Department of Gene Regulation, Stem Cells and Cancer

Centre for Genomic Regulation

Doctoral Thesis 2019

Universitat Pompeu Fabra

The role of PHF19 in myeloid leukemia

Dissertation presented by **Marc Garcia Montolio** for the degree of Doctor in Biomedicine

Work carried out under the supervision of the Dr. Luciano Di Croce in the Epigenetic Events in Cancer Group in the Gene Regulation, Stem Cells and Cancer Program, in the Centre for Genomic Regulation

Todo pasa y todo queda, pero lo nuestro es pasar,
pasar haciendo caminos, caminos sobre el mar

Antonio Machado

ABSTRACT

Polycomb group (PcG) of proteins are a group of highly conserved epigenetic regulators involved in many biological functions such as embryonic development, stem cell self-renewal, cell proliferation, and cancer. PHD finger protein 19 (PHF19) is an associated factor of Polycomb Repressor Complex 2 (PRC2) that has been proposed to regulated its activity in embryonic stem cells. PHF19 has been shown to be up-regulated in different human cancers as well as cancer cell lines. In particular, myeloid leukemia cell lines show increased levels of PHF19, yet little is known about its function. Here, we have characterized the role of PHF19 in myeloid leukemia cell lines. We have demonstrated that PHF19 depletion decreases cell proliferation and induces erythroid differentiation. Mechanistically, we have demonstrated that PHF19 regulates the proliferation of chronic myeloid leukemia cell lines through its interaction with cell cycle regulator p21. Furthermore, we have observed that MTF2, a PHF19 homolog, occupies PHF19 target genes when PHF19 is depleted. Taken together, our results show that PHF19 is a key transcriptional regulator in myeloid leukemic cell lines and suggest that PHF19 inhibition could be a potential target to be explored for myeloid leukemia treatment.

El complejo de proteínas Polycomb (PcG), es un grupo de reguladores epigenéticos altamente conservados que participan en distintas funciones biológicas como el desarrollo embrionario, la auto renovación de las células madre, la proliferación y están involucradas también en cáncer. La proteína *PHD finger 19* (PHF19), es un factor asociado al complejo represor Polycomb 2 (PRC2). PHF19 ha sido propuesta como reguladora de la actividad de PRC2 en células madre embrionarias. También se ha visto que esta sobreexpressada en diferentes canceres y líneas celulares cancerígenas. Nosotros hemos demostrado que la eliminación de PHF19 disminuye la proliferación de las líneas celulares mieloides cancerígenas. Hemos demostrado que la depleción de PHF19 en las células de leucemia crónica mieloide las induce a diferenciarse hacia eritrocitos. Mecánicamente, hemos demostrado que PHF19 regula la proliferación de esta línea celular mediante su interacción con el regulador de ciclo celular p21. Además, hemos observado que MTF2, un homólogo de PHF19, se deposita en aquellos genes donde previamente estaba PHF19. En conjunto, nuestros resultados muestran que PHF19 es un factor transcripcional clave en líneas celulares mieloides y sugieren que la inhibición de PHF19 podría ser una potencial diana para ser explorada para el tratamiento de la leucemia mieloide.

Table of contents

List of figures

DISCUSSION

List of tables

INTRODUCTION

1. Chromatin and Epigenetics

1.1 Chromatin structure

Human genomic information is encoded in the DNA sequence. To fit all this information into the cell nucleus, almost two meters of DNA need to be well organized and tightly packaged, forming a macromolecular complex called chromatin. The first layer of DNA packing structure in the chromatin is the nucleosome: 147 base pairs of DNA are wrapped around octamers of histone proteins [1,2]. The nucleosome is the basic repeated element of the chromatin, and two copies of H2A, H2B, H3 and H4 histones form it. Linker DNA and histone H1 connects nucleosomes between them and modulates its accessibility. Moreover, the N-terminal chain of histones is suitable to suffer covalent modifications that can alter the structure of the chromatin. Then, depending on the modification, chromatin could be more or less accessible for the transcriptional machinery [3], **Figure I.1**.

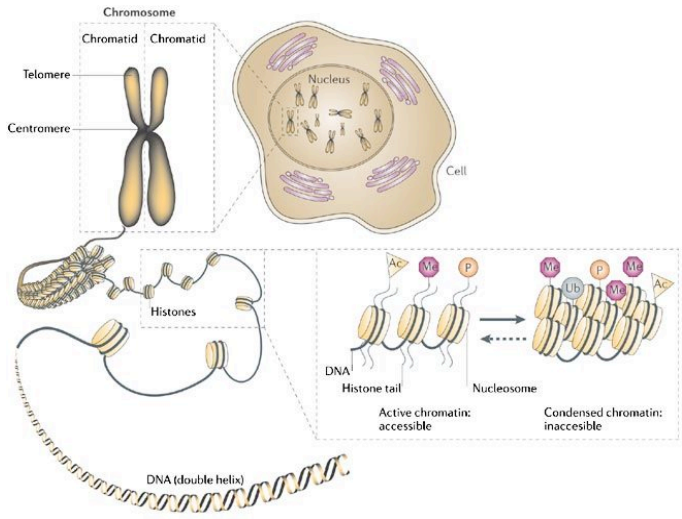


Figure I.1. Chromatin structure. DNA is tightly packaged in the nucleus. The double helix is wrapped around the histones to properly fold and then acquire a higher order of complexity until reaching the chromosome conformation. The N-terminal chain of histones is suitable to suffer covalent modifications that can alter the structure of the chromatin, then making the chromatin more or less accessible for the transcription machinery. Image from Sparmann 2006 [4]

Chromatin organization is essential, not only for a proper DNA compaction but also for basic cellular functions such as DNA repair, transcription and replication. Proper control of gene transcription is essential in multicellular organisms. For example, cell identity is determined by sets of genes whose expression or repression should be tightly coordinated. Therefore, genes that have to be expressed are present in the euchormatin, or the region of the nucleus where chromatin is more open and accessible to the transcription machinery. On the other hand, those genes that should be repressed are situated in the heterochromatin, which is the highly condensed chromatin area of the nucleus, less accessible for the transcription factors and therefore silenced

[5,6]. Moreover, there is an extra layer of organization within the heterochromatin, the constitutive and facultative heterochromatin [7]. Constitutive heterochromatin is this chromatin condensed, with poorly expressed genes, in all the cell types from a specific organism. Constitutive heterochromatin is mainly composed by tandem and transposon repeats. About a 6.5% of the human genome is considered to be constitutive heterochromatin. For instance, in humans, chromosome Y is known to have large regions of constitutive heterochromatin [8]. Facultative heterochromatin, are regions of the DNA packaged that instead of being conserved in all cell types within an organism, the genes condensed in this region can vary its expression and chromatin status depending on the cell fate. The formation of facultative heterochromatin is often ligated to the morphogenesis and differentiation of the cell [9]. An example of facultative heterochromatin is the chromosome X inactivation in female mammals, where one chromosome is packaged in facultative heterochromatin and the other X chromosome is conserved as euchromatin, and is expressed [10]. Polycomb group of Proteins (PcG) are key players in this inactivation and maintenance of the facultative heterochromatin [9].

1.2 Epigenetics

The epigenetic term *per se* defines the study of molecules and mechanisms that under the context of identical DNA sequence are able to alter the gene activity and be inherited. This broad definition includes different modifications, and is orchestrated by several converging processes such as DNA methylation, histone variants, histone post-transcriptional modifications or non-coding RNAs [11,12]. Moreover, the concept is continuously being redefined, and recently, mechanisms such as 3D conformation, chromatin states or all those regulatory processes involving molecules known to participate in epigenetic inheritance, even not being in charge of the epigenetic memory per se, are considerate as well as epigenetic features [13].

1.2.1 Histone modifications

As it is mentioned above (**Figure I.1)**, histone tails are suitable for a posttranscriptional modification (PTM), which in turn can alter the chromatin accessibility. Those modifications could happen in different residues and are mainly: acetylation, methylation, sumoylation, phosphorylation and ubiquitination (**Figure I.2)**. It has been described that PTM combinations confer different characteristics to the chromatin, which can modify its accessibility to transcription factors or other chromatin modifiers [14,15]. Some of these modifications are related with inactive chromatin, such as the tri-methylation

of the lysine 27 of the histone H3 (H3K27me3) or the tri-methylation of the lysine 9 of the histone H3 (H3K9me3). On the other hand, other modifications, such as tri-methylation of the Lysine 4 of the histone H3 (H3K4me3) or acetylation of the Lysine 27 of the histone H3 (H3K27ac) are associated with active chromatin. The large number of histone modifications and the set of different multivalent proteins that have specific domains to recognize all these histone features, gives us an idea of how tight controlled is the gene expression. A picture of some of the histone modification is shown in **Figure I.2**.

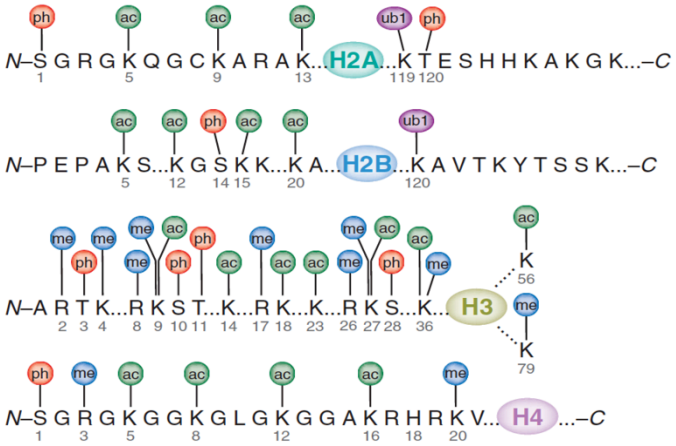


Figure I.2. Histone modifications. Some of the main histone modifications and the localizations are depicted in this figure including acetylation (ac), phosphorylation (ph), ubiquitination (ub1) and methylation (me). Figure from Bhaumik et al. 2007[15].

2. Polycomb Repressive Complex

As mentioned above, regulation of gene expression by epigenetic mechanisms is one of most important features of cell identity and cell-fate choices during development. Polycomb group (PcG) of proteins is a group of evolutionarily conserved chromatin modifying factors, which maintains repressed gene expression states and is a key element for the proper orchestration of gene expression. Polycomb maintains the repression of the target genes *via* recognition and modification of histones tails, through its chromatin remodeling activity [16]. Function of Polycomb was firstly described in *Drosophila*, where mutations in different members of this complex depicted anterior to posterior segment transformation due to ectopic expression of Homeobox (Hox) genes [17]. Although first discoveries where obtained in *Drosophila* and related with Hox clusters genes, nowadays it is known that PcG of proteins regulate a great variety of processes such as chromosome X inactivation, cell cycle, proliferation, and it is involved in stem cell biology and cancer. Classically, PcG of proteins in mammals have been subdivided in 2 complexes, the Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2) [18]. The classical model of action described for these complexes starts with PRC2 recognizing and depositing three methyl groups on the lysine 27 of the histone H3 (H3K27me3), which is subsequently recognized by the

chromobox protein (CBX) members of the PRC1. Then, PRC1 catalyzes the mono-ubiquitination of lysine 119 of histone H3 (H3K119ub), which leads to chromatin compaction [19]. Nonetheless, many others ways of PRC1 and PRC2 recruitment have been described in the last decade [18]. The following section will be focused in the Polycomb Repressive Complex 2 and the different associated factors that modulate its recruitment to the chromatin.

2.1 Polycomb Repressive Complex 2

As mentioned before, PRC2 is a multiprotein complex responsible for depositing, through its catalytic subunit, the mono-, di-, and tri-methyl group on the lysine 27 of the histone H3. PRC2 is composed by three core components essential for the catalytic function, Embryonic Ectoderm Development (EED), Suppressor of Zester 12 (SUZ12) and Enhancer of Zester 2 (EZH2) or its closely related EZH1, which contains the catalytic domain [20-22]. In *Drosophila*, Polycomb proteins are recruited to the DNA by recognizing the Polycomb Response Elements (PREs), regulatory sequences used as a platform for Polycomb binding [23]. However, despite great efforts, there is no a clear consensus about PREs reported in mammals. In contrast, different ways of PRC2 recruitment have been proposed [24]. The complexity underlying PRC2 in mammals resides in the different associated factors that may be able to modulate its

recruitment and activity. Together with RBBP4 or RBBP7, that have a lower stoichiometry and are not necessary for the methyltransferase activity [25,26], several associated factors have been described to interact with PRC2 [22]. Nowadays, PRC2 is thought to form two different subcomplexes (PRC2.1 or PRC2.2), depending on the combination of the partners with whom it associates [27] (**Figure I.3**).

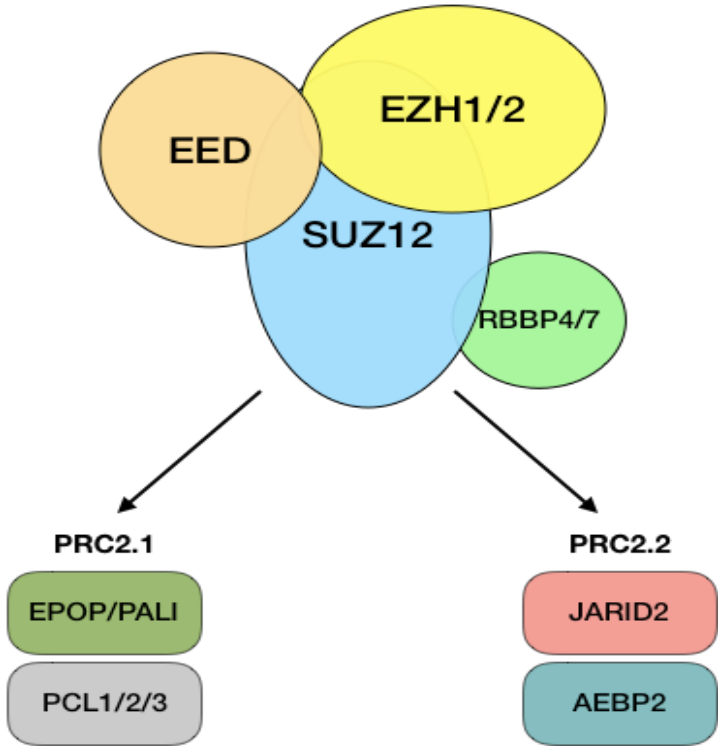


Figure I.3. **PRC2 composition and its associated proteins.** Schematic composition of PRC2. PRC2 can form two alternative subcomplexes, where either it interacts with EPOP or PALI together with one of the Polycomb Like proteins (PCL), forming what is known as PRC2.1. Alternatively, it can associates with JARID2 and AEBP2 thus forming PRC2.2 [27].

2.1.1 JARID2

JARID2 is one of the most studied associated PRC2 factors. It is a member of the Jumonji family group [28], and was initially described as a nuclear protein essential for embryogenesis, heart development and neural tube formation [28,29]. JARID2 enhances H3K27 methylation activity and facilitates PRC2 recruitment to its target genes [30]. Additionally, it has been demonstrated that JARID2 depletion reduces the recruitment of PRC2 to the chromatin in mouse embryonic stem cells (mESCs) [31]. Accordingly, it has also been described that JARID2 stabilizes the interaction of PRC2 with the chromatin due to its nucleosome-binding domain [32]. In addition, more recently, it has been shown that JARID2 interacts with PRC2 through EED, and that its depletion leads to an impaired orchestration of gene expression during cell lineage commitment [33]. Other studies have been done with JARID2 describing its implication in the recognition of the H2A119ub mark, which is deposited by PRC1 [34]. Moreover, its been described that PRC2 recruitment could be regulated by micro RNA and long non-coding RNA together with JARID2 [35]. Altogether, these evidences reflect the complexity surrounding the role of this protein and its function(s) within the PRC2.

2.1.2 AEBP2

AEBP2, similar to JARID2, is a PRC2 associated protein, needed for an optimal enzymatic activity [36]. AEBP2 has been purified together with PRC2 due to its interaction with SUZ12, EED and RBBP4/7 [20]. It has two isoforms, and both have a DNA-binding domain that links PRC2 with the chromatin [37]. Deletion of AEBP2 has been performed in mESCs, and unexpectedly, it does not seem to be essential for PRC2 function. Moreover, seems that its depletion in mESCs slightly increases H3K27me3 signal in PRC2 targets, suggesting an aberrant PRC2 function [38].

2.1.3 PALI

PRC2.1 shows more complexity than PRC2.2, since it has more associated factors that can be part of it. One of them is the PALI proteins family, which is composed by 2 homologs: PALI1 and PALI2, encoded by the *LCOR* and *LCORL* isoforms gene loci, respectively. PALI1 and PALI2 have a CTBP binding domain and a PALI interaction with PRC2 (PIP) domain. It is essential for mouse development and is reported to promote methyltransferase activity of PRC2 *in vivo* and *in vitro* [39].

2.1.4 EPOP

EPOP is a new discovered PRC2 associated factor. EPOP was reported as PRC2 partner responsible for maintaining H3K27me3 at those lineage specific genes essential to be repressed in order to maintain ES cells pluripotency [40]. Expression data also indicates a role of EPOP in the regulation of neuronal cell differentiation during embryonic development [41]. In contrast with the classical function of PRC2 repressing genes, EPOP has been recently described as an Elongin BC linker to PRC2 that helps to sustains low levels of expression of PRC2 targets and modulates JARID2-PRC2 occupancy [42,43].

2.1.5 PCL1/ PHF1

Polycomb like protein (PCL) was reported in *Drosophila* in the early 80s, showing to be involved in recruiting *Ez*, the paralogue of the mammal EZH2, and required for H3K27me3 at PRC2 targets [44-46]. In *Drosophila*, PCL contains two PHD domains and a non-functional Tudor domain. There are three orthologous in mammals: PCL1/PHF1, PCL2/MTF2 and PCL3/PHF19, all of them having a functional Tudor domain able to recognize the H3K36me3, two PHD domains, an Extended Homology (EH) domain and a C-terminal domain or also called chromo-like domain [47]. PHF1 was the first PCL orthologous being identified in humans based on its similarity with the *Drosophila* sequence [48]. Although PHF1 is not

reported to be the main PRC2 recruiter, it is well demonstrated that localizes with PRC2 targets and its depletion reduces levels of methylation on its target genes, thus leading an increase of expression. Moreover, it contributes to increase the enzymatic activity of PRC2 in *in vitro* assays [49,50]. Furthermore, as the others PCL protein does, it has been shown to recruit the PRC2 complex to CpG islands [51]. The Tudor domain of PHF1 is known that interact with H3K36me3 [52], but also was described that could interact with H3K27me3, thus having a dual role in relation with PRC2 [53]. Since PHF1 is not expressed in mESCs [54], most of the studies has been performed in adult cell lines, where has been observed different functions, such as interaction with DNA damage [55]. Moreover, PHF1 has been described to have a role in cancer, preventing the expression of the tumor suppressor like HIC (Hypermethylated in Cancer) and also p53 in osteosarcoma cell lines [56,57].

2.1.6 PCL2/ MTF2

Among the three PCL proteins existing in mammals, MTF2 is the most expressed in mESCs [54]. Since MTF2 is homologous of the others PCL proteins, MTF2 shares all the same domains of PHF1 and PHF19 (described in the next section). In contrast with what it is expected from a Polycomb related protein, depletion of MTF2 in mESCs promotes self-renewal and an undifferentiated state. Moreover, the differentiation capacity is impaired. Knock-down of MTF2 was reported to increase general protein levels of EZH2 and H3K27me3. But in contrast, in the chromatin and especially on MTF2 direct target genes, the depletion causes decrease of PRC2 members and consequent reduction of H3K27me3 mark [54]. Other studies have revealed that MTF2 is necessary for the X chromosome silencing, and that its depletion reduces the PRC2 members presence independently of the Xist RNA [58]. In relation with the recruitment of PRC2 to chromatin through MTF2, it has been recently shown that MTF2 can bind unmethylated CpGs, distinguishing those chromatin regions that are less twisted because of the lack of methyl group [59]. In vivo, a recent report has shown that MTF2 is essential during the process of the hematopoietic development, in particular for erythroid differentiation. MTF2 KO mice were not able to born because of a severe anemia. Mechanistically, it has been proposed that MTF2 inhibits some key genes of the WNT pathway. This WNT inhibition is required for complete erythropoiesis [60].

2.1.7 PCL3/ PHF19

PHD finger protein 19 (PHF19) was the last PCL member identified. Even though, it has been widely studied its relation with PRC2. PHF19, in humans have two main isoforms, so-called short and long PHF19. As described for the other PCL proteins, PHF19 has two PHD domains, a Tudor domain, an Extended Homology (EH) domain and a C-terminal domain, also called chromo-like domain (**Figure I.4)**. The short PHF19 isoform only contains the Tudor domain and one PHD domain [61]. Further studies demonstrated that both isoforms could interact in HEK293 with EZH2 through the Tudor and the PHD2 domain [62]. As for the other PCL family members, the mechanistic role of PHF19 has been studied in mESCs. I was described that PHF19 is an associated protein of PRC2 but without JARID2 [63,64].

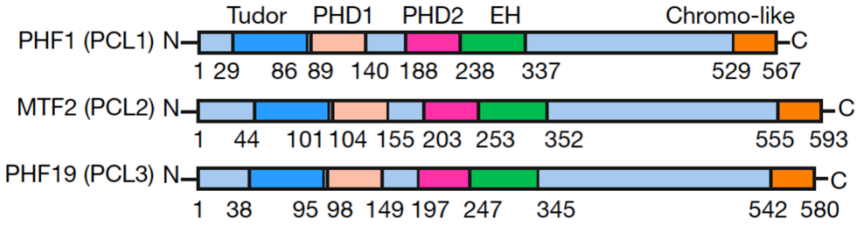


Figure I.4. **Domains of human PCL proteins**. All the PCL proteins have the same domains although the length of each one is different; crystal conformation varies as well between them as Li and Liefke described in 2017 [51].

Genome wide studies have demonstrated that most of PHF19 targets are also targets of the PRC2 member SUZ12, and are decorated with H3K27me3 mark [63-65]. Notably, its depletion in mESCs reduces the SUZ12 binding on those targets in

common with PHF19 and, consequently, generates a reduction in the H3K27me3 mark deposition [64].

With regard to the phenotype that mESCs acquires upon depletion of PHF19, it was reported a spontaneous differentiation and thus a decrease of pluripotency markers [63-65]. However, the function of PHF19 in differentiation has created further debate. *Brien et al 2012* and *Ballaré et al 2012*, found defects in the embryoid body formation when PHF19 was depleted. Moreover, *Ballaré et al 2012* found that the teratomas formed, when PHF19 was depleted, were smaller than the control, furthermore teratomas contained tissues representative of the three germ layers but with an over-representation of ectoderm tissue. On the other hand, *Hunkapiller et al 2012* [63], did not found these differences in embryoid body formation neither in the three germ layers formation.

Regarding the mechanism of how PHF19 links PRC2 with the chromatin, it was described that the Tudor domain contains an aromatic cage essential for the functionality of the PRC2 that joins and subsequently deposit of the H3K27me3 mark [63]. Moreover, this Tudor domain is able to recognize the H3K36me3 mark, a well-known mark associated to active transcription. Notably, the recognition of this transcription mark was found to be necessary for the deposition of the repressive mark H3K27me3 [64-66]. Apart from the recognition of the H3K36me3 mark and the deposition of the H3K27me3, it was shown that PHF19 also interacts in mESCs with

demethylase proteins, such as NO66 or KDM2b, suggesting that those demethylases are needed for erasing the active mark H3K36me3, thus permitting PRC2 to deposit the H3K27me3 mark (**Figure I.5**) [64,65].

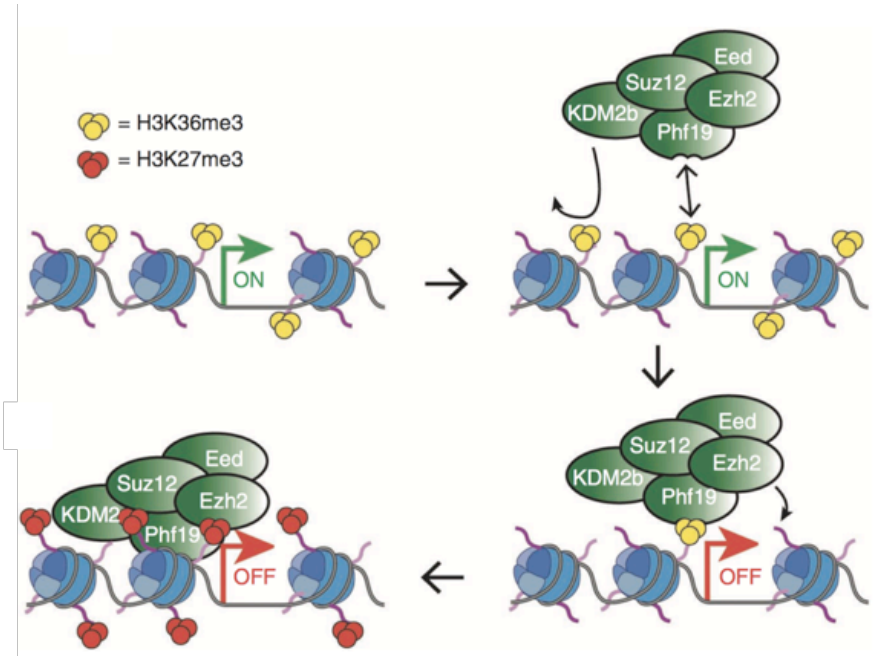


Figure I.5. Recruitment of PRC2 by PHF19. Schematic picture of PHF19 bringing PRC2 and KDM2B to the chromatin. The Tudor domain of PHF19 recognizes the active mark H3K36me3, and then the demethylase erases this one in order to let PRC2 deposit the consequent repressive mark H3K27me3. Image modified for *Ballaré et. al.* 2012 [64].

