Institute of Pharmaceutical Biology/DCAL, Goethe-University of Frankfurt, Biocenter, Frankfurt/Main, Germany

An alternative splice process renders the MLL protein either into a transcriptional activator or repressor

T. RÖSSLER, R. MARSCHALEK

Received November 23, 2012, accepted December 5, 2012

Prof. Dr. Rolf Marschalek, Institute of Pharmaceutical Biology/DCAL, Goethe-University of Frankfurt, Biocenter, Max-von-Laue-Str. 9, 60438 Frankfurt/Main, Germany
Rolf.Marschalek@em.uni-frankfurt.de

Dedicated to Prof. Dr. Theo Dingermann, Frankfurt, on the occasion of his 65th birthday.

Pharmazie 86: 601–607 (2013) *doi: 10.1691/ph.2013.6514*

The biological process of differentiation - from a fertilized egg to a human being - is a consecutive mechanism that leads to the establishment of tissue-specific gene expression, but also to a coordinated shut-down of all those genes that are not necessary for a given cell type. This process is accompanied by post-translational modifications of the chromatin (DNA methylation and covalent histone modifications), also termed the "epigenetic layer". All epigenetic processes are mediated by protein complexes that either mediate specific DNA methylation patterns, or modify nucleosomal proteins in a covalent fashion (acetylation, methylation, phosphorylation and ubiquitinylation). One important player involved in epigenetics is the MLL protein which represents a histone H3 methyltransferase. The *MLL* gene gained much attention because of its frequent genetic rearrangements, thereby creating oncogenic *MLL* fusion genes that cause acute leukemia in pediatric and adult patients. This article is summarizing certain functional aspects about MLL, but is mainly emphasizing on an alternative splice event within the PHD domain. This changes the biological properties of the MLL protein, thereby influencing its ability of being either a transcriptional activator or repressor.

1. Introduction

MLL biology has been fascinating - and still is - for many scientific groups around the world because this protein represents a molecular platform for the assembly of a large, nuclear protein complex that exhibits a very crucial function in eukaryotic cells: the MLL protein complex represents a histone methyltransferase (HMT) that di- and trimethylates lysine-4 residues of the histone H3 protein (Nakamura et al. 2002; Yokoyama et al. 2004). This $H3K4_{me2/3}$ signature can only be found in promotor regions of transcriptionally "active genes". Therefore, all genes that are transcribed in a given cell type need to attract the MLL complex (or a functionally equivalent enzyme) to their promotor region. This is normally being done by the help of transcription factors that are able to recruit the MLL complex. Direct binding of transcription factors like e.g. MENIN1, LEDGF, MYB and GATA-3 to the N-terminus of the MLL protein has already been demonstrated (Yokoyama et al. 2004; Yokoyama and Cleary, 2008; Jin et al. 2010; Nakata et al. 2010).

By contrast, other histone signatures, like e.g. $H3K9_{me2/3}$ or $H3K27_{me2/3}$, are mediating transcriptional silencing (Turner 2002). These signatures are deriving from nuclear protein complexes termed Polycomb repressor complexes I and II (PRCI, PRCII), respectively. The essential core components that execute the histone H3 methylation process are SUV39H1/2 (PRCI; $H3K9_{me}$) and EZH1/2 (PRCII; $H3K27_{me}$). Noteworthy, these repressor complexes can differ in their composition, depending on the differentiation state or the investigated tissue (Kuzmichev et al. 2005; Kato et al. 2007; Müller and Verrijzer, 2009; Schüttengruber and Cavalli, 2009; Bouyer et al. 2011; Zhang et al. 2011).

The PRCI complex - also termed the "Polycomb maintenance complex" - contains several polycomb-group proteins like e.g. PHC1-3 (*polyhomeotic homolog*), SCMH1/2 (*sex comb on midleg homolog*), RING1 (*ring finger protein 1*) and BMI1 (*polycomb ring finger oncogene*). The essential core protein SUV39H1 or 2 (*suppressor of variegation 3-9 homolog*) binds directly to HP1/CBX5, RB1, EVI1, HDAC1 and DNMT1, and thus, interacts also with components of the DNA methylation machinery.

The PRCII complex - also termed the "Polycomb initiator complex" - is composed by SUZ12 (*suppressor of zeste 12 homolog*), EED (*embryonic ectoderm development*) and EZH1 or 2 (*enhancer of zeste homolog*) in conjunction with RBBP4 (*retinoblastoma binding protein 4*), RBBP7 (*retinoblastoma binding protein 7*) and HDAC1/2. This complex is also able to directly interact with DNMT1, DNMT3A and DNMT3B. DNMT3A and DNMT3B are thought to mediate the initial ^{5me}CpG signatures. PRCII is already active in embryonic stem cells where it blocks all differentiation pathways (Lee et al. 2006), while the triumvirate consisting of NANOG, OCT4 and SOX2 regulates the transcription of stem cell maintenance genes (Boyer et al. 2005). Thus, stem cells often carry dual signatures in their promotor regions ($H3K4_{me}$ and $H3K27_{me}$; Cui et al. 2009). This bivalent signature is important as it allows to rapidly turn-on or turn-off genes during development (Weishaupt et al. 2010).

