell* **1**, 578–591 (2007).
27. A. Wilson, M. J. Murphy, T. Oskarsson, K. Kaloupek, M. D. Bettes, G. M. Oser, A. C. Pasche, C. Knabenhans, H. R. MacDonald, A. Trumpp, c-Myb controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* **18**, 2747–2763 (2004).
28. J. E. Balmer, R. Blomhoff, Gene expression regulation by retinoic acid. *J. Lipid Res.* **43**, 1773–1808 (2002).
29. F. K. B. Lauridsen, T. L. Jensen, N. Rapin, D. Aslan, A. S. Wilhelmson, S. Pundhir, M. Rehn, F. Paul, A. Gladi, M. S. Hasemann, P. Serup, I. Amit, B. T. Porse, Differences in cell cycle status underlie transcriptional heterogeneity in the HSC compartment. *Cell Rep.* **24**, 766–780 (2018).
30. J. M. Bernitz, K. Rapp, M. G. Daniel, D. Shcherbinin, Y. Yuan, A. Gomes, A. Waghray, R. Brosh, A. Lachmann, A. Ma'ayan, D. Papatsenko, K. A. Moore, Memory of divisional history directs the continuous course of primitive hematopoietic lineage commitment. *Stem Cell Rep.* **14**, 561–574 (2020).
31. D. Sun, M. Luo, M. Jeong, B. Rodriguez, Z. Xia, R. Hannah, H. Wang, T. Le, K. F. Faull, R. Chen, H. Gu, C. Bock, A. Meissner, B. Göttgens, G. J. Darlington, W. Li, M. A. Goodell, Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* **14**, 673–688 (2014).
32. H. Celik, W. K. Koh, A. C. Kramer, E. L. Ostrander, C. Mallaney, D. A. C. Fisher, J. Xiang, W. C. Wilson, A. Martens, A. Kothari, G. Fishberger, E. Tycksen, D. Karpova, E. J. Duncavage, Y. Lee, S. T. Oh, G. A. Challen, JARID2 functions as a tumor suppressor in myeloid neoplasms by repressing self-renewal in hematopoietic progenitor cells. *Cancer Cell* **34**, 741–756.e8 (2018).
33. F. Rosenbauer, D. G. Tenen, Transcription factors in myeloid development: Balancing differentiation with transformation. *Nat. Rev. Immunol.* **7**, 105–117 (2007).
34. C. Pridans, M. L. Holmes, M. Polli, J. M. Wettenhall, A. Dakic, L. M. Corcoran, G. K. Smyth, S. L. Nutt, Identification of Pax3 target genes in early B cell differentiation. *J. Immunol.* **180**, 1719–1728 (2008).
35. C. Lien, C.-M. Fang, D. Huso, F. Livak, R. Lu, P. M. Pitha, Critical role of IRF-5 in regulation of B-cell differentiation. *Proc. Natl. Acad. Sci. USA.* **107**, 4664–4668 (2010).
36. H. Y. Kueh, E. V. Rothenberg, Regulatory gene network circuits underlying T cell development from multipotent progenitors. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4**, 79–102 (2012).
37. A. Bigas, L. Espinosa, Hematopoietic stem cells: to be or Notch to be. *Blood* **119**, 3226–3235 (2012).
38. E. Dzierzak, S. Phillipsen, Erythropoiesis: Development and differentiation. *Cold Spring Harb. Perspect. Med.* **3**, a011601 (2013).
39. C. Palast, A. Bergeron, V.-P. Lavalée, J. Yeh, P. Gendron, G. L. Norddahl, J. Krosl, J. Boivin, E. Deneault, J. Simard, S. Imren, G. Boucher, K. Eppert, T. Herold, S. K. Bohlander, K. Humphries, S. Lomieux, J. Hebert, G. Sauvageau, F. Barabé, GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo. *Blood* **127**, 2018–2027 (2016).
40. S. W. K. Ng, A. Mitchell, J. A. Kennedy, W. C. Chen, J. M. Leod, N. Ibrahimova, A. Arruda, A. Popescu, V. Gupta, A. D. Schimmer, A. C. Schuh, K. W. Yee, L. Bullinger, T. Herold, D. Görlich, T. Büchner, W. Hiddemann, W. Berdel, B. Wörmann, M. Cheok, C. Preudhomme, H. Dombret, K. Metzeler, C. Buske, B. Löwenberg, P. J. M. Valk, P. W. Zandstra, M. D. Minden, J. E. Dick, J. C. Y. Wang, A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016).
41. J. L. M. Rothberg, H. B. Maganti, H. Jraide, C. J. Porter, G. A. Pallidwor, C. Cafariello, H. L. Battaian, S. T. Khan, T. J. Perkins, R. F. Paulson, C. Y. Ito, W. L. Stanford, Mtf2-PRC2 control of canonical Wnt signaling is required for definitive erythropoiesis. *Cell Discov.* **4**, 21 (2018).
42. M. Beringer, P. Pisano, V. D. Carlo, E. Blanco, P. Chammas, P. Vizán, A. Gutiérrez, S. Aranda, B. Payer, M. Wierer, L. D. Croce, EPOF polycomb-like1 links Elongin and Polycomb in pluripotent stem cells. *Mol. Cell* **64**, 645–658 (2016).
43. J. Qiu, D. Papatsenko, X. Niu, C. Schaniel, K. Moore, Divisional history and hematopoietic stem cell function during homeostasis. *Stem Cell Rep.* **2**, 473–490 (2014).
44. M. N. F. Morcos, T. Zerjatke, I. Glauche, C. M. Munz, T. Ge, A. Petzold, S. Reinhardt, A. Dahl, N. S. Anstee, R. Bogeska, M. D. Milsom, P. Säwén, H. Wan, D. Bryder, A. Roers, A. Gerbaulet, Continuous mitotic activity of primitive hematopoietic stem cells in adult mice. *J. Exp. Med.* **217**, e20191284 (2020).
45. H. Geiger, G. de Haan, M. C. Florian, The ageing haematopoietic stem cell compartment. *Nat. Rev. Immunol.* **13**, 376–389 (2013).
46. S. Hayashi, T. Tenzen, A. P. McMahon, Maternal inheritance of Cre activity in a Sox2Cre deleter strain. *Genesis* **37**, 51–53 (2003).
47. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
48. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimental, S. L. Salzberg, J. L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
49. N. A. O'Leary, M. W. Wright, J. P. Brister, S. Ciufo, D. Haddad, R. M. Veigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Broer, V. Chevrenin, J. Choi, E. Cox, O. Ermoeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, Y. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully,