Regarding the role of PHF19 in human diseases, both isoforms has been observed to be overexpressed in many different cancers and cell lines [61,62]. In melanoma, it has been described that PHF19 control the balance between the growth and the invasiveness, since depletion of PHF19 reduces the proliferation rate, but increases the trans-endothelial migration capacity [67]. Moreover, PHF19 has also been found to be overexpressed in patient tumors data in

Glyoblastoma multiforme [68]. In Glioblastoma has also been reported recently that PHF19 is enrolled in the regulation of the WNT signaling pathway. Moreover, depletion of PHF19 reduces the proliferation and migration capacity of glyoblastoma cells [69]. Other studies in Hepatocellular Carcinoma (HCC), showed that PHF19 is overexpressed and that its targeting with microRNAs reduces *in vitro* and *in vivo* the migration capacity, invasion and proliferation [70]. A recent study shows PHF19 potential role in avoiding senescence when it is activated through AKT pathway in the context of $CD8^+$ T cells [71]. All these studies reflect the different roles that PHF19 could have depending on the cellular context.

3. Leukemia

Hematological malignancies are a group of diseases related with the blood system. Among them, leukemia is the generic name for all types of cancers related with the accumulation of undifferentiated cells called blasts, either in the blood system, in the bone marrow or in the lymph nodes [72]. According to WHO, leukemia can be subdivided depending on the cell lineage of origin in myeloid or lymphoid, and depending on the aggressiveness and the speed of development, can be acute or chronic [73] (**Table I.1**).

Myeloid leukemias are types of leukemias where the cancerous cells derive from the myeloid precursors. Myeloid precursors give rise, in normal condition, to erythrocytes, platelets, monocytes or granulocytes. On the other hand, Lymphocytic leukemias the origin of cell are the precursors of the lymphocytes [74].

Acute leukemia is characterized by the fast growth of immature blood cells. These cells go into the bloodstream and disseminate to other organs. An immediate treatment reaction is needed because of the faster capacity of those cells to collapse the bone marrow. Children are the most common patients of this type of leukemia. Regarding chronic leukemia, it is characterized by an abnormal increase of relatively mature blood cells that slowly accumulates in the

bone marrow or hematopoietic organs. Usually, a chronic phase of a leukemia leads to a blast crisis over the years, resembling to an acute leukemia. Chronic is more common in older patients, although it can also affect at young ages [75,76].

Table I.1 General classification of leukemia

Cell of origin	Acute	Chronic
Myeloid precursors	Acute myeloid leukemia (AML)	Chronic myeloid leukemia (CML)
Lymphoid precursors	Acute lymphoblastic leukemia (ALL)	Chronic lymphoblastic leukemia (CLL)

3.1 Myeloid leukemia

As is mentioned before myeloid leukemias give rise from myeloid precursors. Myeloid leukemia basically includes acute myeloid leukemia (AML) and chronic myeloid leukemia (CML).

AML is an heterogeneous clonal disorder characterized by infiltration of the bone marrow, blood, and other tissues by proliferative, clonal, abnormally differentiated and immature myeloid progenitors cells [75]. The classification of AML is always under-revision and continuously appears new classifications, depending on the morphology, phenotype and genetic background. One of the most commonly used is the

French-American-British (FAB) group classification (**Table I.2**) [77].

Type	Name	Cytogenetics	%of cases
M0	Acute myeloblastic leukemia with minimal differentiation		3 %
M1	Acute myeloblastic leukemia without maturation		15-20%
M2	Acute myeloblastic leukemia with maturation	t(8;21) (q22;q22),t(6;9)	25-30%
M3	Acute promyelocytic leukemia (APL)	t(15;17)	5-10%
M4	Acute myelomonocytic leukemia	inv(16) (p13q22),del(16q)	20 %
M4eo	Acute myelomonocytic leukemia with abnormal eosinophils	inv(16), t(16;16)	5-10%
M5	Acute monoblastic (M5a) and Acute monocytic leukemia (M5b)	del (11q), t(9;11), t(11;19)	2-9%
M6	Acute erythroid leukemia		3-5%
M7	Acute megakaryocytic leukemia	t(1;22)	3-12%

Table I.2. FAB classification of AML.

The disease incidence increases with age, having adult patients poorer outcome than younger patients. Treatment has not changed substantially in the last years. Induction of chemotherapy with differentiation agents like cytarabine and anthracycline are the main approach. Nonetheless, new treatments with higher doses of daunorubicin are under study. Moreover, since new scientific advances are being achieved, treatments against new agents depending on the

AML genetic landscape are now a day a major challenge [75]. An important genetic characteristic of AML is the high prevalence of chromosomal rearrangements that usually gives a prognostic significance. These translocations are traduced in abnormal fusion proteins, usually transcription factors, which cause differentiation arrest. This feature in combination with other genetic mutations that disrupts genes controlling proliferation, results in the uncontrolled growth of immature cells [78].

CML is a clonal disorder of the hematopoietic system characterized by the unregulated growth of non-functional myeloid, erythroid, and platelets in the peripheral blood, and a marked myeloid hyperplasia in the bone marrow [76]. CML accounts for 15–25% of all adult leukemias and 14% of leukemias overall (including the pediatric population, where CML is less common). CML was the first type of leukemia associated to a chromosomal translocation, the Philadelphia chromosome (t(9;22)(q34;q11.2)). The translocation resulted in a fusion protein so called BCR-ABL, which generates a tyrosin-kinase activity that activates signaling cascades in charge of regulation cell cycle, speeding the cell division. Nonetheless, BCR-ABL also inhibits DNA repair, making the cell more susceptible to have extra genetic mutations [79]. CML disease is often divided in three phases. (1) The chronic phase, which is the asymptomatic phase and where most of the patients are diagnosed (around the 90%). (2) The

accelerated phase, in this stage patients have some symptoms such as increase white blood, thrombocytosis, unusual number of blasts in the peripheral blood, co-occurrence of more mutations a part of the fusion protein, and resistance to the classical tyrosine kinase inhibitor (Imatinib). (3) The final phase is characterized by a blast crisis, where the disease behaves like acute myeloid leukemia with rapid progression and poorer survival. The mechanisms of progression, and cytogenetic evolution to blast crisis remain largely unknown. Genetic instability as a consequence of the Philadelphia chromosome translocation might be responsible for additional chromosomal aberrations or mutations frequently seen in blast crisis. As is mentioned before, Imatinib is nowadays the most selected drug for the chronic phase treatment [80], but new studies against other target genes mutated due to the genetic instability are under investigation [81]. Moreover, almost 30% of the patients develop resistance to the tyrosine kinase inhibitor Imatinib, and it is in this context where a second line of treatment focused on epigenetic mutations has had more relevance in the last years [82].

Although not being classified as leukemia per se, Myelodisplasic Syndromes (MDS) are a very heterogeneous group of myeloid disorders characterized by peripheral blood cytopenias and increased risk of transformation to acute myeloid leukemia (AML) [83]. It is also an asymptomatic

disease in early stages. The progression of MDS to AML is a good example of the carcinogenesis multi-steps paradigm, in which a series of mutations occurs in an initially normal cell and transforms it into a cancer cell [84]. Moreover, MDS is characterized by frequent epigenetic abnormalities, including mutations in DNA methylation controllers, such as DNMT3a or TET2, or histone modifiers, such as EZH2 [85,86]. Because of its epigenetic nature, treatments with DNA methylation inhibitors are highly responsive. In addition, other epigenetic acting drugs are being explored, such as inhibitors of histone deacetylases [87].

3.2 PRC2 in myeloid leukemia

As mentioned before, epigenetic mutations has been studied in the last years in the context of myeloid leukemias, particularly for a potential target treatment in cases of relapse [82,87,88]. PRC2 members are misregulated in many cancer types, and very often in leukemias, and its function can vary depending on the cellular context [89].

EZH2 is one of the most studied proteins in the context of myeloid leukemia. Its location in the chromosome 7 of the genome is a hot spot placement for chromosomal deletions and somatic mutations [90]. Moreover, EZH2 mutations have been observed in patients with myeloid malignancies, and epidemiologic data clearly correlates deletion of EZH2 with poorer outcome in MDS [91 92,93]. This could reinforce the idea

of EZH2 being a tumor suppressor gene. On the contrary, it has also been reported that EZH2 is overexpressed in chronic myeloid leukemia initiating cells, and that its inactivation inhibits the cell growth [94]. This result, together with another study claiming that EZH2 inhibitor sensitizes CML cells to Tyrosine Kinases inhibitors, open a research line where EZH2 could be used as a target for myeloid leukemia treatment [95]. Further studies have been performed in this direction, and it has been recently reported that the oncogene MYCN regulates EZH2 expression in CML cells and this leads to a repression of p21, a well-characterized gene that arrests cell cycle, thus leading to an increase in proliferation and block in differentiation [96]. Reinforcing the idea of EZH2 as an oncotarget, it was also described that in MLL-AF9 rearrangement protein dependent leukemia, EZH2 is necessary for blocking the differentiation pathway, thus increasing the leukemogenicity [97,98]. Altogether these results show that EZH2 can have different roles and it has to be considered in a different way depending on the cellular context. Both loss of function and gain of function of EZH2 perturb the normal hematopoietic system, and the H3K27me3 mark that EZH2 deposits must have tight a regulation for a proper cell behavior [99]. In line with it, a novel study suggested that depending whether the AML is in an initial stage or in a maintenance phase, EZH2 could be a tumor suppressor or a facilitator of the disease in already established AML [100]. Although there is still controversy about the targeting of

EZH2, in 2014, SKI-606 a drug inhibiting this PRC2 component was approved for the treatment of adult patients positive for the Philadelphia chromosome CML and resistant or intolerant to the previous therapies [101].

As happens with EZH2, SUZ12 has been also found to have a dual role depending on the type of disease. In myelodisplasic neoplasms, SUZ12 locus has been frequently found deleted, and loss of function mutations of SUZ12 has also been observed [102,103]. On the other hand, the role of SUZ12 changes when the myeloid leukemia is due to the fusion protein PML-RAR. SUZ12 interacts with the fusion protein, and then leads to a repression on PML-RAR targets genes. Moreover, depletion of SUZ12 not only reduces the general methylation status of the chromatin, but also leads to a decrease in proliferation and an increased differentiation. In addition, when PML-RAR dependent cells are treated with retinoic acid, a standard treatment for PML-RAR dependent leukemia, SUZ12 levels are reduced at PML-RAR target genes, and there is a consequent reduction of H3K27me3 thus leading to a cell differentiation [104].

There is also controversy regarding the role of EED in myeloid leukemia. It has been described that mice knock-out for EDD are not viable, but its haploinsufficiency causes an MDS-like disease. Moreover, haploinsufficiency of EED confers a proliferative advantage that together with the overexpression of an oncogene like *Evi1* promotes

progression to AML [105]. In addition, point mutations in the aromatic cage responsible of the interaction of EED with EZH2, has been reported as recurrent in human myeloid disorders reducing the propagation of H3K27me3 mark [106,107]. On the other hand, EED inactivation impaired *ex vivo* cell growth of hematopoietic precursors transfected with the MLL-AF9 fusion myeloid leukemia oncoprotein. EED depletion up-regulated tumor suppressor *p16ink4a* and *p19arf*, and at the same time indirectly downregulated MYC targets genes in MLL-AF9 leukemia [98]. Noteworthy, other studies have found that also in MLL-AF9 system, not only depletion of EDD, but also EZH2 and SUZ12 induced cell cycle arrest and differentiation [108]. So, although PRC2 core components are prone to be targeted for myeloid leukemia treatment, it is important to consider the specificity of each type of leukemia and its developmental stage.

Relevant to for the PhD project presented here, during the development of normal hematopoietic system, PHF19 expression decreases through differentiation, as is shown in **Figure I.6**. It is worth to remark that until now, the role of PHF19 in myeloid leukemia has not been assessed.

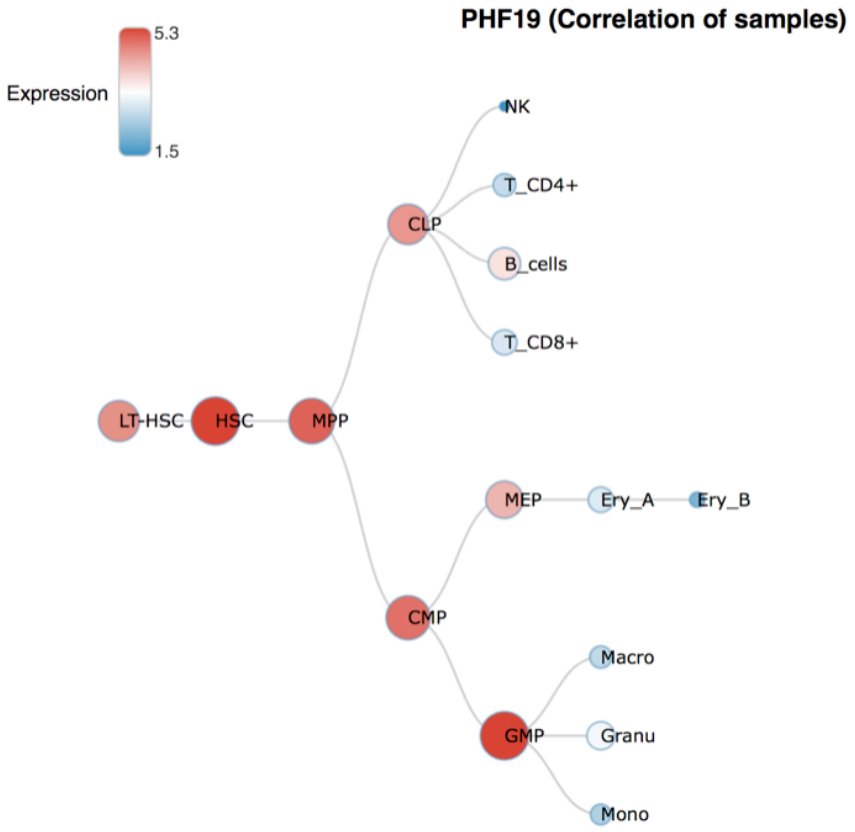


Figure I.6. PHF19 expression through hematopoietic development. Schematic picture showing the expression levels of PHF19 from the apex of the hematopoietic tree where resides the niche of the hematopoietic stem cells, to the differentiated subsequent cell types. Data extract from *Bagger, F et.al.* 2018[109] in reference to the paper from *Lara-Estiaso et.al.* 2014[110].

3.3 p21 in myeloid leukemia

One of the hallmarks of cancer is the misregulation of the cell cycle, which leads to cell over-proliferation and tumor overgrowth [111]. In leukemia, the deregulation of cell cycle genes contributes to the development of acute and chronic

myeloid leukemia by promoting the hyperproliferation of undifferentiated blasts [112,113]. Moreover, hyperproliferation bears an increased probability to accumulate genetic mutations that, as mentioned before, are key for the disease progression. Then, to better understand leukemogenesis, it is crucial to determine the role of proto-oncogenic pathways in charge of the regulation of cell cycle in myeloid malignancies. Regarding the cell cycle progression, it has been proposed that p53, the classic tumor suppressor, regulates, among others, the expression of p21, a major cell cycle regulator of the transition between G1 to S phase [114]. In a tumor environment, the expression of p21 is repressed in order to facilitated self-renewal and cell division. It has been described that epigenetic inhibition carried by PRC2 over p21 is required for the tumor progression in many cancers but also in myeloid leukemia [96,115].

In AML, it has been observed a misregulation of this p53-p21 pathway [116]. Moreover, it has been also reported that in AML patients and in patient derived cell lines, p53 is mutated or deleted [117,118]. In this sense, it has been also shown that p21 is frequently deleted in the AML1-ETO dependent AML, to promote leukemogenesis [119]. In other systems like PML-RAR dependent AML, it has been observed that p21 is involved in the cell cycle arrest and cell differentiation when cells are treated with Retinoic Acid [120].

The relevance of p21 in arresting cell cycle is not exclusive of AML, but it is also important in CML. Loss of function of p53

in CML patients appears to be a common feature and is associated with CML progression [121,122]. Thus, upon the disruption of the p53-p21 pathway, CML cells can grow uncontrolled leading to a blast crisis [123]. In addition, when p53 losses his function, the cell cycle regulation to avoid tumor progression relies in p21 [124-127]. Therefore, when PRC2 adopts its oncogenic role, multiple evidences highlight its mechanistically contribution in preventing the expression of p21 in CML [96]. Then tackling PRC2 to induce the expression of p21 in CML could potentially be a treatment approach.

OBJECTIVES

Research in our group focuses in the investigation of the epigenetic and transcriptional mechanisms involved in differentiation and cancer. Particularly, we are interested in understanding the role of Polycomb group of proteins and its associated factors in myeloid leukemia. We and others have described the role of PHF19 in ESC, but little is known regarding its role in leukemia.

The general objective of this thesis is to describe the role of PHF19 in the context of myeloid leukemia. To achieve that we focused on the following main aims:

1. Investigate the outcome of PHF19 depletion myeloid leukemia cell lines.
2. Identify the mechanism by which PHF19 regulates transcription.

RESULTS

1. PHF19 depletion impairs cell growth in AML cell lines

Previous works from several laboratories including ours have shown the potential therapeutic strategy of targeting PRC2 in myeloid leukemia [95,104]. As mentioned in the introduction, the role PHF19 in myeloid leukemia has not been addressed yet. According to the Protein Atlas database, Myeloid Leukemia cell lines (above all K562, NB4 and HL60 cell lines) have a higher expression of PHF19 with respect to most of the other cancerous cell lines **Figure R.1**.

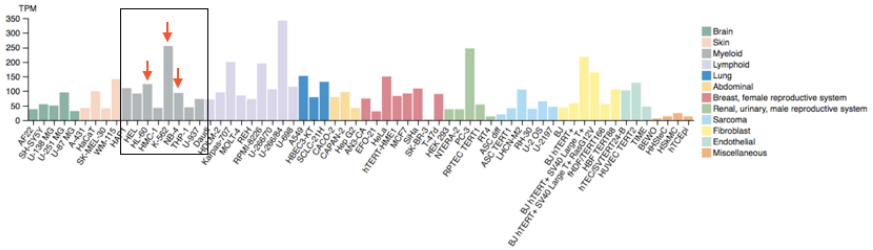


Figure R.1. PHF19 expression data from cancerous cell lines. Average values of Transcript per Million (TPM) from 2 independent RNA-seq data set. Red arrows point to the three different cell lines used, NB4, HL60 and K562. Figure modified from Protein Atlas web page (www.proteinatlas.org).

To explore the role of PHF19 in these cell lines, we generated two independent short-hairpin RNAs (shRNA) that efficiently reduced PHF19 mRNA levels, sh#1 and sh#2. We infected with lentiviruses the two independent shRNAs, and cells were selected with puromycin (**Figure R.2)**.

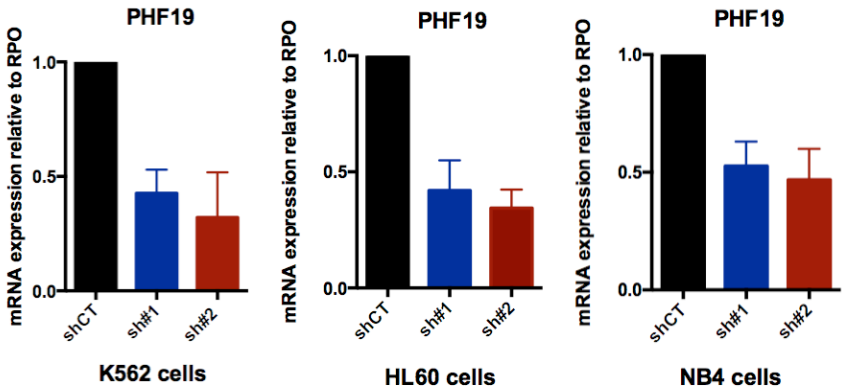


Figure R.2. PHF19 depletion in Myeloid Leukemia cell lines. Panel of real time PCR (RT-PCR) analysis of control and PHF19 depleted myeloid leukemia cell lines. Expression values were normalized by the expression of housekeeping gene RP0 and is represented relative to shCT. Errors bars are the standard deviation (SD) of four independent experiments.

Based on our previous experience on depletion of PRC2 core components in myeloid leukemia cell lines, we wondered whether the depletion of PHF19 similarly compromises the proliferation capacity of the myeloid cell lines. These results presented in **Figure R.3** demonstrated that depletion of PHF19 affected cell growth of several myeloid cell lines.

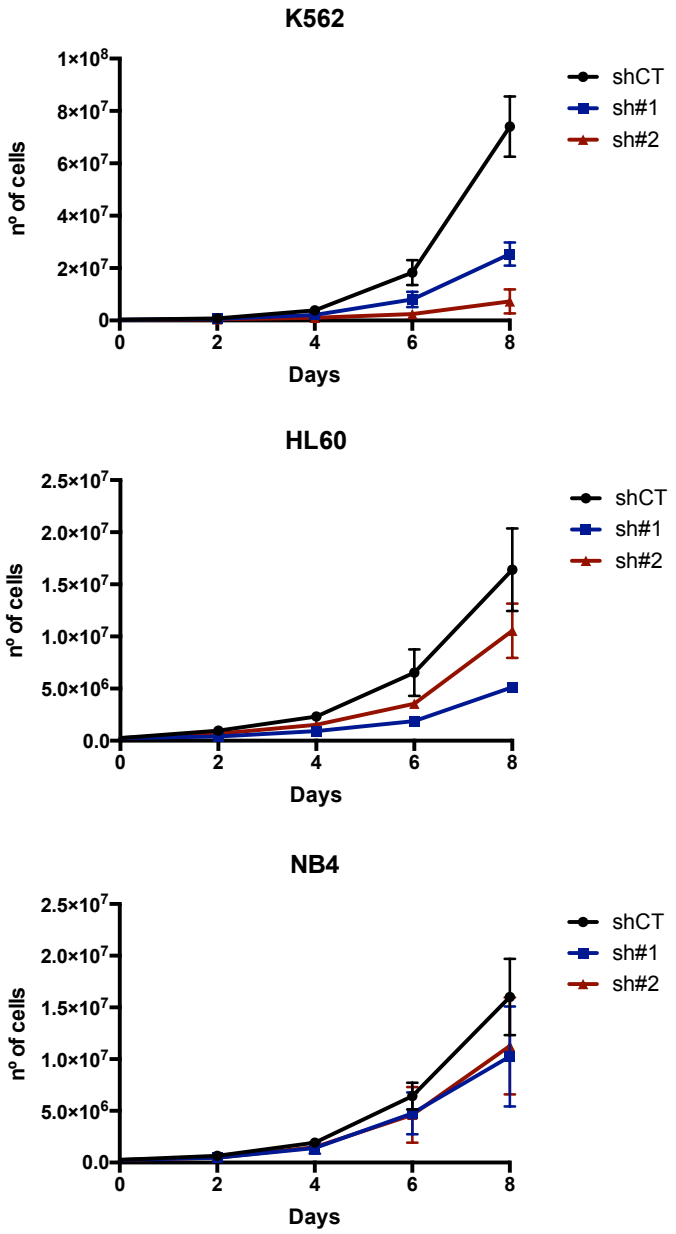


Figure R.3. PHF19 depletion caused a decrease of cell growth. Accumulative growth curve of shCT, sh#1 and sh#2. Y-axis shows number of cells and X-axis shows days after selection. Data are means and standard deviation of three independent experiments.

To further understand the basis of the observed decrease in cell growth caused by the depletion of PHF19, and to investigate whether this decrease in proliferation could be due to an increase in cell death, we next assessed the staining of the apoptotic maker Anexin V by flow cytometry. We observed that PHF19 depletion did not induce apoptosis in myeloid leukemia cell lines, cell death control were the non-infected cells under selection of puromycin. (**Figure R.4).**

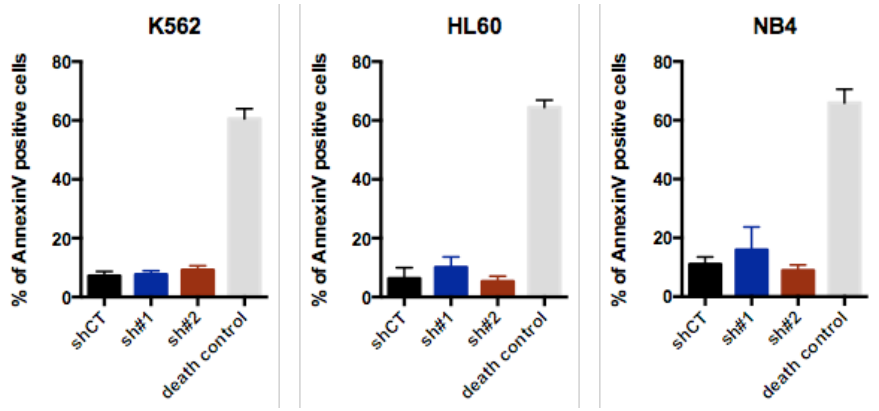


Figure R.4. PHF19 depletion did not increase cell death. Apoptosis assay assessed by Anexin V staining of shCT, sh#1, sh#2. Cell death control was the non-infected cells under the same selection of Puromycin, during 3 days. Y-axis shows percentage of positive Anexin V cells, in three independent experiments, and standard deviation between them.

Among the three cell lines, K562 was the cell line with the highest PHF19 expression (**Figure R.1**). Moreover, the effect of PHF19 depletion in cell growth was more evident in this cell line (**Figure R.3**). Interestingly, we could observe that K562 PHF19 depleted pellets were reddish when compared to the control (shCT) **Figure R.5** (see next section)**.**

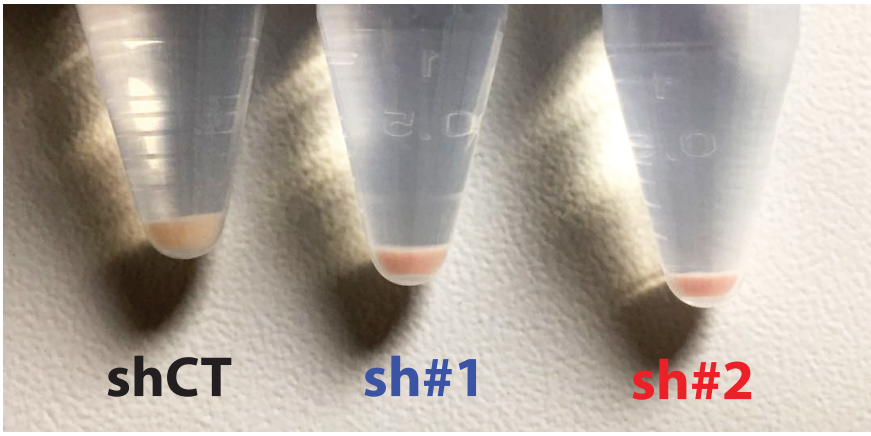


Figure R.5. Color of pellets after PHF19 depletion. Picture taken after five days of selection with puromycin once have been infected with shRNAs. Pellets related to sh#1 and sh#2 are visually more reddish than the shcontrol after 2 washes of PBS.

We thus decided to follow up with the experiments in K562 cell line to further characterize the role of PHF19. We have observed that PHF19 depletion decreases cell growth of CML cell line K562, while the cell death percentage was not increased. To further characterize the phenotype caused by the depletion of PHF19, we assessed a BrdU staining and a cell cycle analysis (**Figure R.6)**. The results obtained showed that PHF19 depletion triggers a cell cycle arrest and a decrease of BrdU incorporation.

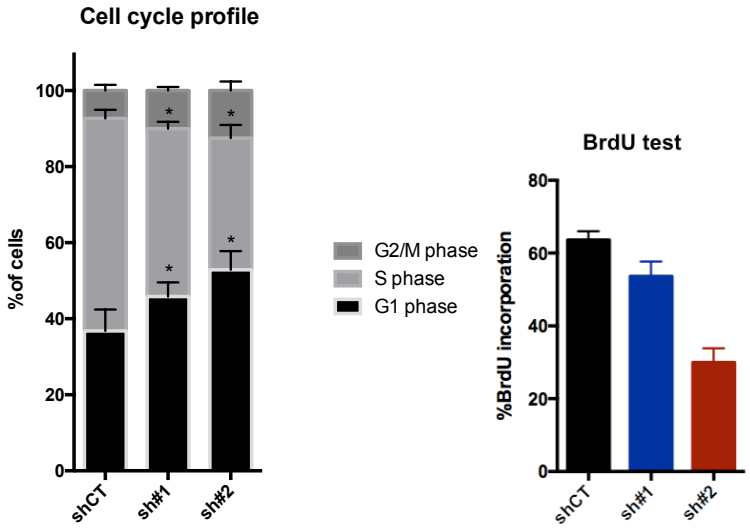


Figure R.6. PHF19 depletion arrests K562 cell cycle. Left panel show the cell cycle profile of the K562 after being depleted by sh#1 and sh#2. Bars are the average of the different phases from the cell cycle (G1, S and G2/M) of five independent experiments. Error bars are the standard error of the mean (SEM). Significance has been proven by a paired student t-test. Right panel shows the mean + standard deviation (SD) of the percentage of cells incorporating BrdU in three independent experiments.

Overall, our results showed that PHF19 depletion in myeloid cell lines, and especially in the chronic myeloid leukemia cell line K562, compromised the cell proliferation due to an arrest on the cell cycle, but without inducing apoptosis.

2. PHF19 depletion promotes K562 erythroid differentiation

To better understand the molecular mechanisms underlying the effect of PHF19 depletion in K562, and to analyze the transcriptomic changes upon PHF19 depletion, we next performed RNA sequencing (RNA-seq) in PHF19-depleted and control K562. We sequenced RNA from two different time points (two days and four days after puromycin selection) of shCT, PHF19 sh#1 and sh#2 from two independent experiments. **Figure R.7** shows the principal component analysis (PCA) from four replicates of each condition, which demonstrated the homogeneity between among replicates: all genes expressed led to a clustering of the samples in a PCA plot irrespectively of the time point, but depending on the shRNA used.

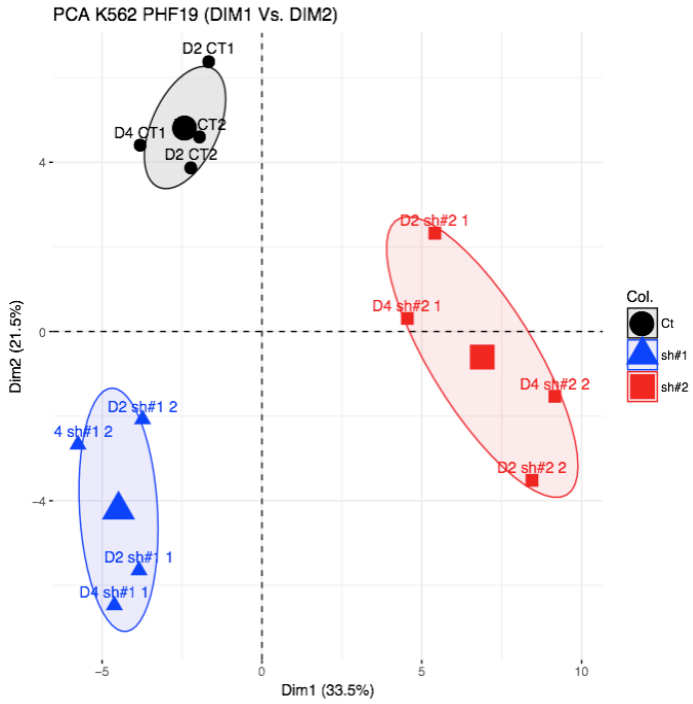
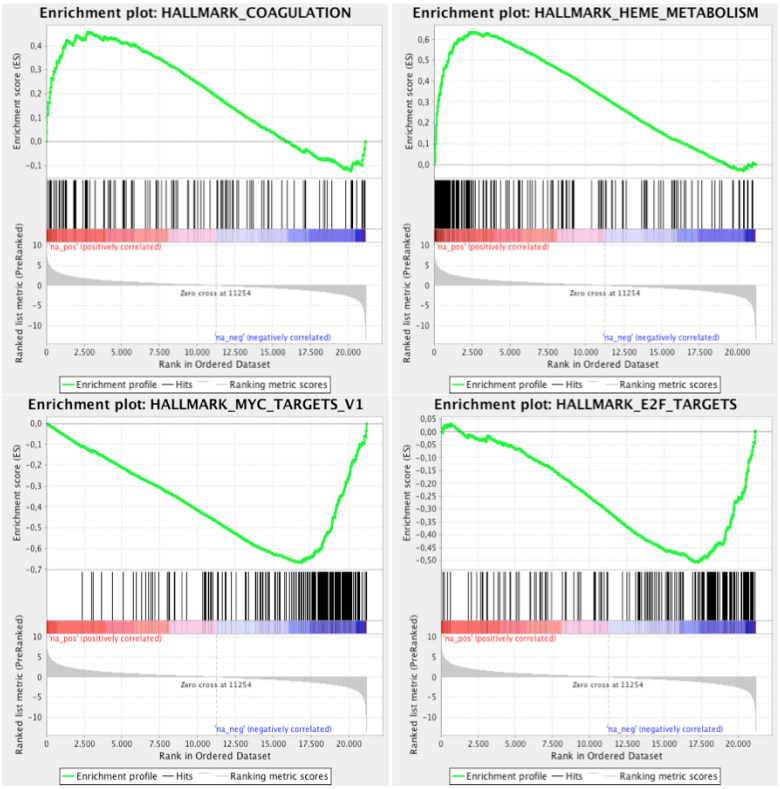


Figure R.7. PCA analysis of 12 RNA-seq samples of control and depleted PHF19 in K562. Each dot in the plot represents the expression of all genes in one RNA-seq experiment. According the PCA, replicates of the same condition were clustered in groups. For instance, replicates of sh#1 are depicted in blue while sh#2 samples are colored in red and the shCT in black.

Since the conditions sh#1 and sh#2 determined the clustering in **Figure R.7**, we ranked the transcriptome according to the ratio between each knock-down (KD) and the corresponding control. For each shRNA, we performed a Gene Set Enrichment Analysis (GSEA). In **Figure R.8** we reported several of the categories identified using the GSEA analysis. Notably, we observed that genes with enriched expression in the PHF19 depletion condition where related with hemoglobin metabolism and coagulation. It is worth to mention the visual phenotype we reported in **Figure R.5**, where it can be clearly

observed that cells upon knock-down acquired a reddish coloring. On the other hand, genes that were more expressed in the control condition, were related with targets of the well-known proto-oncogene MYC and cell cycle transcription factor E2F. Regarding this last category, genes under regulation of E2F transcription factor, are more expressed in control condition than in KD. Since E2F is a cell cycle regulator, the fact that those genes regulated by E2F are less expressed in KD condition is in agreement with the cell cycle arrest and the reduction of proliferation that is observed in K562 cell line in the previous chapter (**Figure R.3 and R.6**).

sh#1



sh#2

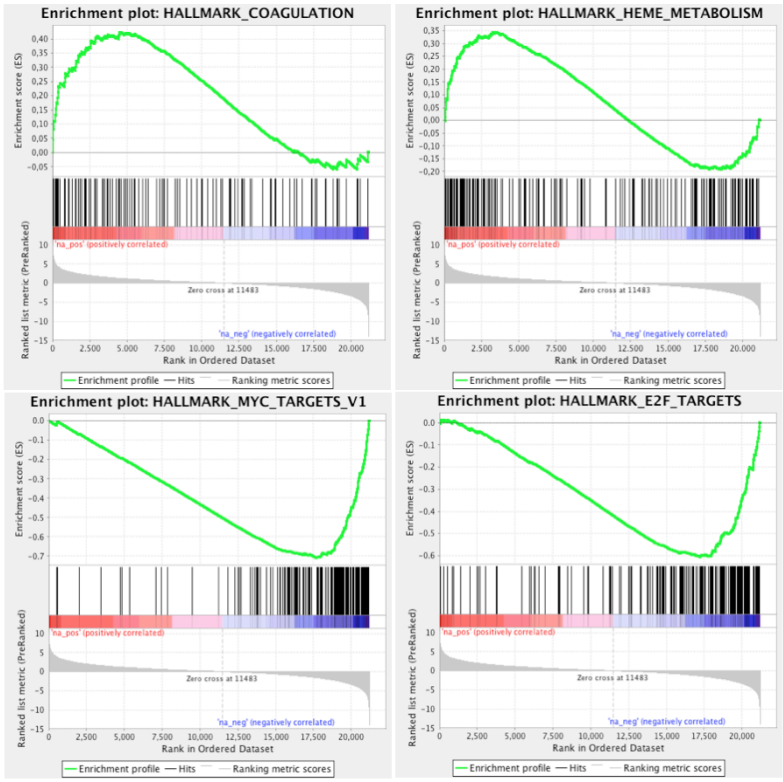


Figure R.8. GSEA analysis of ranked transcriptome of each knock down in comparison with the control samples. Each condition (sh#1 and sh#2) is obtained from 4 replicates. Both conditions were interrogated against the Hallmark ontology with the GSEA software. In all cases p-value was lower than 0,001.

To study more in detail the results observed in **Figure R.8**, we performed a differential gene expression analysis of sh#1 and sh#2 against shCT RNA-seq data (adjusted p-value <0,1 and fold change 1,2). We observed that there were more genes up regulated (619 for sh#1 and 2112 for sh#2) than down regulated (398 for sh#1 and 526 for sh#2). Moreover, when overlapping differential express genes of each comparison, we identified 281 genes up-regulated, and 35 down-regulated. Interestingly, ontology analysis done with

EnrichR platform [128,129] from up-regulated genes showed that depletion of PHF19 up regulates genes related with the erythroid differentiation, this was in line with previous results done with GSEA analysis **Figure R.9**. Furthermore, gene ontology (GO) analyses indicated that up-regulated genes scored as genes involved in categories, such as Daunorubicin metabolism involved in erythroid differentiation metabolism. Genes related with early erythroid differentiation (CD71+ cells) were also highlighted. In this category we find up-regulated genes relevant to erythroid differentiation such as GYPA or KEL. Moreover, genes under regulation of transcription factors GATA2 and GATA 1 where up regulated as well. Over all, RNA-seq data suggests that depletion of PHF19 could be involved in the differentiation of K562 cells towards the erythroid linage.

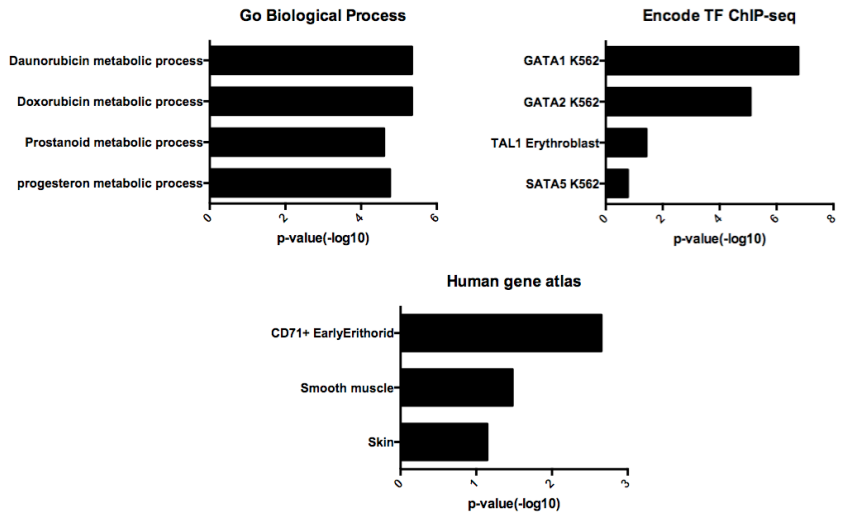


Figure R.9. Gene Ontology analysis of genes up-regulated following PHF19 knockdown. Ontology from the Up-regulated genes was interrogated with EnrichR platform. p-values obtained from EnrichR [129] software are plotted as –log10.

In order to validate these initial results obtained by RNA sequencing, we performed quantitative PCR of genes related with erythroid differentiation and cell cycle. As predicted, genes related with cell cycle such as CDKN1A/p21 and Cyclin D1 were found to be miss-regulated. On the other hand, key genes for the erythroid differentiation pathway are up regulated, such as GYPA, KEL, TAL1 and γ-Globin **Figure R.10**.

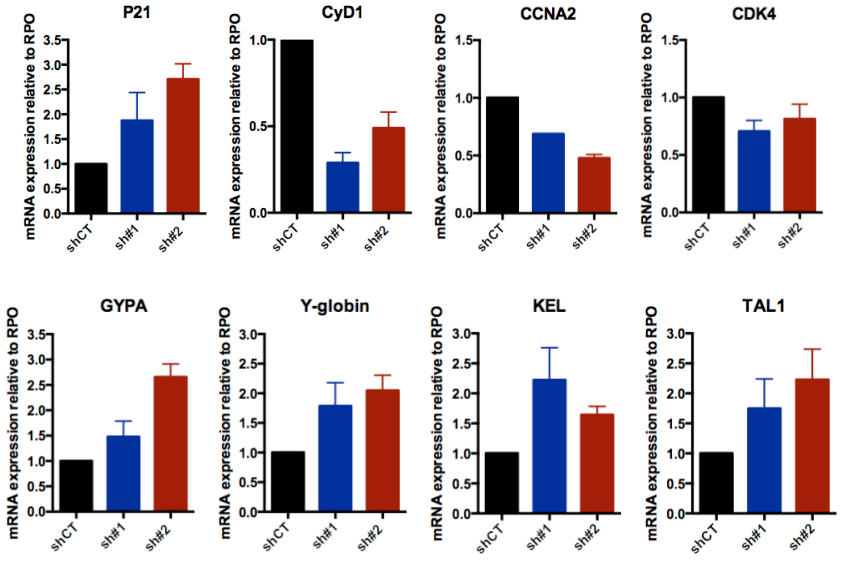


Figure R.10. PHF19 depletion deregulates cell cycle genes as well as erythroid differentiation genes. RT-PCR of key genes deregulated in cell cycle and erythroid differentiation pathways. Results are shown relative to shCT and normalized to the housekeeping gene RP0. Errors bars are the standard deviation (SD) of four independent experiments.

To further validate the erythroid differentiation pathway that K562 cells acquired upon deletion of PHF19, and to phenotypically corroborate that cells underwent differentiation, we performed FACS analysis for the erythroid specific cell surface marker CD235a [130] (**Figure R.11**). As a positive control we induced K562 cells with Ara-C, a well-known compound that induces differentiation towards the erythroid linage in this cell line [131,132]. These results corroborated our previous observations, indicating that PHF19 depletion causes a differentiation towards the erythroid linage.

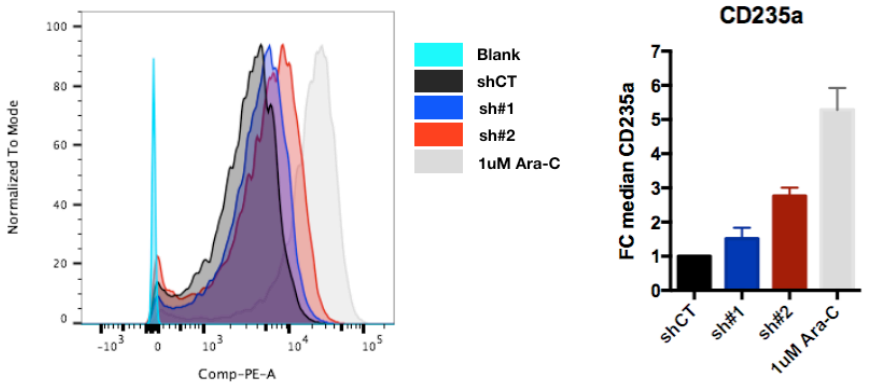


Figure R.11. PHF19 depletion increases CD235a erythroid cell surface marker. Left panel shows a representative plot of the gain of signal when PHF19 is depleted. Right panel shows the average of the median CD235a intensity staining normalized to the control after 72h of selection. Positive control is induced by 1uM of Ara-C. Error bars correspond to the standard deviation (SD) of three independent experiments.

K562 cell line can also differentiate into the megakaryocyte lineage [133]. In order to investigate whether PHF19 depletion causes a general differentiation process, or whether it is

specific for erythroid lineage, we differentiated K562 cells into the megakaryocyte lineage, by treating with phorbol myristate acetate (PMA) [132]. After treating K562 with PMA for 24h we wanted to evaluate the levels of CD61, a cell surface marker specific for megakaryocyte lineage induced under PMA treatment [130]. Results presented in **Figure R.12** point out that cells depleted for PHF19 reached lower level of megakaryocyte marker CD61 than the control did when treating with PMA. These results suggested that PHF19 depletion not only differentiate K562 towards the erythroid pathway, but also prevents the differentiation towards megakaryocyte lineage.

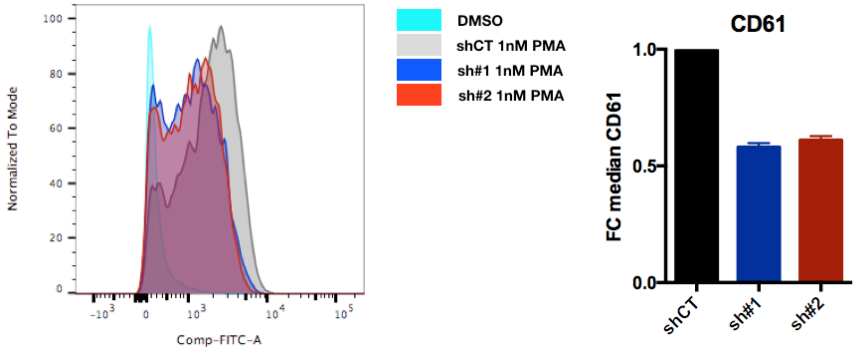


Figure R.12. PHF19 depletion prevents megakaryocyte differentiation induced by PMA. Left panel shows a representative plot of the reduction level of CD61 staining when PHF19 is depleted after induction of PMA. Right panel shows the average of the CD61 median intensity staining normalized to the shCT after 24h of 1nM PMA treatment. Error bars correspond to the standard deviation (SD) of three independent experiments.

To further investigate the potential therapeutic effect of depleting PHF19 in K562, we wondered whether lower doses of Ara-C combined with PHF19 depletion could reach the

levels of differentiation that 1uM of Ara-C does. We then treated the cells with 10nM concentration of Ara-C during 72h, and then by FACS we interrogated the CD235a intensity staining level. The results presented in **Figure R.13** indicate a cooperative effect when we treated PHF19 depleted K562 cells with 100 times less of Ara-C than our control of 1uM of concentration.

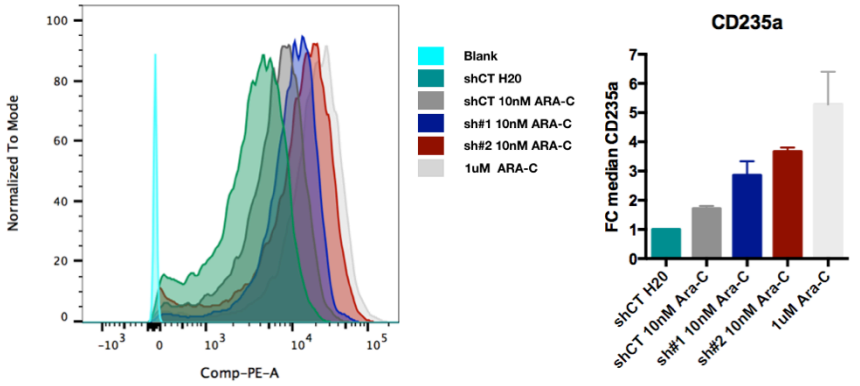


Figure R.13. Cooperative effect of lower doses of Ara-C combined with depletion of PHF19. Right panel shows a representative plot of the gain of signal when PHF19 is depleted and 10nM of Ara-C is added. Left panel shows average of the median intensity of staining of CD235a normalized to the control situation after 72h of Ara-C treatment and PHF19 depletion. Positive control is induced by 1uM of Ara-C. Error bars correspond to the standard deviation (SD) of three independent experiments.

3. PHF19 regulates p21 expression

As an interactor of PRC2, PHF19 is a chromatin associated factor. We thus reasoned that chromatin localization of PHF19 could help us to better understand the phenotype previously observed (i.e. decreased proliferation, erythrocyte differentiation). We took advantage of the chromatin inmunnoprecipitation followed by deep-sequencing (ChIP-seq) technique to identify PHF19 target genes. Since attempts to inmunoprecipitate PHF19 using protocols established in ESC cells did not work for K562 cell line, we used a commercial kit (ChIP-IT high sensitivity kit, Active Motif), which is optimized to give superior results when performing ChIP against low abundant proteins or antibodies with low binding affinities. We performed ChIP-seq adding 1% of *Drosophila* chromatin from the total IP to use it as spike-in in order to reduce the technical variation and sample processing bias (see methods), for both knockdowns and the control. Sequencing results with number of reads and genes are summarized in **Table R.1.**

	shCT	sh#1	sh#2
Mapped Reads (H/F)	34,164962/ 1,643854	36,526085/ 3,376993	37,171735/ 2,117170
Peaks	5621		
Genes	888		

Table R.1. Summary of mapped reads, peaks, and target genes identified for PHF19 using ChIP-seq. First row shows mapped Human (H) and Drosophila (F) reads. Second row shows the number of peaks observed after peak calling. Third row shows the number of genes where PHF19 is localized after association of peaks to genes.

After the sequencing, we processed the reads with peak calling program MACS1 against the input. We then processed the reads with the same peak calling but this time against our condition sh#2, which had the lowest levels of PHF19 when analyzing the mRNA levels. Moreover, since we sequenced the knock-down samples, we considered that “true” peaks could be identified as those were PHF19 is absent upon knock-down, rather than using the input, which presented higher background. To refine our peak identification, we did a manual curating selection of peaks, and we selected those peaks that had a TAG value (average number of reads per peak) higher than 100. We ended up with 5621 peaks that according to the gene association algorithm followed (see material and methods section), from -2.5 Kb upstream of the transcription start site (TSS) until the transcription ending site (TES), corresponded to 888 target genes. PHF19

preferentially localized at the transcription start site (TSS) of genes, including proximal, distal regions and 5'UTR (**Figure R.14**).