Downloaded from <http://advances.sciencemag.org/> on April 5, 2021

Acknowledgements

Probablemente lo mejor de este momento es pararme a pensar que si he conseguido estar escribiendo estas palabras es gracias a que estas personas han formado parte de mi vida. Por esto, y por todo lo que he aprendido en este proceso, me siento increíblemente afortunada y creo que nunca podría dar las gracias lo suficiente.

Aún así, aunque se queden cortas, tengo que decir gracias, gracias y gracias al laboratorio Di Croce, es una suerte haber formado parte de este grupo durante estos años y no puedo creer que ya hayan pasado. Gracias a Luciano, por acogerme y dejarme ser parte de él, por crear este equipo. Gracias a Sergi y a Pedro, por supervisarme, enseñarme, y ayudarme tantísimo en estos años. No habría llegado aquí sin vosotros. Sergi, gracias por tu paciencia conmigo, por hacer esfuerzos extra para hacerme ver las cosas de otro modo, por preocuparte sinceramente por mi. Y Pedro, gracias por creer en mi y darme confianza en mis primeros pasos, por enseñarme que los mejores científicos son también los más humildes. Gracias al PHF19 team, porque en los últimos años me he dado cuenta de lo que disfrutaba nuestras reuniones y discusiones, y las he echado de menos miles de veces. Gracias a Ara, por interesarse siempre por la vida de nuestra mini familia, por escuchar nuestros problemas más tontos, por siempre dar un consejo. Gracias a Ceci, por las risas y por ser una muestra de lo que es la vocación por la ciencia. A Enrique por siempre darme un feedback y motivarme. A François por llegar para ayudarme, y por lo mucho que me queda por aprender de él. Gracias a Ivo, que empezó en esto conmigo y sin saberlo ha sido mi ejemplo a seguir miles de veces. Vas a conseguir lo que te propongas, no tengo ninguna duda. A Li por estar siempre atenta, alegre, llena de vida. Gracias a Alek y a Gianni por estar siempre dispuestos a ayudar. Gracias a Anna, Marc, Paul, Mar, y Valerio, parece que hace una eternidad que os fuisteis y aún os echo de menos, ojalá hubierais estado siempre. En definitiva, gracias, porque lo mejor de estos años es mirar atrás y ver todo lo que he crecido gracias a vosotros, y lo mucho que me he divertido por el camino. Repetiría mil veces.

Gracias a Leire y Nagore del Hospital Sant Joan de Deu. Por su tiempo, empatía, ayuda, y paciencia. A todo el equipo de facilites del CRG, a Jochen, Eduard, y Eva entre otros, por explicarme siempre lo que no entendía. A Ari, Fra, e Ire, no se que habría hecho estos años sin vosotras. Se que continuamente repetimos lo mismo, pero gracias

Acknowledgements

por tener siempre la palabra que necesitaba, por haber crecido juntas, por ser casa. Estoy orgullosa de las mujeres que sois y espero parecerme a vosotras algún día. Por toda la gente que ha compartido estos años conmigo en el CRG, es increíble la cantidad de gente maravillosa que hay en este centro. Gracias porque muchos sois buenos amigos. Gracias a mis Remos, que aún sin poder comprender del todo lo que les contaba, han celebrado cada logro como suyo, y me han apoyado en los momentos no tan buenos. A Montse, que desde el principio creyó en mí más que yo misma. Cuando no estamos, somos lo que hicimos. Ojalá pudieras ver lo que hiciste conmigo y ojalá hubiera más profesores como tú en este mundo. A Cris, Ana, Lau, Ana Belén, Umberto, César, que nos hemos acompañado en este camino incluso antes de que empezara, os admiro. A María Jesús, gracias por cuidarme mejor de lo que yo supe. A mis compis de piso, a Will, Vale, Mari, Tami, Cata, Lu, Álvaro, y en definitiva a todos los que han compartido conmigo estos años en Barcelona, gracias, porque han sido infinitamente divertidos y me llevo el mejor recuerdo de esta etapa. Y a mi familia, que es preciosa. A mis abuelos, a Mari. A mi madre, por hacerlo todo con amor y sin esperar nada a cambio. A mi padre, por esforzarse en ponerme los pies en la tierra aunque él quisiera empujarme a las nubes. A los dos, por darme todo lo que tengo, por no juzgarme, por quererme de manera incondicional. Ojalá consiga seguir vuestros pasos. Y a mi hermana, que es mi mayor apoyo, mi mayor suerte, nunca podría darte las gracias lo suficiente.