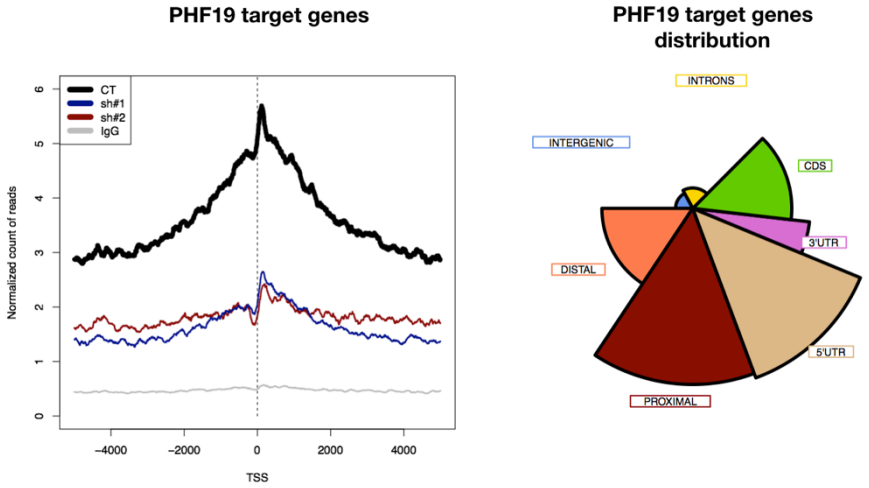
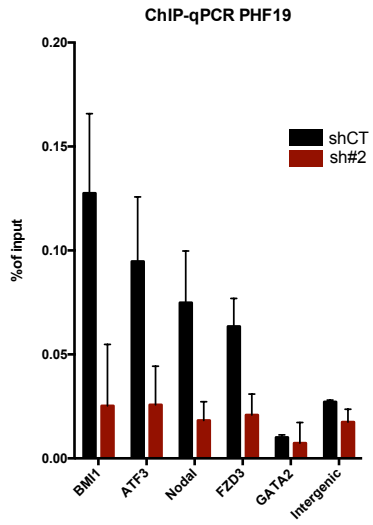


Figure R.14. Genome-wide distribution of PHF19 target genes. Left panel shows a graphical distribution of genes signal, normalized by the number of reads, 5 kb upstream and downstream of TSS for PHF19 target genes. Right panel shows the peak distribution of PHF19 across the genome.

Once we obtained the final list of PHF19 target genes, we wanted to confirm the specificity of the peaks by ChIP-qPCR experiments. We thus prepared chromatin from independent experiments, and inmunoprecipitated PHF19. We also included *Drosophila* chromatin to avoid sample processing bias. We validated several PHF19 targets that had a strong enrichment when visual inspection was performed. Among the 888 genes, we choose for validation BMI1, ATF3, NODAL, FZD3. We also selected GATA2 gene as a negative control since this gene was not present in the final list of PHF19 targets. We also used an intergenic region as a negative control (**Figure R.15**).


ChIP-qPCR PHF19
0.20
0.15
0.10
0.05
0.00
%of input
shCT
sh#2
BMI1
ATF3
Nodal
FZD3
GATA2
Intergenic

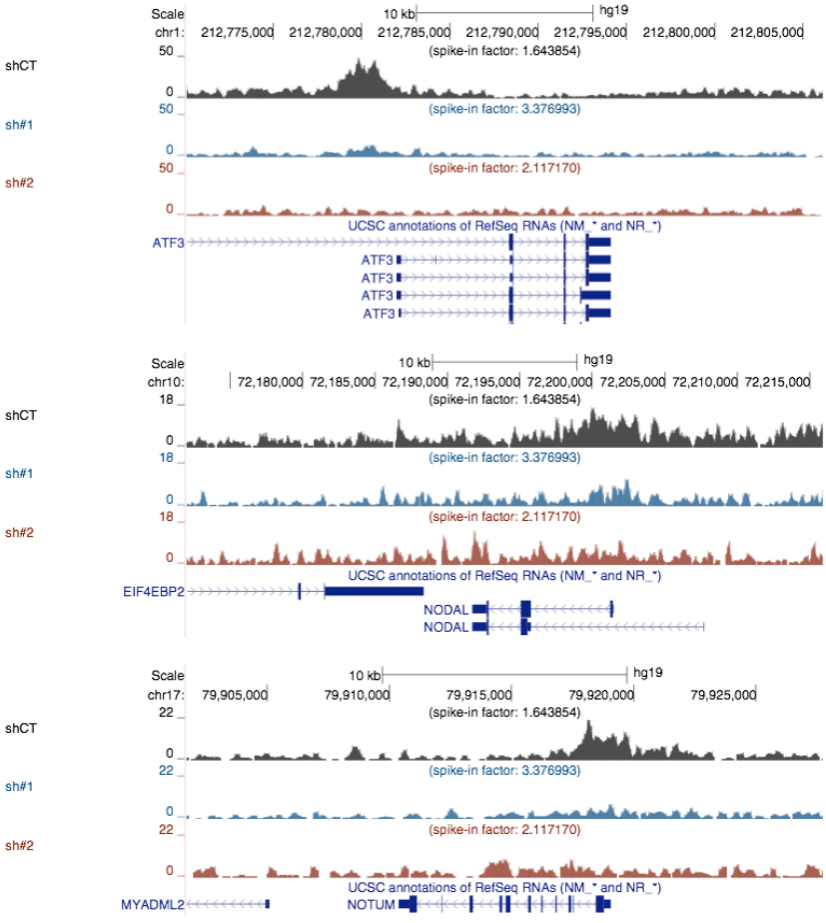

Scale
10 kb
hg19
chr1:
212,775,000
212,780,000
212,785,000
212,790,000
212,795,000
212,800,000
212,805,000
shCT
sh#1
sh#2
(spike-in factor: 1.643854)
(spike-in factor: 3.376993)
(spike-in factor: 2.117170)
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
ATF3
chr10:
72,180,000
72,185,000
72,190,000
72,195,000
72,200,000
72,205,000
72,210,000
72,215,000
EIF4EBP2
NODAL
chr17:
79,905,000
79,910,000
79,915,000
79,920,000
79,925,000
MYADML2
NOTUM

Figure R.15. Validation of PHF19 peaks. The panel above shows validation of several PHF19 target genes. Average values of the % of input normalized to *Drosophila* gene *pbgs* of at least 2 independent experiments. Error bars show the standard deviation (SD). The panels below show some examples from the UCSC genome browser.

To further analyze the genome distribution of PHF19, but also its function(s), we tested whether PHF19 target genes were also PRC2 targets genes. We took advantage of ENCODE ChIP-seq available data in K562 for PRC2 core components and for H3K27me3 histone modification. As reported in **Figure R.16,** we observed 7822 common genes between EZH2, SUZ12 and H3K27me3. When we overlapped the common PRC2 target genes with the PHF19 targets genes we observed that 94% of PHF19 target genes (833/888) were also targets of PRC2 core components, and had the repressive mark too. This suggested us that PHF19 function is directly connected with the PRC2 function.

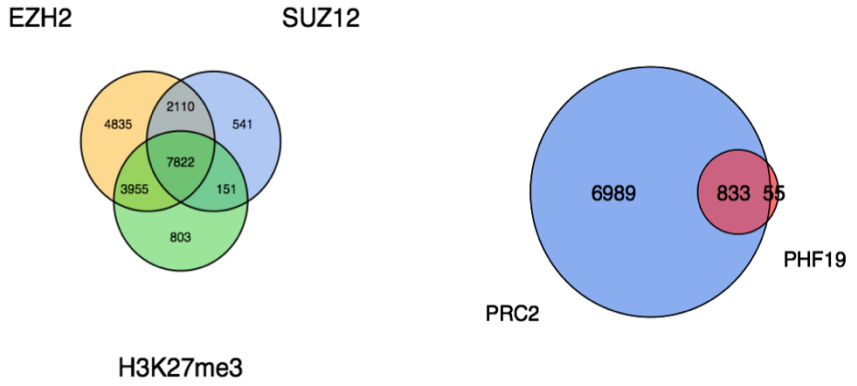


Figure R.16. Comparison of PHF19 targets and PRC2 targets. Left panel shows the intersection between EZH2, SUZ12 and H3K27me3 targets genes. Right panel shows a diagram with the intersection of previous targets in common from left panel, called PRC2, and PHF19 target genes.

Then, in order to better understand the functionality of PHF19 target genes, we interrogated with EnrichR platform the gene ontology of the 888 PHF19 target genes (**Figure R.17**).

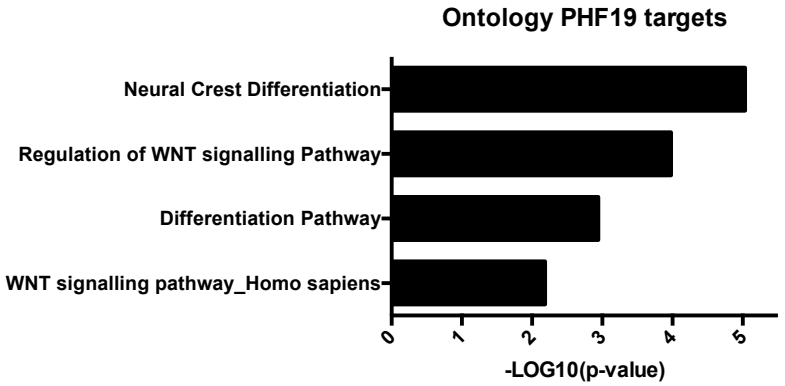


Figure R.17. Gene enrichment analysis of PHF19 target genes. PHF19 targets were enriched in differentiation pathways and implicated in the WNT signaling pathway or in its regulation. p-values obtained from EnrichR[129] software were plotted as –log10. Categories are obtained from Wikipathways, GO Biological process and Panther resources.

Since most of PHF19 target genes were also PRC2 targets, as expected we obtained categories related with differentiation processes. Some of the PHF19 target genes belonging to these categories were: SHH, PDGFB, PDGFA, NODAL or FGFR1.

Moreover, PHF19 target genes are also involve in the regulation of the WNT signaling pathway, which has been recently demonstrated to be important during the final steps of erythropoiesis [60]. Some of these genes were: FZD1, FZD3, CTNNB1, WNT2B or NOTUM.

Altogether, gene enrichment analysis indicated that PHF19 together with PRC2 regulated several of the key genes related with the differentiation of K562 cell line.

In order to further characterize the PHF19 target genes, we investigated whether their expression was affected by PHF19 depletion. We took advantage of GSEA, using PHF19 target genes as a geneset to evaluate changes in expression. As reported in **Figure R.18**, PHF19 target genes are mainly up-regulated upon knock-down, according to a classical repressive role of PRC2.

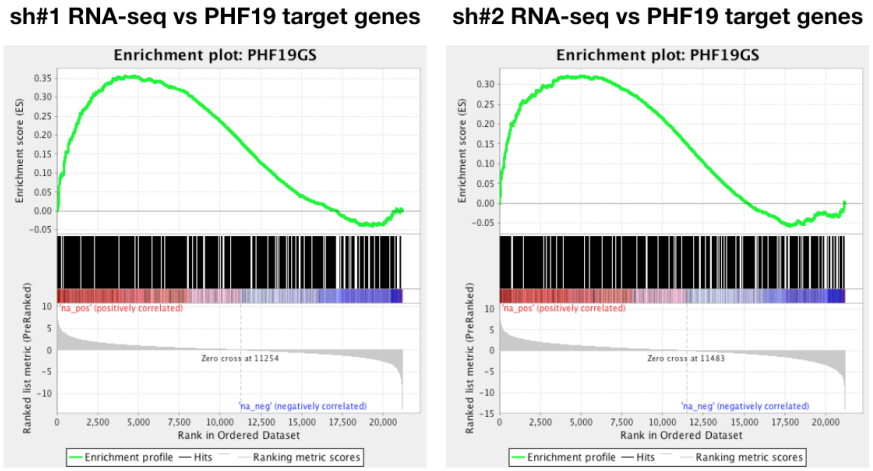


Figure R.18. GSEA of the RNA-seq data against PHF19 target genes. Ranked RNA-seq genes from both knockdowns where interrogated against the subset of PHF19 target genes. In both cases p-value was lower than 0,001. NES sh#1: 1,688. NES sh#2: 1,589.

Notably, from the top25 target genes of PHF19 that have higher expression upon PHF19 depletion, we found CDKN1A/p21 gene. Since p21 is one of the major cell cycle regulators, we validated its expression levels by qPCR, as

well as PHF19 occupancy by ChIP-qPCR, upon PHF19 knock down (**Figure R.19**).

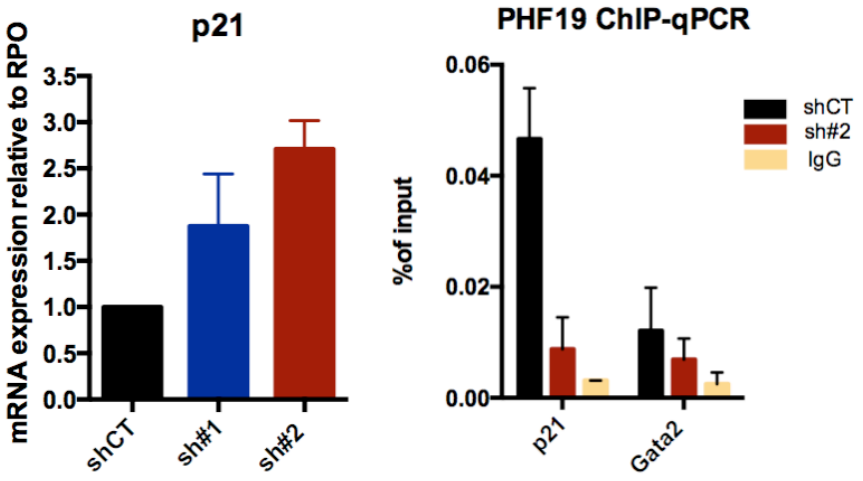


Figure R.19. Validation of p21 as PH19 target. Left panel shows the mRNA expression levels of p21 after PHF19 depletion relative to RPO gene. Error bars are the standard deviation (SD) of 5 independent experiments. Right panel shows average values of the % of input, normalized to the *Drosophila* gene *pgsb*, of PHF19 ChIP-qPCR for p21 genes and GATA2 as a negative control gene. Error bars are the standard deviation (SD) of two independent experiments.

Moreover, although p21 is a direct target of PHF19 and its depletion increases p21 expression, we wondered if its increased expression correlates with an increase at the protein level. As reported in **Figure R.20** we validated that p21 increased at the protein level by western blot in PHF19 depleted K562 cells compared with shCT.

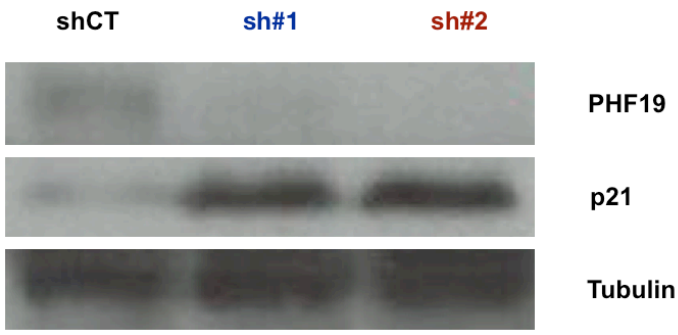


Figure R.20. Increased p21 protein levels after PHF19 depletion. Western blot of K562 cell line, in control and PHF19 depleted cells. Tubulin was used as a loading control.

We next wondered whether the overexpression of p21 was the direct trigger for the differentiation of K562 cells upon depletion of PHF19 (see results reported in Section2). To test that, we overexpressed in K562 p21 using a pBABE plasmid with the aim of assessing whether this was sufficient to trigger erythroid differentiation. As reported in **Figure R.21**, we observed similar overexpression level of p21 than those observed when PHF19 was depleted. Moreover, some key genes related to cell cycle where also miss-regulated. Then we checked the BrdU incorporation and the cell death. Despite the similar expression levels of p21, the BrdU incorporation was completely different between p21 overexpressed condition and PHF19 depletion. There was almost no BrdU incorporation in p21 overexpressed condition, in contrast with the BrdU incorporation obtained with PHF19 knock-down cells. Moreover, p21 overexpression led to a massive cell death in contrast of PHF19 depletion effect, as

measuted by Anexin V staining, a specific marker of cell death.

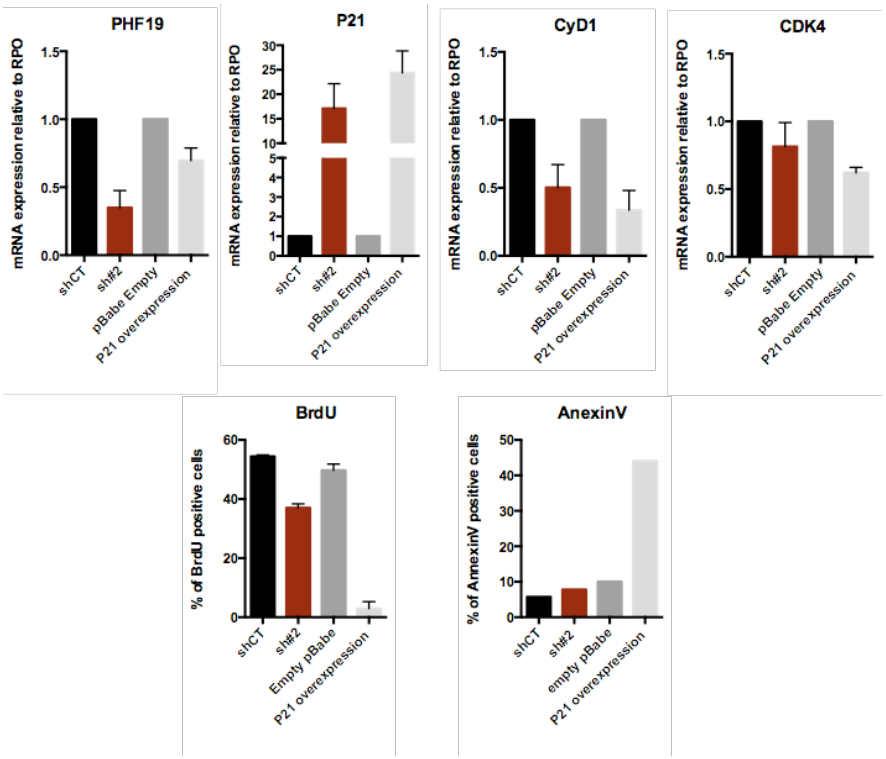


Figure R.21. Decreased BrdU incorporation and increased cell death after p21 overexpression. Top panels show mRNA expression levels of PHF19, p21, Cyclin D1, and CDK4, error bars are the standard deviation (SD) of two independent experiments. Lower left panel, BrdU incorporation of PHF19 depleted cells and P21 overexpressed cells, error bars show the standard deviation (SD) of two independent experiments. Lower right panel shows a cell percentage of Anexin V incorporation of an individual experiment.

In this section we have reported the genomic localization of PHF19, demonstrated how PHF19 directly regulates p21 expression, and how this potentially leads to the impairment of the cell cycle previously described in the Section 1. It is worth to notice that after the analysis of some human data available from patients with chronic myeloid leukemia and acute myeloid leukemia, we have found a clear anti-

correlation between the expression of PHF19 and the expression of P21 in patients with CML independently on the stage of the disease [134] (**Figure R.22**).

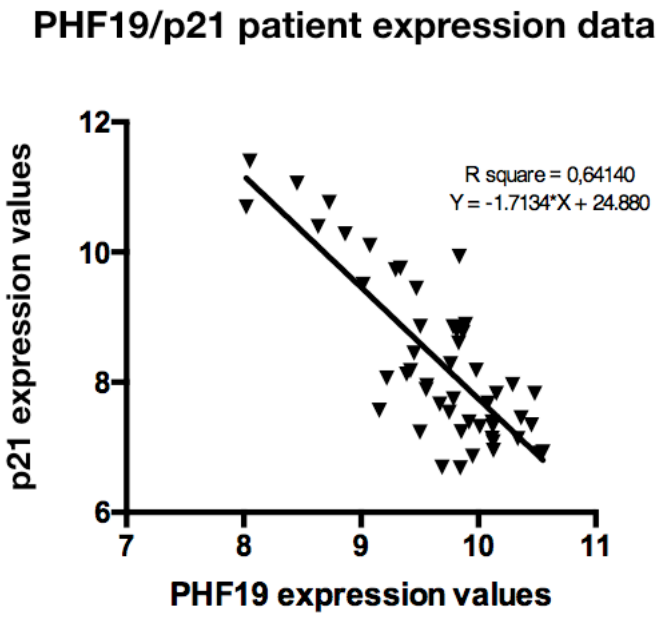


Figure R.22. Correlation of PHF19 and p21 expression data in CML patients. Microarray expression values of CML patients from *Abraham S.A et al* 2016 [134] (GSE47927).

Moreover, we have also shown that the p21 overexpression induces a cell cycle arrest and cell death. This axis of regulation between PRC2, PHF19, and p21 locus, is highlighting the role of Polycomb in regulation the cell cycle in K562 cells. We then hypothesized that the overexpression of p21 after depletion of PHF19 could lead to the erythroid differentiation. Since we obtained massive cell death upon overexpression of p21 in K562 cell line, we wondered whether an additional mechanism (and thus not only p21

overexpression) was triggered by PHF19 depletion, which leaded to erythroid differentiation.

4. PHF19 depletion promotes K562 erythroid differentiation through the regulation of WNT signaling pathway *via* MTF2

We have shown in Section 3 that PHF19 regulates p21 locus, which may account for the cell cycle arrest in K562 upon PHF19 depletion. However, it is still unclear how this depletion of PHF19 also leaded to erythroid differentiation, as mentioned in Section 2. Recently, *Rothberg JLM et al 2018* [60] have reported that in the mouse hematopoietic system, MTF2 is necessary for terminal differentiation of the erythroid lineage. Moreover, they showed that MTF2 regulates this process through the inhibition of the WNT signaling pathway. Therefore, we analyzed the expression levels of MTF2 after depletion of PHF19 (**Figure R.23**). We could observe that the RNA levels of MTF2 were not increased when PHF19 was depleted. Of note, the PHF19 homolog PHF1 is not expressed in this cell line based on our RNA-seq data.

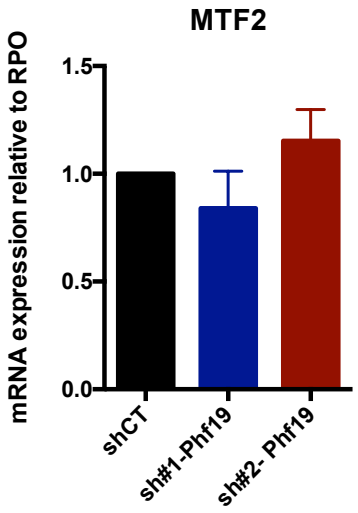


Figure R.23. MTF2 expression after PHF19 knoc-kdown. RT-PCR analysis of control and PHF19 depleted K562 cell line. Expression values were normalized by the expression of housekeeping gene RPO and are depicted relative to shCT. Errors bars are the standard deviation (SD) of four independent experiments.

Since MTF2 had not a significant increase at expression level, we wondered if its genomic localization was different after depletion of PHF19. We performed a ChIP-seq analysis of MTF2, adding 1% of Drosophila chromatin as a spike-in control to compare between control and PHF19 depleted conditions (see Matherial and Methods section). Sequencing results with number of reads and genes are summarized in **Table R.2**.

	shCT	sh#2
Mapped Reads (H/F)	24,757,245/345,956	36,525,149/585,523
Peaks	12,502	18,287
Genes	3246	4423

Table R.2. Summary of mapped reads, peaks, and target genes identified for MTF2 using ChIP-seq. First row shows mapped Human (H) and Drosophila (F) reads. Second row shows the number of peaks measured after peak calling. Third row shows the number of genes where PHF19 is localized after associate peaks to genes. First column shows the numbers for the shCT samples and second column shows the numbers for sh#2 PHF19 samples.

We processed the reads with MACS1 program against the Input and IgG to obtain the peaks. Finally, since input samples presented higher background, we decided to use the peaks obtained running the peak caller against the IgG. Our first observation was that there were more peaks and genes in the PHF19 KD condition than in control condition. As expected, the majority of the genes in the control situation (3027) were present in knock-down condition (**Figure R.23**).

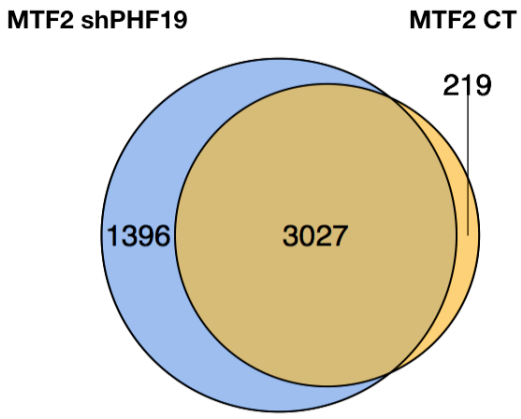


Figure R.24. Venn diagram of MTF2 target genes by ChIP-seq analysis in control and upon PHF19 knockdown. Venn diagram shows the intersection of MTF2 target genes in control and upon PHF19 KD situation. 3027 genes were in common.

Then we wondered whether the PHF19 target genes were also target genes of MTF2. In **Figure R.24** we overlapped the PHF19 target genes with the MTF2 target genes obtained when PHF19 was depleted. We observed that the majority (90.6%) of PHF19 targets genes were also bound by MTF2. Moreover, **Figure R.24** showed that there was a general gain of MTF2 signal in the PHF19 target genes when PHF19 is depleted.

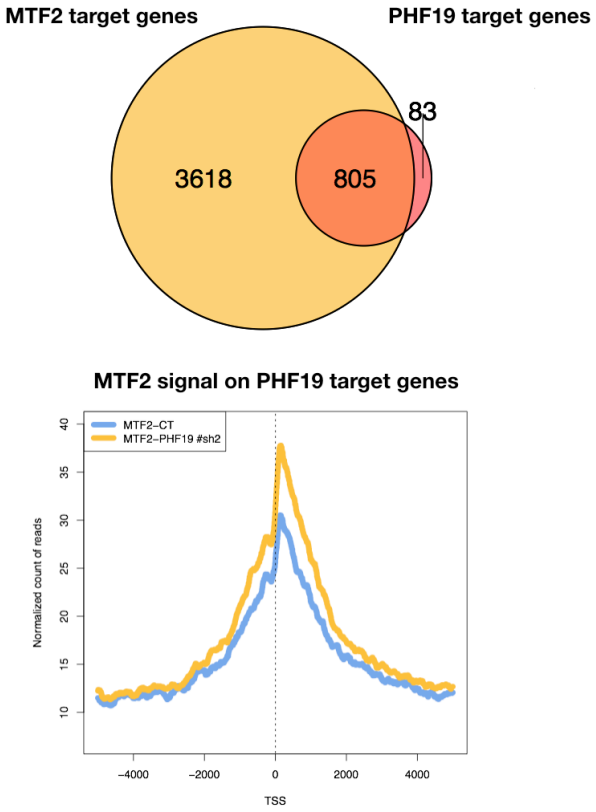


Figure R.25. Correlation of PHF19 and MTF2 targets genes. Upper panel shows the intersection between PHF19 target genes and MTF2 target genes (805 in common). Lower panel shows MTF2 signal in PHF19 target genes upon PHF19 depletion.

We visually inspected in the USCS genome browser the increase of MTF2 signal in those PHF19 target genes (**Figure R.26**).

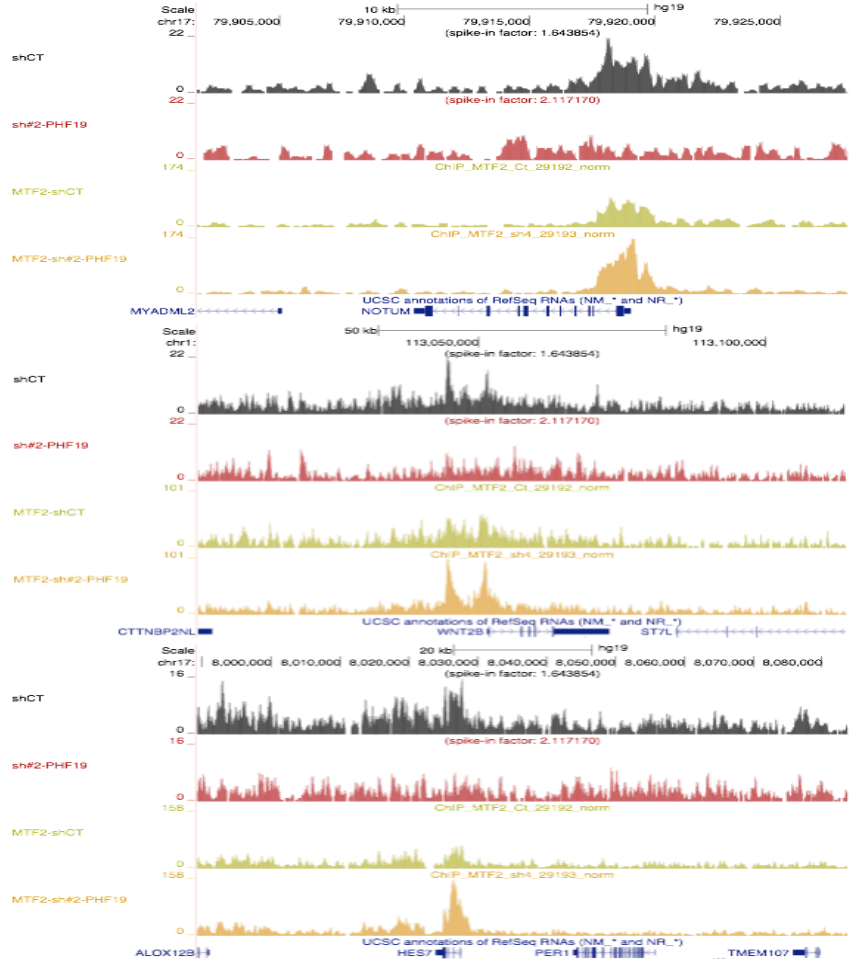


Figure R.26. Gain of MTF2 signal after PHF19 depletion. UCSC screenshots of PHF19 and MTF2 ChIP-seq profile in shCT or sh#2-PHF19 K562 cells. In black PHF19 ChIPseq of the shCT. In red PHF19 ChIPseq of sh#2-PHF19. In green MTF2 ChIPseq of shCT. In Yellow MTF2 ChIPseq of sh#2-PHF19. Numbers at the top left of each graph represent the number of reads for each ChIP-seq in the given gene.

Moreover, in order to confirm the specificity of MTF2 gain of signal, we validated by ChIP-qPCR several genes where the increase of MTF2 was observed. We prepared chromatin from a biological replicate experiment, also using *Drosophila* spike-in to avoid technical bias. We checked NOTUM, FZD3,

PDGFRA, GATA3, CXCR4 genes. We used GATA2 and an intergenic region as negative controls (**Figure R.27**)

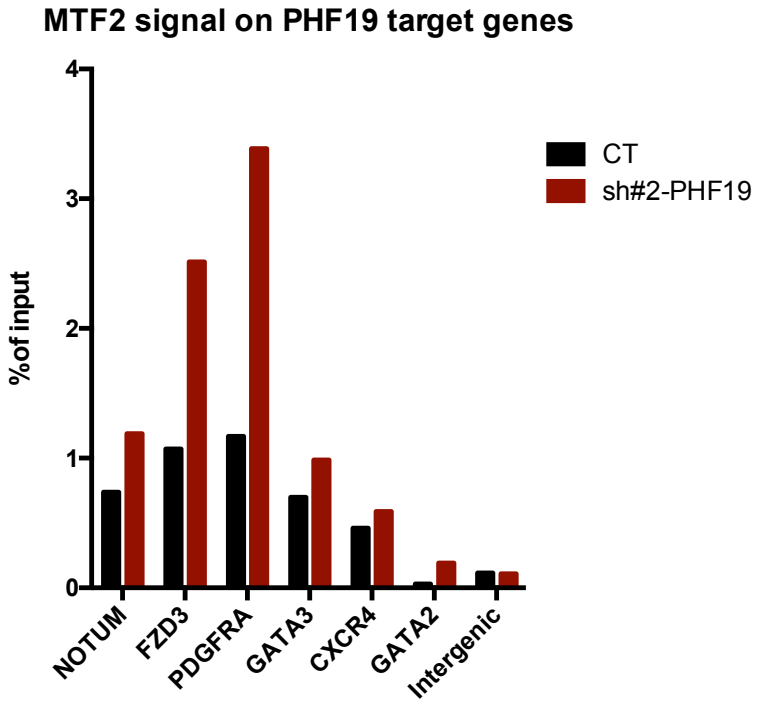


Figure R.27. Validation of MTF2 gain of signal on PHF19 target genes. ChIP-qPCR validation of several PHF19 target genes. Percentage of input normalized to *Drosophila* gene *pbgs* of an independent experiment.

We then wondered whether the increase of MTF2 signal could affect the levels of the histone mark H3K27me3. For this, we performed ChIP-sequencing of H3K27me3 in control and PHF19 knock-down situation. Indeed, ChIP-seq data reported in **Figure R.28** indicated that PHF19 target genes presented increased levels of H3K27me3.

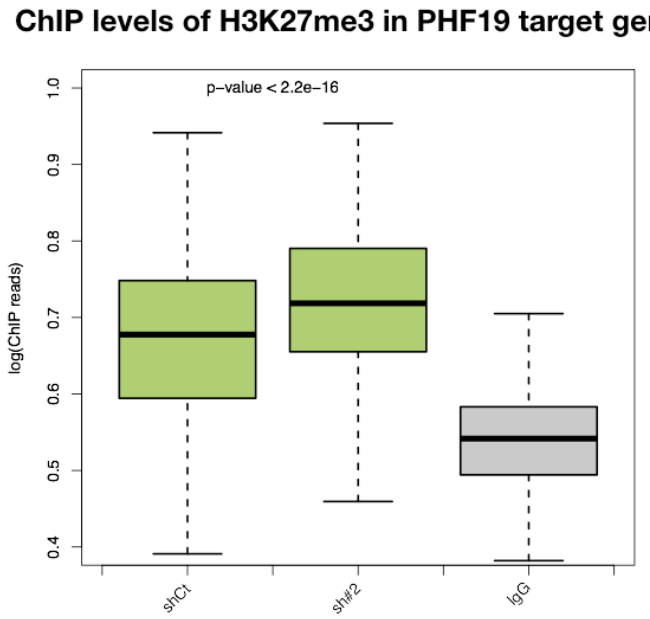


Figure R.28. H3K27me3 ChIP signal in PHF19 target genes. Boxplot showing ChIP signal of H3K27me3 was compared between control condition and PHF19 knockdown in the PHF19 target genes. P-value was obtained with a Wilcoxon non-parametric t-test.

We next performed gene ontology analysis of those genes that gained MTF2 signal after PHF19 depletion. There were 232 genes out of 888 target genes with increased MTF2 signal of at least 1,3 of fold change (FC). As reported in **Figure R.29**, we observed that the majority of genes were related to regulation of WNT signaling pathway. Some of the genes belonging to this category were: FZD1, FZD3, CTNNB1, WNT4, WNT2B or NOTUM.

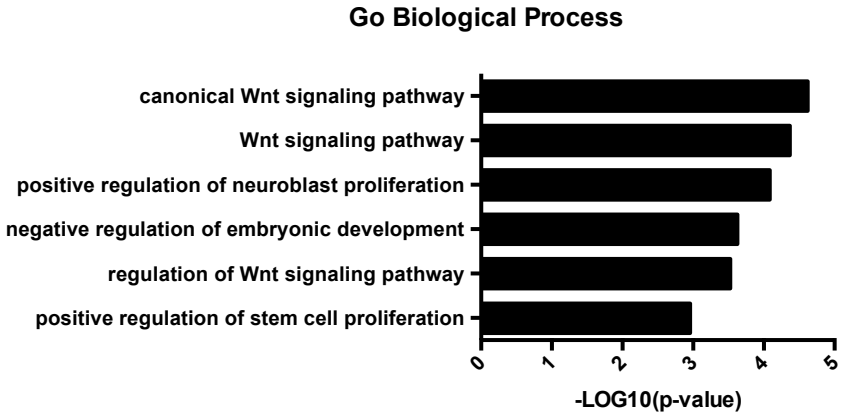


Figure R.29. Gene ontology analysis of PHF19 target genes gaining MTF2 signal. PHF19 target genes that gained MTF2 (of at least 1,3 FC) upon PHF19 depletion are enriched in WNT signaling pathway. P-values obtained from ENRICHR software are plotted as –log10. Categories are obtained from GO Biological process, using the tool Enrichr [129].

In sum, PHF19 target genes gained MTF2 and H3K27me3 signal upon PHF19 depletion. Although these results were in line with previous publications from *Rothberg JLM et al* 2018 [60] reporting that MTF2 repression of WNT targets is needed for complete erythropoiesis, they are in contrast with the general up-regulation of PHF19 targets observed in **Figure R.18** from Section 3. We next analyzed more in detail the expression levels of the PHF19 target genes. Although GSEA (**Figure R.18**) pointed towards to an up-regulation of PHF19 targets, most of these genes were very weakly expressed (RPKM<1) (**Figure R.29**).

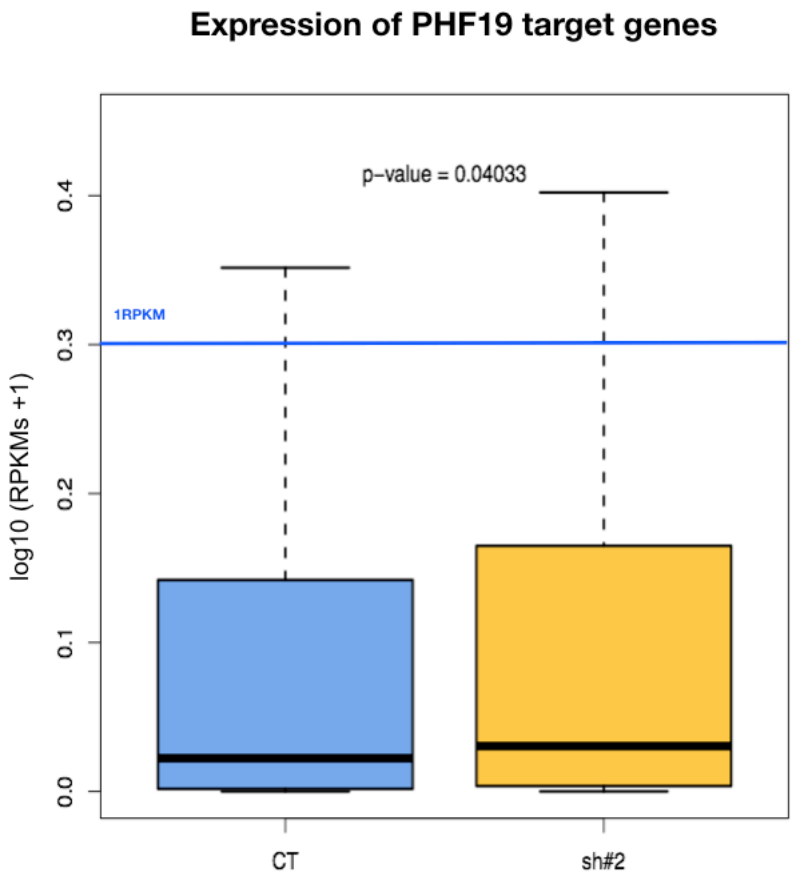


Figure R.30. Comparative expression values of PHF19 target genes. Boxplot showing the expression values of PHF19 target genes in control and PHF19 knockdown situation from RNA-seq data. P-value was obtained using a Wilcoxon non-parametric t-test.

Interestingly, not all the genes accumulated MTF2 after PHF19 depletion. For instance, CDKN1A that we previously showed as significatively up-regulated (**Figure R.19**) did not reveal any significant change in MTF2 signal upon depletion of PHF19 (**Figure R.31**).

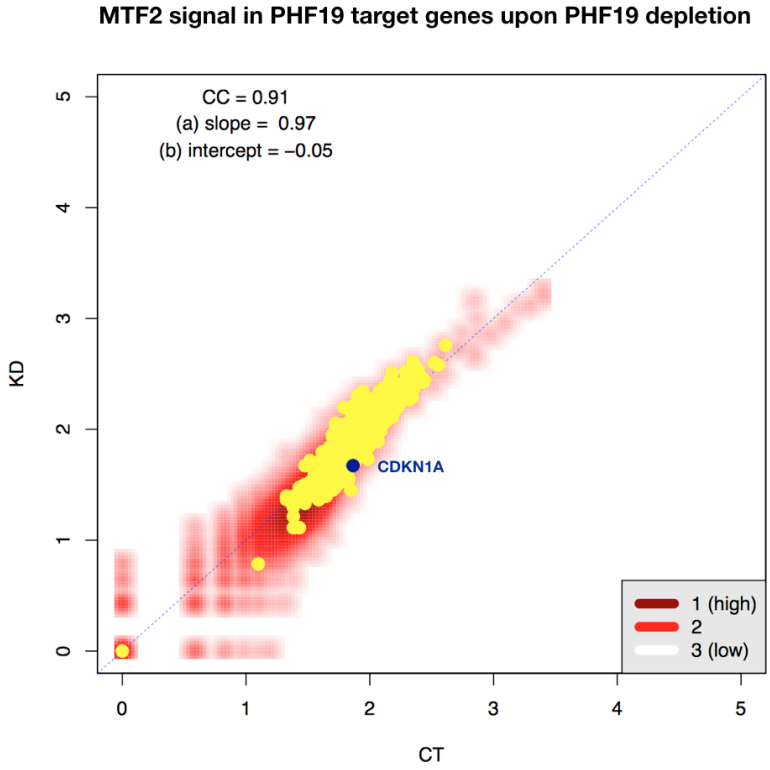


Figure R.31. Scatter plot distribution of ChIPseq MTF2 CT and ChIPseq MTF2 sh#2-PHF19. Each point indicates the maximum value for the normalized number of reads of the MTF2 ChIPseq in the shCT situation, and the MTF2 ChIPseq in sh#2-PHF19 situation, within the region -2.5 upstream of the TSS until the end of the gene. In red, we represent this value of the full set of genes annotated in hg19 by RefSeq. In yellow we highlighted the PHF19 target genes in the same plot. CDKN1A is highlighted in blue.

To obtain a further insights for the changes occurring upon PHF19 depletion, we selected the top 50 genes gaining MTF2 signal upon depletion of PHF19 (TOP50), as well as the bottom 50 genes, those that lost MTF2 signal upon depletion of PHF19 (BOTTOM50) (**Figure R.32**).

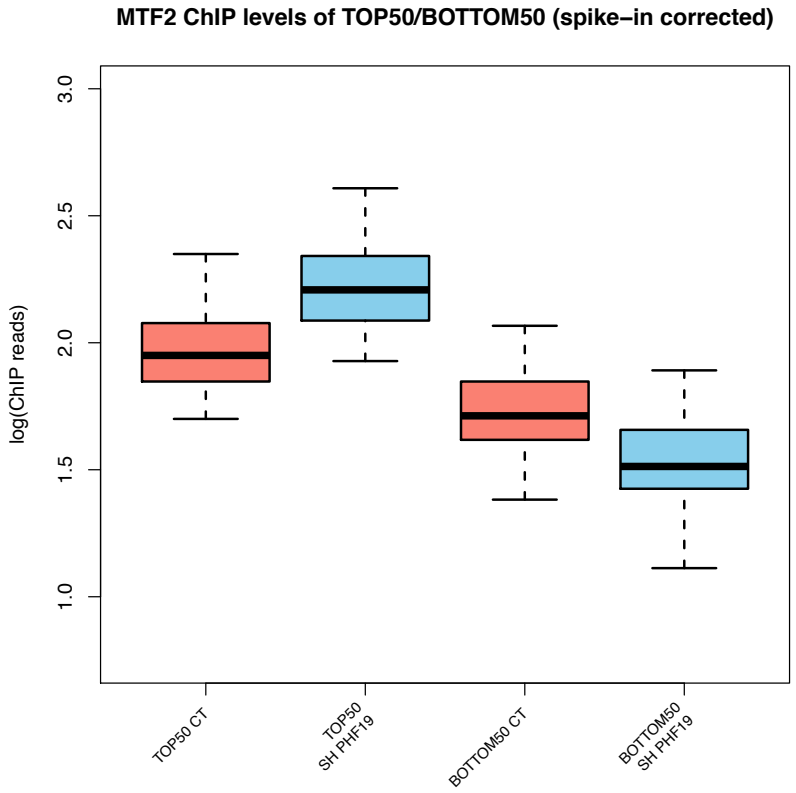


Figure R.32. MTF2 ChIP levels of TOP50 and BOTTOM50 genes. Boxplot showing MTF2 ChIP levels in the subset of TOP50 genes gaining MTF2 after PHF19 depletion, and the BOTTOM50 genes gaining MTF2 after PHF19 depletion.

We next analyzed the expression values of these two subsets of genes generated based on the MTF2 increase or decrease of signal after depletion of PHF19. As reported in **Figure R.33,** genes that gained more MTF2 were those genes less expressed initially and which expression was actually not altered after depletion of PHF19. On the other hand, when we focused on those genes that strongly lost MTF2 after PHF19 depletion, we observed that these genes were initially more expressed (above the 1RPKM threshold), and also were

having a tendency to be more expressed when PHF19 was depleted.

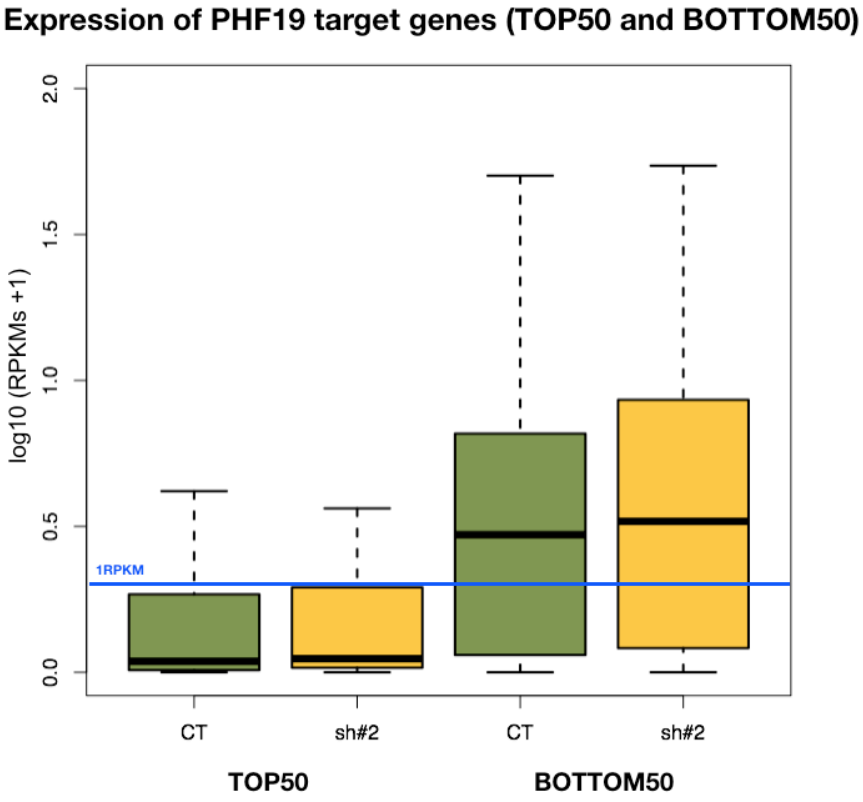


Figure R.33. Comparative expression values of TOP50 and BOTTOM50. Boxplot showing the expression values of the TOP 50 PHF19 target genes gaining MTF2 signal after PHF19 depletion, and BOTTOM 50 target genes loosing MTF2 signal after PHF19 depletion.

More interestingly, we analyzed in detail the PHF19 levels of these TOP50 and BOTTOM50 genes in the basal situation: In **Figure R.34** we observe that TOP50 genes gaining MTF2 had more PHF19 than BOTTOM50 genes.

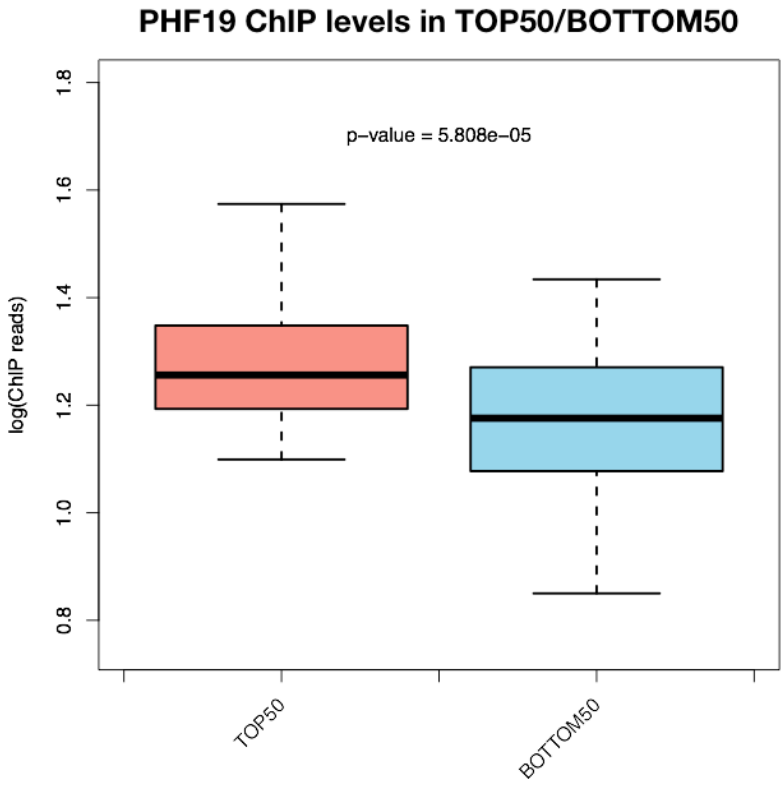


Figure R.34. PHF19 ChIP levels of TOP50 and BOTTOM50. Boxplot showing PHF19 basal ChIP levels of the TOP50 and BOTTOM50 genes gaining MTF2 after PHF19 depletion. P-value was obtained doing Wilcoxon non-parametric t-test.

Of note, the BOTTOM50 subset of genes that loose MTF2 after PHF19 depletion, initially also had lower levels of PHF19 signal. One of the genes observed in this group was p21 that as we have reported before, was up-regulated after PHF19 depletion.

Thus, in this section we have observed that PHF19 depletion in K562 cell line do not increase the expression of its homolog MTF2, but it has an impact of on MTF2 redistribution

at PHF19 target genes. Interestingly, MTF2 tends to increase in those target genes that had a very low level of expression, and which are precisely related with erythropoiesis. On the other hand, in higher expressed PHF19 target genes, MTF2 increased occupancy was not detected, and gene derepression was observed.

DISCUSSION

1. PHF19, homeostasis and differentiation

PHF19 is an associated PRC2 protein, which has been mainly studied in the context of mESCs. It has been described that PHF19 interacts with PRC2 and helps to the proper deposition of the repressive mark H3K27me3 [63,64]. Moreover, PHF19 depletion caused spontaneous differentiation in mESCs. In contrast, depletion of PRC2 core members do not trigger this spontaneous differentiation, but their presence is important for a proper differentiation [135,136]. When PHF19 is analyzed in the context of adult homeostasis, we and others have observed that the expression of PHF19 is mainly localized in adult stem cells, and decreases in more committed progenitors or differentiated cells [110,137,138]. This downregulation of PHF19 is also observed in the hematopoietic system, where the hierarchy of differentiation is well characterized (**Figure I.6**) [110].

There are several hypotheses on the origin of the cancer stem cells and tumor formation, but the main consensus is that cancer cells resemble in some features, such as self-renewal, or the capacity to give raises specialized cells, to normal stem cell [139,140]. In this regard, one of the main characteristics of the cancer cells is the mis-regulation of the cell cycle, which consequently leads to acquisition of high proliferative status. Furthermore, this capacity of growing in a non-regulated manner, also leads to a decreased capacity of differentiation [111]. In this sense, some of the therapies for

treating cancer patients are focused on elimination of these over-proliferating cells, but also forcing the cancer cells to differentiate. Of note, this therapeutic approach is very relevant for leukemia treatment [141].

Apart from anti-proliferative and differentiating strategies, another level of research in the field of leukemia has been focused on those cells that are escapers (*i.e.* resistant to the conventional treatments). It has been proposed that those cells that lead the relapse in leukemia are the ones that apart from being resistant to classical treatment, accumulate multiple genetic mutations. These acquired mutations provide to cancer cells with new genetic characteristics that should be considered for the following rounds of treatment [142-144].

Epigenetic factors have an impact on the chromatin and therefore on gene expression. With the aim of tackling cancer cells, drugs targeting epigenetic factors are nowadays hotspot in research. Indeed, there are several drugs targeting PRC2 core component EZH2, which are already under study for treating not only leukemia, but also other types of cancer [145]. Related with that, in our lab we have observed that targeting SUZ12 could be a strategy for treating myeloid leukemia dependent on PML-RARa oncoprotein, in combination with the retinoic acid treatment [104].

All these results leaded us to hypothesize that PHF19 could be also used as a target in myeloid malignancies. Moreover, since we observed that PHF19 is overexpressed in myeloid leukemia cell lines (**Figure R.1**). We thought that using these

cell lines would be a great system to understand the role of PHF19 in myeloid leukemia.

We took advantage of the acute and chronic myeloid leukemia cell lines HL60, NB4 and K562, and our first observation was that myeloid leukemia cell lines were growing less when PHF19 was depleted. Then we wanted to investigate that this decrease in cell growth was not due to an increase of cell death. This was done because sometimes infections with lentiviruses could lead to an off-target cell death, masking the real results. We observed that cells were not dying because of the depletion of PHF19, but they were indeed proliferating less. Moreover, to our surprise, we noticed that K562 cells pellets presented a reddish color when PHF19 was depleted, even after several PBS washes. Since one of the differentiation pathways that K562 cells could take is erythrocyte, we wanted to further investigate the possibility that K562 cells were not only arresting the cell cycle, but also differentiating after PHF19 depletion.

We further validated that PHF19 depletion was decreasing the cell growth rate by analyzing the cell cycle profile. Moreover RNA-seq data, and RT-PCR analysis validated this observation: we indeed demonstrated that key cell cycle genes were misregulated.

When analyzing the RNA-seq data we also noticed that K562 cells were differentiating towards the erythroid lineage. Indeed, we could corroborate that some genes related with erythroid differentiation were up-regulated. A part from gene

expression, we functionally validated the erythroid differentiation using cell surface markers (**Figure R.11**). Interestingly, the differentiation was specific, since forced megakaryocyte differentiation was partially impaired in PHF19 knock down cells (**Figure R.12**).

Our data support that PHF19 is indispensable for K562 aberrant growth as its depletion triggers cell cycle arrest and enhances differentiation. With these evidences regarding the role of PHF19 in K562, we also tried to force differentiation towards the erythroid lineage treating K562 cells with Ara-C while depleting PHF19, to explore the combinatorial therapeutic capacity of targeting PHF19. We could observe that there was not a synergistic effect but an additive effect (**Figure R.13**).

All the experiments were done in a cell line that, although resembles a myeloid leukemia, is an immortalized cell line with its obvious limitations. Further investigation is needed to corroborate that PHF19 inhibition could be used as a target for a leukemic treatment. In this sense, we tried to perform initial experiments with a xenograft model trying to validate *in vivo* what we observed *in vitro*. Our aim was to corroborate in an *in vivo* model that, after depletion of PHF19, K562 cells were not able to grow and/or invade adult organs. Although is not reported in this thesis, we used stable K562 cells lines with a luciferase reporter, and we injected them in an inmunodepresive C57BL/6 mouse. Due to manipulation

issues, we could not properly report the effect of PHF19 depletion in a xenograft model.

Worthy of note, our lab has generated a mouse PHF19 knockout to study the role of PHF19 in normal homeostasis (**see ANNEX**). In order to further investigate the role of PHF19 in leukemogenesis we could purify hematopoietic precursors from bone marrow of PHF19 knockout mice and to transduce them with oncoproteins, such as PML-RAR or BCR-ABL, as *Minucci et al* reported in 2002 [146]. This will allow us to study the effect of PHF19 deletion in transformed hematopoietic precursors. Moreover, these cells can be transplanted into recipient mice in order to investigate whether the lacking of PHF19 modulates *in vivo* the oncogenic capacity of these cells. In this regard, viruses to transfect hematopoietic precursors cells has been already generated, and preliminary experiments are now being carried on in the lab.

It is worth to note that although this experiment would inform us on the role of PHF19 in the leukemia formation, this still would not resemble what we are observing *in vitro* with already leukemic established cell lines. An alternative experiment that could be performed is to generate an inducible knockout for PHF19, and then after infecting bone marrow cells with leukemic oncogenes, triggering the deletion of PHF19. This would allow us to study *ex vivo* and *in vivo* the phenotype of depleting PHF19 in a more solid leukemic system.

2. Epigenetic inhibition of p21 by PHF19.

It has been reported that Polycomb in homeostasis regulates key target genes related with cell cycle and therefore is important for controlling cell fate and development [4]. Moreover, in cancer, since the cell cycle is misregulated, it has been described that Polycomb could inhibit the expression of cell cycle regulators and tumor suppressor genes [147]. It has also been described that leukemic cancer cells can be treated by inhibiting their cell cycle progression. Then, arrested cells can undergo either senescence or differentiation [112,114].

When we analyzed the RNA-seq data obtained in K562 cell line upon PHF19 depletion, we realized that in accordance with what we previously described phenotypically, cell cycle genes were misregulated. To further understand mechanistically how depletion of PHF19 leads to cell cycle arrest and differentiation, we performed a ChIP-seq analysis. Most of the PHF19 target genes had the tendency to increase their expression after PHF19 depletion, and among them p21 was one of the most overexpressed. We validated p21 overexpression at protein level, and, more importantly, the enrichment of PHF19 at p21 locus. This suggests that there is a direct link between PHF19, p21 and the cell growth arrest observed.

It has been reported that overexpression of p21 arrest myeloid leukemic cells in G1 phase [112]. Worthy of note, a way

to promote differentiation of leukemic cells is arresting their cell cycle [148]. This differentiation after cell cycle arrest has been observed when retinoic acid is applied to PML-RAR dependent leukemic cells [104,149]. Altogether, we hypothesize that high levels of PHF19 are preventing the cell cycle arrest of the K562 cells by inhibiting p21.

It is known that the tumor suppressor p53 is considered to be the guardian of the genome through multiples mechanisms. When arrest in G1-phase is required, p21 is activated by p53 [112]. On the other hand, p53 also inhibits many other proteins in an indirect way. It was described in a genome-wide meta-analysis using many different cancer cell lines, that p53 interact with PHF19 through DREAM complex (DP, RB-like, E2F4 and MuvB) [150,151]. Interestingly p53 protein is truncated and not functional in K562 cell line [152], so in our system p53 would not be able to inhibit PHF19. This observation could explain the fact that PHF19 is overexpressed in K562 when comparing with other cancer cell lines. Moreover, it was already described that EZH2 is targeting p21 locus in K562 cell lines [96]. Altogether, we hypothesize that PHF19 would be acting as an oncogene, targeting PRC2 to the chromatin and preventing the expression of p21, which may lead to uncontrolled cell growth.

The anticorrelation about PHF19 and p21 in clinical samples (**Figure R.22**) corroborates our hypothesis. Yet, we are not able to state that the differentiation phenotype towards the erythroid pathway when PHF19 is depleted is exclusively

because of the induction of p21. We tried to overexpress p21 in our cell model in order to know if this was enough to differentiate K562 towards the erythroid pathway. However, massive death due to the high levels of p21 hampered a clear conclusion. An alternative experiment could be to generate a conditional expression of p21 in K562 cell line that would temporally induce the expression and let us analyze if those cells undergo to differentiation because of p21 expression. Worthy of note, it was described that in K562 the overexpression of p21 drives differentiation towards the megakaryocyte lineage instead of the erythroid [148]. This observation is allowing us to hypothesize p21 induction upon depletion of PHF19 is necessary, but not sufficient to induce the observed erythroid differentiation (**Figure D.1**). Another way to tackle this problem would be to reduce p21 expression upon PHF19 depletion and assess whether differentiation was still triggered even in the absence of cell cycle arrest. We are in the process of preparing plasmids to do the knock down of p21 and combine it with PHF19 knock down.

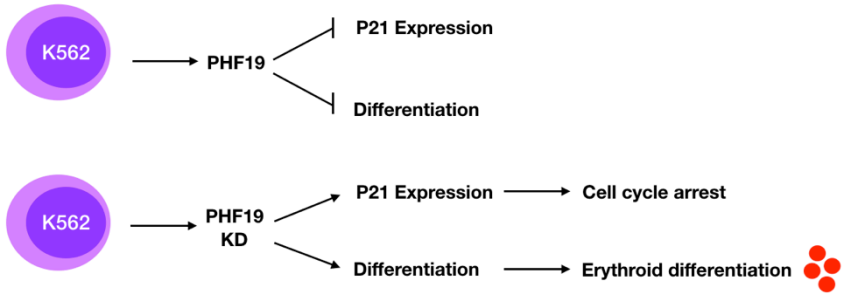


Figure D.1. Effect of PHF19 depletion in K562. Top panel: In a basal situation, PHF19 is highly expressed in K562 thus blocking the expression of P21 and preventing differentiation. Lower panel: When PHF19 is depleted in K562, P21 is

overexpressed arresting the cell cycle in G1 phase, therefore this leads to differentiate. Moreover this differentiation is specific towards the erythroid lineage.

It is worth to mention that all the mechanistic experiments have been performed in the K562 cell lines, and that the erythroid differentiation should not be considered for the other cell lines used in this thesis. Nonetheless, when PHF19 was depleted in NB4 and HL60 cell line, their proliferation decreased and at the RNA level, p21 was also up-regulated. Furthermore, reports from our lab indicate that depletion of SUZ12, a core component member of Polycomb, leaded differentiation of NB4 and HL60 cell lines [104]. Altogether, suggesting that PHF19 could be considered as targetable protein to induce cell cycle arrest and differentiation in NB4 and HL60 cell lines.

3. Mechanism behind depletion of PHF19

Recently, it has been published that MTF2, a PHF19 homolog, is essential for final erythropoiesis in mouse system. Moreover, the authors propose that this is mechanistically achieved by MTF2-PRC2 inhibition of the WNT signaling pathway [60]. To further understand the mechanistic behind the differentiation towards the erythroid lineage, we performed a ChIP-seq of MTF2 in control and PHF19 depletion condition. We observed that the majority of PHF19 targets were also targets of MTF2, and that after

PHF19 depletion there was a gain of MTF2 signal on those targets. Our first idea was that PHF19 was competing with MTF2 for the same target genes. Ontology of PHF19 target genes highlighted that some of the targets are related with the WNT signaling pathway. Since the majority of target genes were increasing MTF2 signal, we validated some of them related with WNT signaling pathway. We also performed ChIP-seq of the H3K27me3 mark and observed that PHF19 target genes also gained signal of H3K27me3. Altogether indicating that the erythroid differentiation, as *Rothberg et al* 2008 proposed [60], is due to the deposition of MTF2 at those PHF19 targets genes related with WNT signaling pathway. Nonetheless, more experiments should be performed to clarify why is MTF2 necessary, but not PHF19, for the repression of WNT genes that facilitates the erythroid differentiation since the homology of their sequence is quite high.

Worthy of note, we realized that not all the PHF19 target genes, even being targets of MTF2, were gaining MTF2 signal. Due to this observation, we selected the 50 target genes that were gaining the most MTF2 signal and those 50 target genes that were not increasing MTF2 signal. Looking carefully at the PHF19 target genes, we observed that those target genes where the increase in MTF2 was higher, were also initially more enriched for PHF19, and vice-versa. From these data and considering that they compete for the same targets, we conclude that there is a complex coordination

between PCL homologs and that the turnover between them after PHF19 KD is different depending on the initial signal of PHF19.

It should be also taken under consideration that between PHF19 target genes the initial expression level was different. As expected, target genes that had less PHF19 signal were initially more expressed, but at the same time were the most de-repressed after PHF19 depletion. On the other hand, those genes that had more signal of PHF19 were less expressed initially, and after PHF19 depletion they were not increasing their expression. All this data previously described allows us to hypothesize if MTF2, and therefore PRC2, are able to inhibit those targets that had higher levels of expression before PHF19 depletion (**Figure D.2**). Of note, these data is also reflecting some degree of expression permissiveness that low levels of PHF19 have on its target genes.

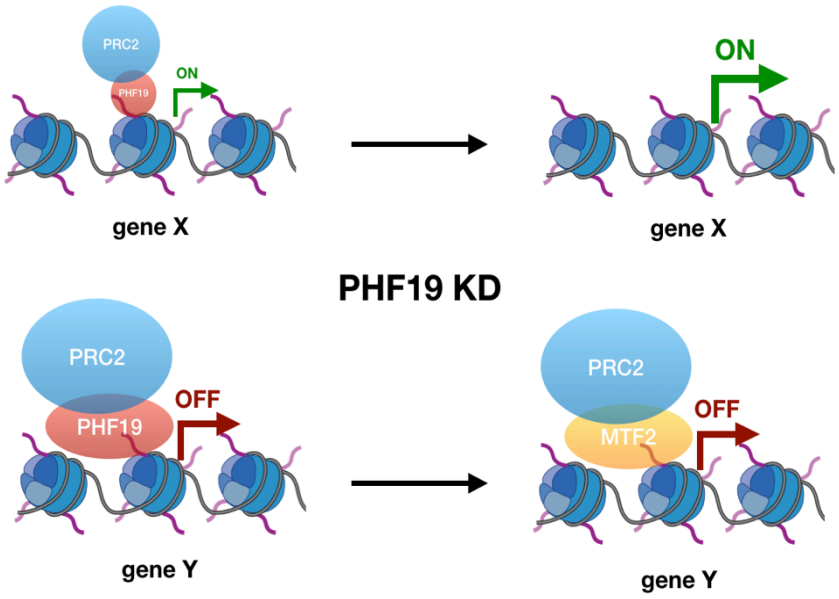


Figure D.2. PHF19-MTF2 turnover. Schematic picture about the turnover between PHF19 and MTF2 on PHF19 target genes. Our data indicates that in those target genes where PHF19 is more present, when PHF19 is depleted MTF2 replaces it and keep the gene inhibition. But in those target genes where signal of PHF19 is lower, and there is a lower levels of expression, then when PHF19 is depleted MTF2 is not able to replace it and the gene expression is full.

Different PRC2 complexes have been described depending on the associated factors (PRC2.1 and PRC2.2) [24,27]. EPOP or PALI, together with PCLs forms the complex PRC2.1 (**Figure I.3**). Our laboratory has previously reported that EPOP, together with PRC2, forms a complex interacting with Elonging BC in mESCs. This complex is allowing low levels of expression in the PRC2 genomic targets [42]. Then is not strange to think that, within PRC2.1, different PCL homologs may also account for different activities. Our results indicate that depending on the levels of PHF19, PRC2 is more or less permissive with respect to expression. Further analysis

should be performed in our system to fully understand this PHF19-MTF2 turnover and the low expression permissiveness on the PHF19-PRC2 targets.

4. Targeting PRC2 for myeloid leukemia treatment.

There is not just a single oncogenic insult driving myeloid leukemia. As described before, apart from the main oncogenic driver, such as a chromosome translocation, there are also accumulative mutations that increase the aggressiveness and facilitate the relapse capacity of leukemic cancer cells. Epigenetic and PRC2 mutations have been widely studied in myeloid leukemia [86,89,153]. Therefore, targeting PRC2 members could be considered for tackling those leukemic cells that have escaped to the classic treatment, or for targeting the leukemic initiating cells [94,95].

In this thesis we have explored the role of PHF19 and its potential use as a target for treating myeloid leukemia, especially CML. We have described a direct mechanism of PHF19 regulating cell cycle and preventing differentiation. Moreover, we have observed that there is a clear anticorrelation between the expressions of p21 and PHF19 in chronic myeloid leukemia patient data. Thus, the inactivation of PHF19 would be an approach to arrest leukemic cancer cells and induce them to differentiate. Targeting PRC2

components for cancer treatment has been a major challenge in last years. However, inhibiting the whole complex could have some unexpected consequences in some other tissues or even in the same leukemic cells, depending on the cellular stage of the disease [145,154]. Moreover, it has to be taken in consideration that PRC2 can be targeted to the chromatin through many different factors [22]. Therefore, it becomes crucial to study the accessory factors that lead PRC2 to the chromatin. In this thesis we have described the role of one of them, PHF19, in a particular leukemic context. Since PHF19 is an epigenetic factor and has a role in the chromatin, his function could be reversible. Then, further experiments analyzing which PHF19 domains are responsible for the binding of PRC2 to the chromatin would be necessary for a long-term drug development. Nevertheless, we have to be aware of the molecular limitations of tackling PHF19, since we will be affecting the regulation of just a subset of genes but not the main oncogenic driver. Therefore, combination of already existing drugs and the inhibition of PHF19 would be an approach to improve the patient therapies. For instance, for CML patients, tyrosine kinase inhibitors (TKI's) such as Imatinib are the standardized treatment [155]. Imatinib is bound to the BCR-ABL fusion protein inhibiting the phosphorylation activity. This results in the inhibition of proliferative signal to the nucleus and led the leukemic cell to the induction of apoptosis [155]. Although it is already highly effective, more then 80% of the Imatinib treated patients in chronic phase are

alive after 10 years [156], based in our observations combinatorial treatment with lower doses of Imatinib and depletion of PHF19 could be explored in order to increase the efficiency of the CML treatment.

CONCLUSIONS

Derived from the results presented in this thesis we can draw the following conclusions:

1. PHF19 depletion inhibits cell growth in different myeloid leukemia cell lines.

2. PHF19 depletion induces differentiation towards the erythroid lineage in chronic myeloid leukemia cell line K562.

3. PHF19 depletion prevents K562 differentiation towards the megakaryocytic lineage induced by PMA.

5. PHF19 is enriched at p21 locus and its depletion increases p21 expression leading to cell cycle arrest.

6. PHF19 competes with MTF2 for the same target genes.

7. Upon PHF19 depletion, MTF2 signal is increased in those target genes that needed to be repressed for the erythroid differentiation.

8. Expression levels of PHF19 target genes anticorrelate with PHF19 signal.

9. Upon PHF19 depletion, MTF2 increases in those targets were PHF19 was highly present, and do not increase, in those targets where PHF19 was at lower levels.

MATHERIALS AND METHODS

1. Cell culture, infection, proliferation, apoptosis and differentiation methods

Cell culture

NB4, HL60 and K562 cells were cultured at 37ºC and 5% of CO_2 in RPMI medium supplemented with 10% of fetal bovine serum. HEK293T and PLATINUM A cells were cultured at 37ºC and 5% CO_2 in DMEM supplemented with 10% of fetal bovine serum.

To monitor cell growth, leukemic cell lines were seeded at approximately 2,5 x 10^5 cells / ml for the NB4 and HL60 cell lines and 3 x 10^5 cells / ml for the K562 cell line, then cells were counted and diluted every two days.

For ARA-C treatment, ARA-C was dissolved in water and then dissolved in cell culture medium to obtain the desired concentration.

For PMA treatment, PMA was dissolved in dimethyl sulphoxide (DMSO), and then dissolved in cell culture medium to obtain the desired concentration.

Generation of the lentiviral vector for PHF19 knock down

PLKO.1 lentiviral vector was used for the expression of shRNAs against PHF19. The following oligos were ordered from SIGMA and used to clone shRNA specific for PHF19 into pLKO.1 vector:

Fw-shPHF19#1:

CCGG***AAGCTTCCATCCACATGTGTT***CTCGAG***AACACAT GTGGATGGAGCTT***TTTTTG

Rv-shPHF19#1:

AATTCAAAAA***AAGCTTCCATCCACATGTGTT***CTCGAG***AA CACATGTGGATGGAAGCTT***

Fw-shPHF19#2:

CCGG***GCCACACATTTGAGAGCATCA***CTCGAG***TGATGCT CTCAAATGTGTGGC***TTTTTG

Rv-shPHF19#2:

AATTCAAAAA***GCCACACATTTGAGAGCATCA***CTCGAG***TG ATGCTCTCAAATGTGTGGC***

Then oligos were annealed by incubating the following reaction at 95ºC 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 ddH_2O. 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 2ul of ligations was transformed into competent bacterial cells and plated in LB agar plates with 100ug/ml of ampicillin resistance. Collonies 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.

Cell transfection and infection (Lentivirus vector)

For the production of viruses to do the knock-downs, HEK293T were plated at a density of 2 x 10^6 cells in a p10 plate. The day after, the following cocktail was made for each transfection; 10ug of pLKO.1 of plasmid, 6ug of pCML-dR8.91, 5ug of pCMV-VSGV, 62ul of 2M $CaCl_2$ and up to 500ul of H_2O. The cocktail was then added drop by drop to 500ul of 2X HBS while bubbling the mix using a Pasteur pipette. The mixture was incubated 15 minutes at room temperature before adding it to the HEK293T plate drop by drop. Cells were incubated with the transfection mix for 14-16 hour. Then medium was replaced by fresh medium and cells were incubated 24h.

Medium with the lentiviral particles were harvested and filtered through a 45.um filter into a 15 ml tube. Fresh medium was added to the HEK293T cell plate and medium with lentiviral particles were harvested again the following day. For the infection, a non-treated 6-well plate was used. Two rounds of infection were performed according to the days the medium was harvested. 5 x 10^5 of cells was plated in 1ml of medium, and then 1ml of the medium with lentivirus was added. Then cells were spininoculated (1000g, 90min, 32ºC) in presence of protamine sulfate followed by an additional 3

hours incubation of 37ºC and 5% CO_2. After the two rounds of infection, cells were selected with 1ug/ml of puromycin (Sigma).

Cell transfection and infection (Retrovirus vector)

For the production of viruses to do the overexpression, Platinum A cells were plated at a density of 2 x 10^6 cells in a p10 plate. The day after, the following cocktail was made for each transfection; 10ug of pBabe plasmid, 62ul of 2M $CaCl_2$ and up to 500ul of H_2O. Then the protocol followed was equal to the previously described for pLKO.1 plasmid. PBabe plasmid with p21 clonned was given by Bill Keyes lab from IGMBC at Strasburg.

Cell growth curve

2,5 x 10^5 cells/ml were seeded in a non-treated 6 well plate. Then cells were counted every two days and seeded again the same number of cells for each condition. Number of cells was counted using Countess^{tm} chamber slides and trypan blue staining to exclude death cells. The differences were counted in an accumulative way.

Cell Apoptosis assay

The cell apoptosis assay was performed using Violet Annexin V / Dead Cell Apoptosis Kit (Invitrogen) according to the manufacture's protocol. Post staining, cells were analyzed by flow cytometry.

Cell proliferation

Cells were treated with 10uM of BrdU solution during 30 minutes and then analyzed the BrdU incorporation using BrdU Flow Kit (BD Pharmingen) according to manufacture's protocol. The percentage of BrdU-positive cells was analyzed by flow cytometry.

Cell cycle

1 x 10^6 cells were washed with cold PBS and resuspended in 0,9ml of EDTA-PBS (EDTA 5mM). Then cells were permeabilized adding 2,1ml of cold Ethanol 100% dropwise while the mixture was softly shacking in the vortex. Cells were kept at 4ºC until the following day. The following day cells were resuspended in PI staining buffer that contains 955ul of PBS + 30ul of solution A (38ul Sodium citrate 0'5M + 562ul Propidium Iodide 500ug/ml) + 2ul of RNAse A (Thermofisher). Cells then were incubated at 37ºC for 1hour and then analyzed by flow cytometry. Cell cycle profiles were analyzed with ModFit LT™ software.

Cell differentiation

K562 cell lines either in normal conditions or after Ara-C or PMA treatment were rinsed twice with PBS and incubated for 45mins again CD235a-PE or CD61-FITC (antibodies table), at 4ºC and protected from light. After washing twice with cold

PBS, cells were suspended in PBS with DAPI and then analyzed by flow cytometry. Analysis was performed using Flowjo Software.

2. Gene Expression methods

qPCR-RNA analysis

RNA was extracted with RNAeasy Mini Kit (Qiagen) according to the manufacture's protocol. Then cDNA was generated from 1ug of RNA with the First Strand cDNA Synthesis Kit (Fermentas) according to the manufacture's protocol. cDNA was diluted to 100ul of water and 2ul 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*.

RNA-seq analysis

RNA was extracted with RNAeasy Mini Kit. Single-end RNA-sequencing was performed using 1ug RNA. Samples were obtained in two different days and two indendent biological replicates were secuenced from the three conditions shCT sh#1 and sh#2. The CRG Genomics Unit performed the quality control and the library preparation. Libraries were sequenced using Illumina HiSeq2000 sequencer.

The RNA-seq samples were mapped against the hg19 human genome assembly using TopHat [157] with the option –g 1 to discard those reads that could not be uniquely mapped in just one region. DESeq2 [158] was run to quantify the expression of every annotated transcript using the RefSeq catalog of exons and to identify each set of differentially expressed genes. Expression values shown in the boxplots correspond to the average RPKMs across the different replicates in each condition.

GSEA of the pre-ranked lists of genes by DESeq2 stat value was performed with the GSEA software [159]. Human genes were ranked by the ratio between KD and CT RNA-seq expression.

3. Protein analysis method

Western blot

Cell pellets were washed twice with PBS, then pellets were resuspended in hypotonic buffer (10mM TrisHCL p7,4 + 10mM KCL + 15mM $MgCl_2$) with inhibitors (cOmplete EDTA-free tablet from Roche50x + PMSF 1mM + Na-orthovanadate 1mM + B-glycerophosphate 20mM) during 10 minutes in ice. Then pellets were centrifuged during 5 minutes at 4ºC at 700g. The pellet was resuspended in nuclear lysis buffer (300mM NaCL + 50mM Hepes pH 7,5 + 0,5% NP40 + 2,5mM

$MgCl_2$) with inhibitors (cOmplete EDTA-free tablet from Roche50x + PMSF 1mM + Na-orthovanadate 1mM + B-glycerophosphate 20mM) and Benzonase (50 units/ 500ul of buffer). Then pellets were centrifuged during 30 minutes at 4ºC at 16000g. The supernatant was considered the protein extract of the nuclear protein. Protein extract were quantified by Bradford assay (Bio-Rad). Around 70ug of protein were taken and diluted in 5x Laemli buffer and heated for 5 minutes in 100ºC. Protein samples were loaded in a NuPAGE™ 4-12% Bis-Tris precast gel from Invitrogen and then were run in running buffer from Invitrogen at 100V. Proteins were transferred onto nitrocellulose membranes at 300mA for 70 minutes at 4ºC in transfer buffer (25mM Tris-HCL pH 8,3, 200mM glycine, 20% methanol). Trasnfered proteins were blocked with 5% of milf in TBS-Tween (10mM Trsi-HCl pH 7,5, 100mM NaCl and 0,1% Tween-20) for 30mins at room temperature. Then membranes were incubated overnight with the primary antibody (in 5% of milf in TBS-Tween) at 4ºC (see antibody table). The following day, membranes were washed twice for 10 minutes with TBS-Tween followed by incubation of the secondary antibody conjugated to the horseradish peroxidase (1:5000), diluted in TBS-Tween, for 1h at room temperature. Then membranes were washed twice with TBS-Tween at room temperature. Then, proteins were detected by enhanced chemilluminescence reagent (Pierce ECL Western Blotting Substrate, Thermo Scientific).

Chromatin immunoprecipitation (ChIP) and ChIP-seq

25 x 10^6 cells were washed twice with PBS. Then cells were sinlge crosslinked or double crosslinked depending if ChIP was against proteins or histone mark. For the double crosslinking cells were resuspended in 10ml of PBS 1mM $MgCl_2$ + 40ul of ChIP Cross-link Gold from Diagenode and shaking during 30 minutes at room temperature. Then cells were washed twice with cold PBS. Cells then were resuspended in the fixation buffer according to manufacture's protocol from Active Motive. Once cells were crosslinked, pellets were prepared for sonication as ChIP-IT High Sensitivity (53040) protocol recommends. Cells were sonicated during 80 cycles (30 seconds ON/ 30 seconds OFF) with a Bioruptor (Diagenode) at maximum output. After sonication, cells were centrifuged at maximum speed at 4ºC for 20mins to pellet the cell debris. To check the chromatin size, 25 ul of chromatin was diluted with 175 ul of TE + 1 u of RNAse A and incubated during 30 minutes at 37ºC. Then chromatin was incubated for 30 minutes at 55ºC plus 2ul of Proteinase K. Afterwards chromatin was incubated 2 hours at 80ºC. Then chromatin was precipitated as the manufacturers protocol described. Chromatin was analyzed in the Nanodrop. Then around 800ng were analyzed in a 1% agarose gel to corroborate that chromatin was between 200-500 bp of size. Then the inmunoprecipitations was performed according to Active Motiv protocols from the ChIP-IT High Sensitivity kit adding 1% of total IP of *Drossophila* chromatin. Then Ab used

for the IP were, the protein of interest plus the Ab of the drosophila.

When ChIP was performed against histone modifications, the protocol followed was the same as previously described but without doing the previous crosslink with ChIP Cross-link Gold from Diagenode.

ChIP-seq for all samples were performed using 2-10ng of precipitated ChIP DNA followed by sequencing library preparation, quality control and quantification. Libraries were sequenced in the CRG sequencing unit with the Illumina Hiseq2000 sequencer.

For the bioinformatics analysis, ChIPseq samples normalized by spike-in were mapped against a synthetic genome constituted by the human and the fruit fly chromosomes (hg19 + dm3) using Bowtie with the option -m 1 to discard reads that did not map uniquely to one region [160].

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 [161].

Then visual inspection of the target genes was performed and to define the subsets FDR value was used, as is described in Results section.

The genome distribution of each set of peaks was calculated by counting the number of peaks fitted on each class of region according to RefSeq annotations [162]. Distal region is the region within 2.5 Kbp and 0.5 Kbp upstream of the transcription start site (TSS). Proximal region is the region

within 0.5 Kbp upstream of the TSS. UTR, untranslated region; CDS, protein coding sequence; intronic regions, introns; and the rest of the genome, intergenic. Peaks that overlapped with more than one genomic feature were proportionally counted the same number of times.

Calculating the genome distribution of all features in the full genome we generated Spie charts, and the R caroline package was used to combine the piechart of each set of peaks with the full genome distribution [163].

Each set of target genes was retrieved by matching the ChIPseq peaks in the region 2.5 Kbp upstream of the TSS until the end of the transcripts as annotated in RefSeq.

Reports of functional enrichments of GO and other genomic libraries were generated using the EnrichR tool [129].

The aggregated plots showing the average distribution of ChIPseq reads around the TSS of each target gene were generated by counting the number of reads for each region according to RefSeq and then averaging the values for the total number of mapped reads of each sample and the total number of genes in the particular gene set.

The aggregated plots of ChIPseq samples containing spike-in were generated by counting the number of reads mapped in human around the TSS for each gene and then averaging these values for the total number of reads mapped on the *Drosophila* genome and the number of targets of the gene list, as previously described [164].

Boxplots showing the ChIPseq level distribution for a particular ChIP experiment on a set of genomic peaks were calculated by determining the maximum value on this region at this sample, which was assigned afterwards to the corresponding peak. The values of the samples including spike-in were corrected by the number of *Drosophila* reads mapped of the sequencing experiment.

Each point on the scatterplots of ChIPseq intensities between two conditions was calculated by determining the maximum value of the sample inside each peak at each condition.

The UCSC genome browser was used to generate the screenshots of each group of experiments along the manuscript [165].

4. List of primers used

Expression mRNA-qPCR oligos from 5' to 3':

CCNA2 FW: ACCCAGGGTTCTCAGAATGG
CCNA2 RV: CTTGGATGCCAGTCTTACTCA

CDK4 FW: GAAACTCTGAAGCCGACCAG
CDK4 RV: AGGCAGAGATTCGCTTGTGT

CCND1 FW: GCCGAGAAGCTGTGCATC
CCND1 RV: CCACTTGAGCTTGTTCACCA

GYPA FW: CAAACGGGACACATATGCAG
GYPA RV: TCCAATAACACCAGCCATCA

KEL FW: ACCATGGGGAGACTGTCCT
KEL RV: GGGCTTCCTACACATCACCT

p21 FW: CAGCTGCCGAAGTCAGTTCC
p21 RV: GTTCTGACATGGCGCCTCC

PHF19 FW: CAGCAGAAAAGGCGAGTTTATAG
PHF19 RV: CTCCAGGCTGAGGTGAAGTC

RP0 FW: TTCATTGTGGGAGCAGAC
RP0 RV: CAGCAGTTTCTCCAGAGC

TAL1 FW: CATGGTGCAGCTGAGTCCT
TAL1 RV: GGTGGTGAACATAGGGAAGG

Y-GLOBIN FW: CACAAAGCACCTGGATGATCT
Y-GLOBIN RV: AAACGGTCACCAGCACATTT

Genomic ChIP-qPCR oligos from 5' to 3':

ATF3 FW: GTGGGTGGTCTGAGTGAGGT
ATF3 RV: CACAGTTTGGTAATTTGGGGTAG

BMI1 FW: CGGCCCCACGTTTTAGTT

BMI1 RV: CCGTTTCATCCAAATCTACCTCT

CXCR4 FW: GTATATTGGGCGGGAGTGTC
CXCR4 RV: ACACGAGGATGGCAAGAGAC

GATA2 FW: ATAATTTTAAATGCCCCGTTTTC
GATA2 RV: CCCTTTATCCTGTGTGACTGG

GATA3 FW: CACGTGGAGAGAATTAGGAGGA
GATA3 RV: GTTTCTTTCAGAGGGAGTCTGC

INTERGENIC FW: ACAGGATAAAGTTGGCATAACCA
INTERGENIC RV: CAACAAAACCGTTTGGAATACAT

NODAL FW: GCGACTTCCTTACTCGACCTC
NODAL RV: TCTCTTCCTGCTTCTTGAGGA

NOTUM FW: CCGAGGCTGGGCTTATTT
NOTUM RV: GGGAAGAAAAGGCGATGC

p21/CDKN1A FW: ATGTCATCCTCCTGATCTTTTCA
p21/ CDKN1A RV: AGAATGAGTTGGCACTCTCCAG

PDGFRA FW: GGGGTGTCAGTTACAGAAGGTCT
PDGFRA RV: CTGCCTGGATTAAAGTGTTAGGG

5. Table of Antibodies used

Antibody	Specie	Provider	Reference	ChIP	WB	FACS
CD235a	Human	Invitrogen	12-9987-82			1:400
CD61	Human	eBioscience	11-0619-42			1:100
Drossophila	Human	Active Motif	#104597	1ul		
H3K27me3	Human	Millipore	07-449	3ul		
MTF2	Human	Protein Tech	16208-1-AP	5ul		
P21	Human	Cell signaling	#2947		1:1000 (rabbit)	
PHF19	Human	Cell signaling	#77271	5ul	1:1000 (rabbit)	
Tubulin	Human	Abcam	ab7291		1:1000 (mouse)	

Table MM.1 Antibodies and their applications. The antibodies used in this study, their dilution for use and commercial information.

REFERENCES

REFERENCES

1 Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871 (1974).
2 Hansen, J. C. Human mitotic chromosome structure: what happened to the 30-nm fibre? *EMBO J* **31**, 1621-1623, doi:10.1038/emboj.2012.66 (2012).
3 Talbert, P. B. & Henikoff, S. Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* **18**, 115-126, doi:10.1038/nrm.2016.148 (2017).
4 Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**, 846-856, doi:10.1038/nrc1991 (2006).
5 Huisinga, K. L., Brower-Toland, B. & Elgin, S. C. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* **115**, 110-122, doi:10.1007/s00412-006-0052-x (2006).
6 Solovei, I., Thanisch, K. & Feodorova, Y. How to rule the nucleus: divide et impera. *Curr Opin Cell Biol* **40**, 47-59, doi:10.1016/j.ceb.2016.02.014 (2016).
7 Richards, E. J. & Elgin, S. C. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**, 489-500, doi:10.1016/s0092-8674(02)00644-x (2002).
8 Saksouk, N., Simboeck, E. & Dejardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8**, 3, doi:10.1186/1756-8935-8-3 (2015).
9 Trojer, P. & Reinberg, D. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* **28**, 1-13, doi:10.1016/j.molcel.2007.09.011 (2007).
10 Heard, E. Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Curr Opin Genet Dev* **15**, 482-489, doi:10.1016/j.gde.2005.08.009 (2005).
11 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705, doi:10.1016/j.cell.2007.02.005 (2007).
12 Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science* **330**, 612-616, doi:10.1126/science.1191078 (2010).
13 Cavalli, G. & Heard, E. Advances in epigenetics link genetics to the environment and disease. *Nature* **571**, 489-499, doi:10.1038/s41586-019-1411-0 (2019).

14 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395, doi:10.1038/cr.2011.22 (2011).
15 Bhaumik, S. R., Smith, E. & Shilatifard, A. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol* **14**, 1008-1016, doi:10.1038/nsmb1337 (2007).
16 Schuettengruber, B., Bourbon, H. M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34-57, doi:10.1016/j.cell.2017.08.002 (2017).
17 Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-570 (1978).
18 Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. *Sci Adv* **1**, e1500737, doi:10.1126/sciadv.1500737 (2015).
19 Wang, L. *et al.* Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* **14**, 637-646, doi:10.1016/j.molcel.2004.05.009 (2004).
20 Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* **15**, 57-67, doi:10.1016/j.molcel.2004.06.020 (2004).
21 Montgomery, N. D. *et al.* The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr Biol* **15**, 942-947, doi:10.1016/j.cub.2005.04.051 (2005).
22 Vizan, P., Beringer, M., Ballare, C. & Di Croce, L. Role of PRC2-associated factors in stem cells and disease. *FEBS J* **282**, 1723-1735, doi:10.1111/febs.13083 (2015).
23 Hodgson, J. W., Argiropoulos, B. & Brock, H. W. Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Mol Cell Biol* **21**, 4528-4543, doi:10.1128/MCB.21.14.4528-4543.2001 (2001).
24 Laugesen, A., Hojfeldt, J. W. & Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol Cell* **74**, 8-18, doi:10.1016/j.molcel.2019.03.011 (2019).

25 Nekrasov, M., Wild, B. & Muller, J. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *EMBO Rep* **6**, 348-353, doi:10.1038/sj.embor.7400376 (2005).

26 Ciferri, C. *et al.* Molecular architecture of human polycomb repressive complex 2. *Elife* **1**, e00005, doi:10.7554/eLife.00005 (2012).

27 Hauri, S. *et al.* A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep* **17**, 583-595, doi:10.1016/j.celrep.2016.08.096 (2016).

28 Takeuchi, T. *et al.* Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev* **9**, 1211-1222, doi:10.1101/gad.9.10.1211 (1995).

29 Lee, Y. *et al.* Jumonji, a nuclear protein that is necessary for normal heart development. *Circ Res* **86**, 932-938 (2000).

30 Li, G. *et al.* Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev* **24**, 368-380, doi:10.1101/gad.1886410 (2010).

31 Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306-310, doi:10.1038/nature08788 (2010).

32 Son, J., Shen, S. S., Margueron, R. & Reinberg, D. Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev* **27**, 2663-2677, doi:10.1101/gad.225888.113 (2013).

33 Sanulli, S. *et al.* Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol Cell* **57**, 769-783, doi:10.1016/j.molcel.2014.12.020 (2015).

34 Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* **21**, 569-571, doi:10.1038/nsmb.2833 (2014).

35 Kaneko, S. *et al.* Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell* **53**, 290-300, doi:10.1016/j.molcel.2013.11.012 (2014).

36 Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043, doi:10.1126/science.1076997 (2002).

37 Kim, H., Kang, K. & Kim, J. AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res* **37**, 2940-2950, doi:10.1093/nar/gkp149 (2009).
38 Grijzenhout, A. *et al.* Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. *Development* **143**, 2716-2723, doi:10.1242/dev.123935 (2016).
39 Conway, E. *et al.* A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. *Mol Cell* **70**, 408-421 e408, doi:10.1016/j.molcel.2018.03.005 (2018).
40 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* **29**, 229-240, doi:10.1002/stem.578 (2011).
41 De Cegli, R. *et al.* Reverse engineering a mouse embryonic stem cell-specific transcriptional network reveals a new modulator of neuronal differentiation. *Nucleic Acids Res* **41**, 711-726, doi:10.1093/nar/gks1136 (2013).
42 Beringer, M. *et al.* EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. *Mol Cell* **64**, 645-658, doi:10.1016/j.molcel.2016.10.018 (2016).
43 Liefke, R., Karwacki-Neisius, V. & Shi, Y. EPOP Interacts with Elongin BC and USP7 to Modulate the Chromatin Landscape. *Mol Cell* **64**, 659-672, doi:10.1016/j.molcel.2016.10.019 (2016).
44 Duncan, I. M. Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. *Genetics* **102**, 49-70 (1982).
45 Savla, U., Benes, J., Zhang, J. & Jones, R. S. Recruitment of Drosophila Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. *Development* **135**, 813-817, doi:10.1242/dev.016006 (2008).
46 Nekrasov, M. *et al.* Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J* **26**, 4078-4088, doi:10.1038/sj.emboj.7601837 (2007).
47 Friberg, A., Oddone, A., Klymenko, T., Muller, J. & Sattler, M. Structure of an atypical Tudor domain in the Drosophila

Polycomblike protein. *Protein Sci* **19**, 1906-1916, doi:10.1002/pro.476 (2010).

48 Coulson, M., Robert, S., Eyre, H. J. & Saint, R. The identification and localization of a human gene with sequence similarity to Polycomblike of Drosophila melanogaster. *Genomics* **48**, 381-383, doi:10.1006/geno.1997.5201 (1998).

49 Cao, R. *et al.* Role of hPHF1 in H3K27 methylation and Hox gene silencing. *Mol Cell Biol* **28**, 1862-1872, doi:10.1128/MCB.01589-07 (2008).

50 Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V. & Reinberg, D. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol* **28**, 2718-2731, doi:10.1128/MCB.02017-07 (2008).

51 Li, H. *et al.* Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* **549**, 287-291, doi:10.1038/nature23881 (2017).

52 Musselman, C. A. *et al.* Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nat Struct Mol Biol* **19**, 1266-1272, doi:10.1038/nsmb.2435 (2012).

53 Kycia, I. *et al.* The Tudor domain of the PHD finger protein 1 is a dual reader of lysine trimethylation at lysine 36 of histone H3 and lysine 27 of histone variant H3t. *J Mol Biol* **426**, 1651-1660, doi:10.1016/j.jmb.2013.08.009 (2014).

54 Walker, E. *et al.* Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* **6**, 153-166, doi:10.1016/j.stem.2009.12.014 (2010).

55 Chou, D. M. *et al.* A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc Natl Acad Sci U S A* **107**, 18475-18480, doi:10.1073/pnas.1012946107 (2010).

56 Boulay, G. *et al.* Hypermethylated in cancer 1 (HIC1) recruits polycomb repressive complex 2 (PRC2) to a subset of its target genes through interaction with human polycomb-like (hPCL) proteins. *J Biol Chem* **287**, 10509-10524, doi:10.1074/jbc.M111.320234 (2012).

57 Yang, Y. *et al.* Polycomb group protein PHF1 regulates p53-dependent cell growth arrest and apoptosis. *J Biol Chem* **288**, 529-539, doi:10.1074/jbc.M111.338996 (2013).

58 Casanova, M. *et al.* Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. *Development* **138**, 1471-1482, doi:10.1242/dev.053652 (2011).

59 Perino, M. *et al.* MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat Genet* **50**, 1002-1010, doi:10.1038/s41588-018-0134-8 (2018).

60 Rothberg, J. L. M. *et al.* Mtf2-PRC2 control of canonical Wnt signaling is required for definitive erythropoiesis. *Cell Discov* **4**, 21, doi:10.1038/s41421-018-0022-5 (2018).

61 Wang, S., Robertson, G. P. & Zhu, J. A novel human homologue of Drosophila polycomblike gene is up-regulated in multiple cancers. *Gene* **343**, 69-78, doi:10.1016/j.gene.2004.09.006 (2004).

62 Boulay, G., Rosnoblet, C., Guerardel, C., Angrand, P. O. & Leprince, D. Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes. *Biochem J* **434**, 333-342, doi:10.1042/BJ20100944 (2011).

63 Hunkapiller, J. *et al.* Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. *PLoS Genet* **8**, e1002576, doi:10.1371/journal.pgen.1002576 (2012).

64 Ballare, C. *et al.* Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat Struct Mol Biol* **19**, 1257-1265, doi:10.1038/nsmb.2434 (2012).

65 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* **19**, 1273-1281, doi:10.1038/nsmb.2449 (2012).

66 Cai, L. *et al.* An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol Cell* **49**, 571-582, doi:10.1016/j.molcel.2012.11.026 (2013).

67 Ghislin, S., Deshayes, F., Middendorp, S., Boggetto, N. & Alcaide-Loridan, C. PHF19 and Akt control the switch between proliferative and invasive states in melanoma. *Cell Cycle* **11**, 1634-1645, doi:10.4161/cc.20095 (2012).

68 Li, G. *et al.* Altered expression of polycomb group genes in glioblastoma multiforme. *PLoS One* **8**, e80970, doi:10.1371/journal.pone.0080970 (2013).
69 Deng, Q. *et al.* PHF19 promotes the proliferation, migration, and chemosensitivity of glioblastoma to doxorubicin through modulation of the SIAH1/beta-catenin axis. *Cell Death Dis* **9**, 1049, doi:10.1038/s41419-018-1082-z (2018).
70 Xu, H. *et al.* MicroRNA-195-5p acts as an anti-oncogene by targeting PHF19 in hepatocellular carcinoma. *Oncol Rep* **34**, 175-182, doi:10.3892/or.2015.3957 (2015).
71 Ji, Y. *et al.* miR-155 harnesses Phf19 to potentiate cancer immunotherapy through epigenetic reprogramming of CD8(+) T cell fate. *Nat Commun* **10**, 2157, doi:10.1038/s41467-019-09882-8 (2019).
72 Islam, A. The origin and spread of human leukemia. *Med Hypotheses* **39**, 110-118 (1992).
73 Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391-2405, doi:10.1182/blood-2016-03-643544 (2016).
74 Grigoropoulos, N. F., Petter, R., Van 't Veer, M. B., Scott, M. A. & Follows, G. A. Leukaemia update. Part 1: diagnosis and management. *BMJ* **346**, f1660, doi:10.1136/bmj.f1660 (2013).
75 Dohner, H., Weisdorf, D. J. & Bloomfield, C. D. Acute Myeloid Leukemia. *N Engl J Med* **373**, 1136-1152, doi:10.1056/NEJMra1406184 (2015).
76 Sawyers, C. L. Chronic myeloid leukemia. *N Engl J Med* **340**, 1330-1340, doi:10.1056/NEJM199904293401706 (1999).
77 Abdul-Hamid, G. *Acute Leukemia- The Scientist's Perspective and Challenge*. (InTech, 2011).
78 Fialkow, P. J., Janssen, J. W. & Bartram, C. R. Clonal remissions in acute nonlymphocytic leukemia: evidence for a multistep pathogenesis of the malignancy. *Blood* **77**, 1415-1417 (1991).
79 Hehlmann, R., Hochhaus, A., Baccarani, M. & European, L. Chronic myeloid leukaemia. *Lancet* **370**, 342-350, doi:10.1016/S0140-6736(07)61165-9 (2007).
80 O'Brien, S. G. *et al.* Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase

chronic myeloid leukemia. *N Engl J Med* **348**, 994-1004, doi:10.1056/NEJMoa022457 (2003).

81 Sklarz, L. M. *et al.* Genetic Mutations in a Patient with Chronic Myeloid Leukemia Showing Blast Crisis 10 Years After Presentation. *Anticancer Res* **38**, 3961-3966, doi:10.21873/anticanres.12682 (2018).

82 Machova Polakova, K., Koblihova, J. & Stopka, T. Role of epigenetics in chronic myeloid leukemia. *Curr Hematol Malig Rep* **8**, 28-36, doi:10.1007/s11899-012-0152-z (2013).

83 Germing, U., Kobbe, G., Haas, R. & Gattermann, N. Myelodysplastic syndromes: diagnosis, prognosis, and treatment. *Dtsch Arztebl Int* **110**, 783-790, doi:10.3238/arztebl.2013.0783 (2013).

84 Montalban-Bravo, G. & Garcia-Manero, G. Myelodysplastic syndromes: 2018 update on diagnosis, risk-stratification and management. *Am J Hematol* **93**, 129-147, doi:10.1002/ajh.24930 (2018).

85 Issa, J. P. Epigenetic changes in the myelodysplastic syndrome. *Hematol Oncol Clin North Am* **24**, 317-330, doi:10.1016/j.hoc.2010.02.007 (2010).

86 Heuser, M., Yun, H. & Thol, F. Epigenetics in myelodysplastic syndromes. *Semin Cancer Biol* **51**, 170-179, doi:10.1016/j.semcancer.2017.07.009 (2018).

87 Issa, J. P. The myelodysplastic syndrome as a prototypical epigenetic disease. *Blood* **121**, 3811-3817, doi:10.1182/blood-2013-02-451757 (2013).

88 Wouters, B. J. & Delwel, R. Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. *Blood* **127**, 42-52, doi:10.1182/blood-2015-07-604512 (2016).

89 Di Carlo, V., Mocavini, I. & Di Croce, L. Polycomb complexes in normal and malignant hematopoiesis. *J Cell Biol* **218**, 55-69, doi:10.1083/jcb.201808028 (2019).

90 Honda, H., Nagamachi, A. & Inaba, T. -7/7q- syndrome in myeloid-lineage hematopoietic malignancies: attempts to understand this complex disease entity. *Oncogene* **34**, 2413-2425, doi:10.1038/onc.2014.196 (2015).

91 Ernst, T. *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* **42**, 722-726, doi:10.1038/ng.621 (2010).

92 Bejar, R. *et al.* Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med* **364**, 2496-2506, doi:10.1056/NEJMoa1013343 (2011).
93 Patel, J. P. *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* **366**, 1079-1089, doi:10.1056/NEJMoa1112304 (2012).
94 Xie, H. *et al.* Chronic Myelogenous Leukemia- Initiating Cells Require Polycomb Group Protein EZH2. *Cancer Discov* **6**, 1237-1247, doi:10.1158/2159-8290.CD-15-1439 (2016).
95 Scott, M. T. *et al.* Epigenetic Reprogramming Sensitizes CML Stem Cells to Combined EZH2 and Tyrosine Kinase Inhibition. *Cancer Discov* **6**, 1248-1257, doi:10.1158/2159-8290.CD-16-0263 (2016).
96 Liu, L. *et al.* MYCN contributes to the malignant characteristics of erythroleukemia through EZH2-mediated epigenetic repression of p21. *Cell Death Dis* **8**, e3126, doi:10.1038/cddis.2017.526 (2017).
97 Tanaka, S. *et al.* Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. *Blood* **120**, 1107-1117, doi:10.1182/blood-2011-11-394932 (2012).
98 Neff, T. *et al.* Polycomb repressive complex 2 is required for MLL-AF9 leukemia. *Proc Natl Acad Sci U S A* **109**, 5028-5033, doi:10.1073/pnas.1202258109 (2012).
99 Lund, K., Adams, P. D. & Copland, M. EZH2 in normal and malignant hematopoiesis. *Leukemia* **28**, 44-49, doi:10.1038/leu.2013.288 (2014).
100 Basheer, F. *et al.* Contrasting requirements during disease evolution identify EZH2 as a therapeutic target in AML. *J Exp Med* **216**, 966-981, doi:10.1084/jem.20181276 (2019).
101 Gambacorti-Passerini, C. *et al.* Bosutinib efficacy and safety in chronic phase chronic myeloid leukemia after imatinib resistance or intolerance: Minimum 24-month follow-up. *Am J Hematol* **89**, 732-742, doi:10.1002/ajh.23728 (2014).
102 Klampfl, T. *et al.* Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. *Blood* **118**, 167-176, doi:10.1182/blood-2011-01-331678 (2011).

103 Brecqueville, M. *et al.* Mutations and deletions of the SUZ12 polycomb gene in myeloproliferative neoplasms. *Blood Cancer J* **1**, e33, doi:10.1038/bcj.2011.31 (2011).

104 Villa, R. *et al.* Role of the polycomb repressive complex 2 in acute promyelocytic leukemia. *Cancer Cell* **11**, 513-525, doi:10.1016/j.ccr.2007.04.009 (2007).

105 Ikeda, K. *et al.* Maintenance of the functional integrity of mouse hematopoiesis by EED and promotion of leukemogenesis by EED haploinsufficiency. *Sci Rep* **6**, 29454, doi:10.1038/srep29454 (2016).

106 Ueda, T. *et al.* EED mutants impair polycomb repressive complex 2 in myelodysplastic syndrome and related neoplasms. *Leukemia* **26**, 2557-2560, doi:10.1038/leu.2012.146 (2012).

107 Ueda, T. *et al.* Propagation of trimethylated H3K27 regulated by polycomb protein EED is required for embryogenesis, hematopoietic maintenance, and tumor suppression. *Proc Natl Acad Sci U S A* **113**, 10370-10375, doi:10.1073/pnas.1600070113 (2016).

108 Shi, J. *et al.* The Polycomb complex PRC2 supports aberrant self-renewal in a mouse model of MLL-AF9;Nras(G12D) acute myeloid leukemia. *Oncogene* **32**, 930-938, doi:10.1038/onc.2012.110 (2013).

109 Bagger, F. O., Kinalis, S. & Rapin, N. BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Res* **47**, D881-D885, doi:10.1093/nar/gky1076 (2019).

110 Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943-949, doi:10.1126/science.1256271 (2014).

111 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

112 Schnerch, D. *et al.* Cell cycle control in acute myeloid leukemia. *Am J Cancer Res* **2**, 508-528 (2012).

113 Zhou, H. & Xu, R. Leukemia stem cells: the root of chronic myeloid leukemia. *Protein Cell* **6**, 403-412, doi:10.1007/s13238-015-0143-7 (2015).

114 Abbas, T. & Dutta, A. p21 in cancer: intricate networks and multiple activities. *Nat Rev Cancer* **9**, 400-414, doi:10.1038/nrc2657 (2009).

115 Fiskus, W. *et al.* Combined epigenetic therapy with the histone methyltransferase EZH2 inhibitor 3-deazaneplanocin A and the histone deacetylase inhibitor panobinostat against human AML cells. *Blood* **114**, 2733-2743, doi:10.1182/blood-2009-03-213496 (2009).

116 Kojima, K., Konopleva, M., Tsao, T., Nakakuma, H. & Andreeff, M. Concomitant inhibition of Mdm2-p53 interaction and Aurora kinases activates the p53-dependent postmitotic checkpoints and synergistically induces p53-mediated mitochondrial apoptosis along with reduced endoreduplication in acute myelogenous leukemia. *Blood* **112**, 2886-2895, doi:10.1182/blood-2008-01-128611 (2008).

117 Sugimoto, K. *et al.* Frequent mutations in the p53 gene in human myeloid leukemia cell lines. *Blood* **79**, 2378-2383 (1992).

118 Stirewalt, D. L. *et al.* FLT3, RAS, and TP53 mutations in elderly patients with acute myeloid leukemia. *Blood* **97**, 3589-3595, doi:10.1182/blood.v97.11.3589 (2001).

119 Peterson, L. F., Yan, M. & Zhang, D. E. The p21Waf1 pathway is involved in blocking leukemogenesis by the t(8;21) fusion protein AML1-ETO. *Blood* **109**, 4392-4398, doi:10.1182/blood-2006-03-012575 (2007).

120 Casini, T. & Pelicci, P. G. A function of p21 during promyelocytic leukemia cell differentiation independent of CDK inhibition and cell cycle arrest. *Oncogene* **18**, 3235-3243, doi:10.1038/sj.onc.1202630 (1999).

121 Savona, M. & Talpaz, M. Getting to the stem of chronic myeloid leukaemia. *Nat Rev Cancer* **8**, 341-350, doi:10.1038/nrc2368 (2008).

122 Feinstein, E. *et al.* p53 in chronic myelogenous leukemia in acute phase. *Proc Natl Acad Sci U S A* **88**, 6293-6297, doi:10.1073/pnas.88.14.6293 (1991).

123 Ceballos, E. *et al.* c-Myc antagonizes the effect of p53 on apoptosis and p21WAF1 transactivation in K562 leukemia cells. *Oncogene* **19**, 2194-2204, doi:10.1038/sj.onc.1203541 (2000).

124 Jiang, H. *et al.* Induction of differentiation in human promyelocytic HL-60 leukemia cells activates p21,

WAF1/CIP1, expression in the absence of p53. *Oncogene* **9**, 3397-3406 (1994).

125 Radosevic, N., Delmer, A., Tang, R., Marie, J. P. & Ajchenbaum-Cymbalista, F. Cell cycle regulatory protein expression in fresh acute myeloid leukemia cells and after drug exposure. *Leukemia* **15**, 559-566 (2001).

126 Huang, M. J., Cheng, Y. C., Liu, C. R., Lin, S. & Liu, H. E. A small-molecule c-Myc inhibitor, 10058-F4, induces cell-cycle arrest, apoptosis, and myeloid differentiation of human acute myeloid leukemia. *Exp Hematol* **34**, 1480-1489, doi:10.1016/j.exphem.2006.06.019 (2006).

127 Gonzales-Aloy, E. *et al.* miR-101 suppresses the development of MLL-rearranged acute myeloid leukemia. *Haematologica*, doi:10.3324/haematol.2018.209437 (2019).

128 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).

129 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).

130 Ogino, T., Kobuchi, H., Fujita, H., Matsukawa, A. & Utsumi, K. Erythroid and megakaryocytic differentiation of K562 erythroleukemic cells by monochloramine. *Free Radic Res* **48**, 292-302, doi:10.3109/10715762.2013.865840 (2014).

131 Zhang, D., Cho, E. & Wong, J. A critical role for the co-repressor N-CoR in erythroid differentiation and heme synthesis. *Cell Res* **17**, 804-814, doi:10.1038/cr.2007.72 (2007).

132 Cai, J. *et al.* ZNF300 knockdown inhibits forced megakaryocytic differentiation by phorbol and erythrocytic differentiation by arabinofuranosyl cytidine in K562 cells. *PLoS One* **9**, e114768, doi:10.1371/journal.pone.0114768 (2014).

133 Butler, T. M., Ziemiecki, A. & Friis, R. R. Megakaryocytic differentiation of K562 cells is associated with changes in the cytoskeletal organization and the pattern of chromatographically distinct forms of phosphotyrosyl-specific protein phosphatases. *Cancer Res* **50**, 6323-6329 (1990).

134 Abraham, S. A. *et al.* Dual targeting of p53 and c-MYC selectively eliminates leukaemic stem cells. *Nature* **534**, 341-346, doi:10.1038/nature18288 (2016).
135 Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* **27**, 3769-3779, doi:10.1128/MCB.01432-06 (2007).
136 Chamberlain, S. J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **26**, 1496-1505, doi:10.1634/stemcells.2008-0102 (2008).
137 San Roman, A. K., Tovaglieri, A., Breault, D. T. & Shivdasani, R. A. Distinct Processes and Transcriptional Targets Underlie CDX2 Requirements in Intestinal Stem Cells and Differentiated Villus Cells. *Stem Cell Reports* **5**, 673-681, doi:10.1016/j.stemcr.2015.09.006 (2015).
138 Doynova, M. D., Markworth, J. F., Cameron-Smith, D., Vickers, M. H. & O'Sullivan, J. M. Linkages between changes in the 3D organization of the genome and transcription during myotube differentiation in vitro. *Skelet Muscle* **7**, 5, doi:10.1186/s13395-017-0122-1 (2017).
139 Dalerba, P. in *Clinical Oncology* 97-107 (2020).
140 Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat Med* **23**, 1124-1134, doi:10.1038/nm.4409 (2017).
141 Nowak, D., Stewart, D. & Koeffler, H. P. Differentiation therapy of leukemia: 3 decades of development. *Blood* **113**, 3655-3665, doi:10.1182/blood-2009-01-198911 (2009).
142 Shlush, L. I. *et al.* Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* **547**, 104-108, doi:10.1038/nature22993 (2017).
143 Laverdiere, I. *et al.* Leukemic stem cell signatures identify novel therapeutics targeting acute myeloid leukemia. *Blood Cancer J* **8**, 52, doi:10.1038/s41408-018-0087-2 (2018).
144 Passegue, E. & Weisman, I. L. Leukemic stem cells: where do they come from? *Stem Cell Rev* **1**, 181-188, doi:10.1385/SCR:1:3:181 (2005).
145 Wen, Y., Cai, J., Hou, Y., Huang, Z. & Wang, Z. Role of EZH2 in cancer stem cells: from biological insight to a therapeutic target. *Oncotarget* **8**, 37974-37990, doi:10.18632/oncotarget.16467 (2017).

146 Minucci, S. *et al.* PML-RAR induces promyelocytic leukemias with high efficiency following retroviral gene transfer into purified murine hematopoietic progenitors. *Blood* **100**, 2989-2995, doi:10.1182/blood-2001-11-0089 (2002).

147 Pasini, D. & Di Croce, L. Emerging roles for Polycomb proteins in cancer. *Curr Opin Genet Dev* **36**, 50-58, doi:10.1016/j.gde.2016.03.013 (2016).

148 Munoz-Alonso, M. J. *et al.* p21Cip1 and p27Kip1 induce distinct cell cycle effects and differentiation programs in myeloid leukemia cells. *J Biol Chem* **280**, 18120-18129, doi:10.1074/jbc.M500758200 (2005).

149 Fang, Y. *et al.* The ubiquitin-proteasome pathway plays essential roles in ATRA-induced leukemia cells G0/G1 phase arrest and transition into granulocytic differentiation. *Cancer Biol Ther* **10**, 1157-1167, doi:10.4161/cbt.10.11.13556 (2010).

150 Fischer, M., Grossmann, P., Padi, M. & DeCaprio, J. A. Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Res* **44**, 6070-6086, doi:10.1093/nar/gkw523 (2016).

151 Engeland, K. Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. *Cell Death Differ* **25**, 114-132, doi:10.1038/cdd.2017.172 (2018).

152 Law, J. C., Ritke, M. K., Yalowich, J. C., Leder, G. H. & Ferrell, R. E. Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. *Leuk Res* **17**, 1045-1050 (1993).

153 Lindsley, R. C. *et al.* Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood* **125**, 1367-1376, doi:10.1182/blood-2014-11-610543 (2015).

154 Safaei, S. *et al.* Double sword role of EZH2 in leukemia. *Biomed Pharmacother* **98**, 626-635, doi:10.1016/j.biopha.2017.12.059 (2018).

155 Sacha, T. Imatinib in chronic myeloid leukemia: an overview. *Mediterr J Hematol Infect Dis* **6**, e2014007, doi:10.4084/MJHID.2014.007 (2014).

156 Kalmanti, L. *et al.* Safety and efficacy of imatinib in CML over a period of 10 years: data from the randomized CML-

study IV. *Leukemia* **29**, 1123-1132, doi:10.1038/leu.2015.36 (2015).

157 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

158 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

159 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

160 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

161 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

162 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745, doi:10.1093/nar/gkv1189 (2016).

163 Stafoggia, M. *et al.* Spie charts, target plots, and radar plots for displaying comparative outcomes of health care. *J Clin Epidemiol* **64**, 770-778, doi:10.1016/j.jclinepi.2010.10.009 (2011).

164 Orlando, D. A. *et al.* Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep* **9**, 1163-1170, doi:10.1016/j.celrep.2014.10.018 (2014).

165 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1101/gr.229102 (2002).

ABREVIATIONS

Ac: Acetylation

ALL: Acute Lymphoid Leukemia

AML: Acute Myeloid Leukemia

AraC: Cytarabine: cytosine B-D-arabinofuranoside

CBX: Chromobox protein

CCNA2: Cyclin A2

CDK4: Cyclin Dependent Kinase 4

ChIP: Chromatin inmunoprecipitation

ChIP-seq: Chromatin inmunoprecipitation followed by sequencing

CLL: Chronic Lymphoid Leukemia

CML: Chronic Myeloid Leukemia

CXCR4: C-X-C Motif Chemokine Receptor 4

CyD1: Cyclin D1

DMSO: Dymethyl Sulphoxide

EED: Embryonic Ectoderm Development

EH: Extended Homology

EZH1/2: Enhancer of Zester 1 or 2

F: *Drosohpila*

FAB: French-American-British classification

FACS: Fluoresence-activated cell sorting

FC: Fold Change

FZD3: Frizzled Class Receptor 3

GATA2: GATA binding factor 2

GATA3: GATA binding factor 3

GO: Gene Ontology

GSEA: Gene Set Enrichment Analysis

GYPA: Glycophorine Alpha

H: Human

H3K4me3: Histone H3 trimethylated at lysine 4

H3K9me3: Histone H3 trimethylated at lysine 9

H3K27ac: Histone H3 acethylated at lysine 27

H3K27me3: Histone H3 trimethylated at lysine 27

H3K119ub: Histone H3 monoubiquitination at lysine 119

HCC: Hepatocellular carcinoma

HIC: Hypermethylated in Cancer

HOX: Homeobox

KD: knock-down

MDS: Myelodisplasic Syndrome

Me: Methylation

mESCs: mouse Embrionic Stem Cells

NEB: New England Biolabs

PCA: Principal Component Analysis

PCG: Polycomb Group of proteins

PCL: Polycomb like protein

PDGFRA: Platelet Derived Growth Factor Receptor Alpha

PHF19: PHD finger protein 19

PIP: PALI interaction with PRC2 domain

PMA: Phorbol Myristrate Acetate

PRC1: Polycomb Repressor Complex 1

PRC2: Polycomb Repressor Complex 2

PREs: Polycomb Response Elements

PTM: Post-transcriptional modifications

RNA-seq: RNA sequencing

RPKM: Reads per kilo base per million mapped reads

RT-PCR: Real time PCR

SD: standard deviation

SEM: standard error of the mean

shRNAs: short hairpins RNAs

SUZ12: Suppressor of Zester 12

TAL1: T-Cell Acute Lymphocytic Leukemia Protein 1

TES: Transcription Ending Site

TKI: Tyrosine Kinase Inhibitor

TPM: Transcript per kilo base Million

TSS: Transcription Starting Site

ACKNOWLEDGMENTS

Se hace muy difícil pensar en estos cuatro años y no acordarme de muchas de las personas que me han acompañado en este camino. Espero no dejarme a nadie, y si lo hago, que no me lo tengan en cuenta. Es raro escribir estas palabras cuando parece que sea ayer cuando empecé este largo proceso. Nunca olvidaré el primer día que llegué, viendo a todos aquellos científicos que luego se iban a convertir en compañeros, y que uno a uno me fueron dando sus ideas y el ánimo necesario para luego emprender este PhD. También reconociendo alguna que otra cara ya conocida (eh Anna ;D).

Primero de todo, me gustaría darle las gracias a Luciano por darme la oportunidad de formar parte de este grupo plagado de gente brillante. También por darme el soporte y *feedback* necesario en aquellos momentos donde lo necesitaba. Por tener siempre una palmada en el hombro y ver las cosas fáciles cuando los demás lo vemos complicado. Hacer este PhD es sin duda el proceso mas gratificante de mi vida y estoy seguro que no hubiese sido lo mismo en otro laboratorio. *Grazie mille.*

Tengo que hacer una mención especial a Pedro, que sin ser mi supervisor, siempre ha sido mucho más que eso. Sonrío al escribir estas palabras y recordar como recién llegados por la mañana, especialmente los Lunes, nos poníamos al día de política, del día a día en general y de futbol. Me quedo con

nuestro entender el Barça de forma diferente, y nuestros debates sobre toda la actualidad culé. Y las risas, muchas risas, pero también seriedad y concentración cuando tocaba. Gracias Por ese *cassette* de estopa que lo guardo como oro en paño. También por tener siempre un consejo personal y científico cuando más lo necesitaba. Por creer casi más que yo en mi proyecto en los momentos difíciles, y porque en tus momentos libres, con lo que ya tenías encima, siempre has sacado un rato para dedicármelo a mi. Muchas gracias amigo por todo y ya sabes donde estoy para un buen debate futbolístico, de la vida o lo que haga falta.

Al resto de laboratorio muchas gracias a todos y cada uno por ser como sois y por ayudarme siempre que lo necesitaba. Me siento un gran afortunado de haber podido conoceros y trabajar con vosotros.

A las más veteranas solo puedo tener buenas palabras para ellas. Arantxa, eres el alma de este laboratorio y sin ti el *PHF19 team* no existiría. Como dijo Piqué, contigo empezó todo. Me llevo todas las charlas, que son unas cuantas, entre ratones y cultivos celulares. Cecilia, muchas gracias por tu paciencia conmigo, por tus ‘’ Montolio, nos toca compras y *check-in list* ‘’ y por ayudarme y aguantarme con todas las ChIPs (si, incluso aquellas que aprendieron a nadar ;D).

Sergi, muchas gracias por tus consejos, risas de despacho tontas, y sobretodo por que ante mis preguntas (algunas absurdas) siempre decías aquello de: ‘’ ¿por qué no lo buscas tu? ’’ o me respondías con otra pregunta para que le diese vueltas. Sin duda, he aprendido muchísimo de ti. Enrique, gracias por toda la ayuda en los análisis, también por tener siempre un hueco para mi cuando te decía que solo eran 5 minutos y resultaban ser 30. Y por esas palmaditas en el hombro, y tu ver las cosas siempre de forma optimista.

A mis compañeros PhDs muchas gracias por estar ahí siempre, por reír conmigo, por aguantar mis bromas y mis cantes. Anna, ets més que una amiga, una confident i sempre estaré allà on em necessitis. Les hem passades de tots colors i sempre hem estat un al costat de l’altre recolzant-nos. Moltes gràcies. Paul, tu eres un chico brillante, me ha encantado compartir contigo todos estos años. Durante los años nos hemos ido conociendo más y de verdad que me llevo un amigo contigo. Eres muy grande. Al resto de PhDs (Aleksandra, Francesca, Isa, Ivano y Mar) que os habéis ido incorporando más tarde, me encantaría haber estado con vosotros todos los años, pues os siento como una pequeña gran familia y me lo he pasado en grande con todos vosotros. De todos y cada uno he aprendido algo y me llevo un ‘’cachito’’. Sois geniales.

Hay otros que ya se fueron del laboratorio pero que han formado parte también de mi tesis, con los cuales me encantaría haber estado mucho más tiempo con ellos, como Valerio (eh! En tu cara!), Alex (t'he trobat a faltar els últims dos anys), Gloria, Lluís, Malte, Miriam…A vosotros también muchas gracias por todo. Otro que tampoco está, pero del cual me acuerdo mucho es Joao, mi compañero de despacho, *o mais grande do mundo*.

También me gustaría agradecer a todos aquellos que han estado conmigo desde hace ya 10 años. Empezamos el camino juntos en la facultad de Biología y cada uno tomó su rumbo, pero siempre hemos estado juntos y lo seguimos estando. Habéis sido y sois muy importantes en mi vida, y os quiero agradecer a todos por haber estado ahí siempre. Con vosotros al fin del mundo.

No me quiero olvidar de mis amigos de toooooda la vida (acento Guille Gimenez). Seguramente sin el balonmano no hubiese podido hacer esta tesis, ni esa carrera, ni tantas otras cosas. Pero estoy seguro, que sin vuestra amistad todo esto no lo podría haber llevado a cabo. Siempre estáis ahí para lo que sea y os lo agradezco. Rajadas always!

Es imposible que deje de nombrar a algún que otro amigo/a, tanto del balonmano como de la vida en general, pero ellos ya saben quienes son, los de siempre y alguno/a más.

Siempre me han sacado una sonrisa y siempre han estado para echar una ‘’charleta’’ amanerada con alguna copa o cerveza cuando lo he necesitado.

Y por último, y no menos importante, gracias a mi familia. A mis padres, por todo. Por darme todo lo que tengo y hacer de mi lo que soy y lo que seré. Por ser mis referentes vitales. Me encantaría poder ser la mezcla perfecta de vosotros dos, y llegar a conseguir la mitad de lo que habéis logrado en vuestras vidas. Gracias también por bajarme de las nubes, aconsejarme y ayudarme siempre. A mi tío José y tía Nuria, porque si, porque sois los mejores y lo sabéis. Y a primos, por que no tengo hermanos pero os tengo a vosotros. También mis abuelos por que sin entender nada me han dado siempre su apoyo incondicional.

I a tu, a tu també, per ser una incondicional al meu costat, completar-me i fer-me feliç.

ANNEX

During my PhD I have collaborated to determine the role of PHF19 in the hematopoietic system, and resulting from that we have submitted to Cell Stem Cell the following publication:

The Polycomb associated factor PHF19 controls hematopoietic Stem Cell status and differentiation

Pedro Vizán, Arantxa Gutiérrez, Isabel Espejo, **Marc García-Montolio**, Martin Lange, Ana Carretero, Atefeh Lafzi, Holger Heyn, Marisa de Andrés, Enrique Blanco, Luciano Di Croce

Summary

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 show that genetic deletion of *Phf19* increases HSC identity. 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 the molecular level, PHF19 controls deposition of the histone repressive mark H3K27me3, which accumulates in blood lineage-specific genes. Our results provide novel insights into

how epigenetic mechanisms determine HSC identity, control differentiation, and are key for proper hematopoiesis.