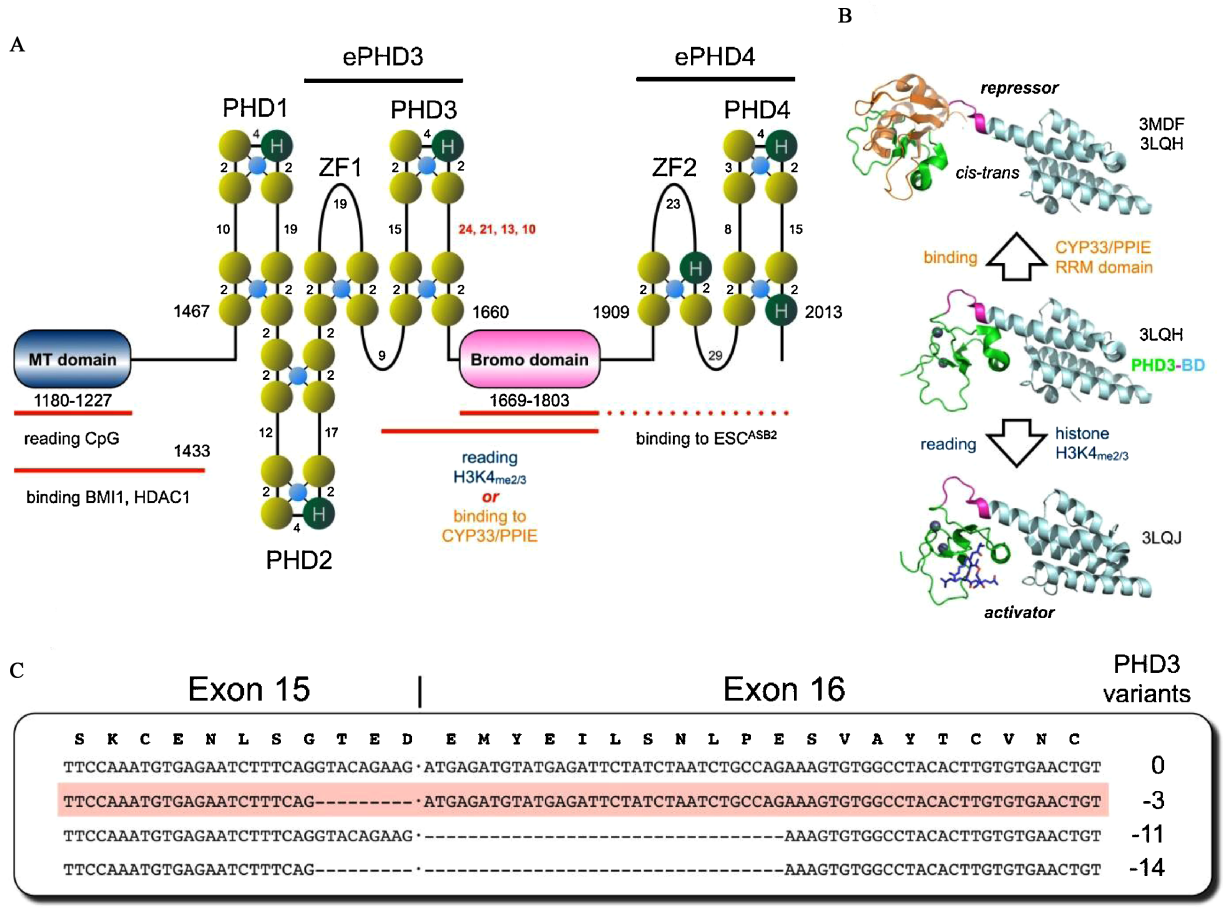


Fig. 1: The PHD domain of the human MLL protein. A. The complete structure of the PHD domain of the MLL protein is depicted. Yellow and green spheres represent cysteine- and histidine-residues, respectively. The numbers represent the amount of spacing amino acids. The smaller blue circles represent Zn^{2+} ions that are coordinating the structure of the PHD or zinc finger domains. The PHD domain is flanked by the MT-domain (depicted in blue; aa 1180-1227) and the BD-domain (depicted in pink; aa 1669-1803). The PHD domain is comprised by PHD1 (aa 1467-1512), PHD2 (aa 1515-1563), ZF1 (aa 1566-1592), PHD3 (aa 1602-1660), ZF2 (aa 1909-1929) and PHD4 (aa 1969-2013). ZF1 and PHD3 are termed also ePHD3, while ZF2 and PHD4 are termed ePHD4. The MT domain is known to bind hemi-methylated CpG dinucleotide motifs. This region is also important for binding BMI1, HPC2 and HDAC1. PHD3-BD is known to be involved in $H3K4_{me3}$ reading, or, CYP33/PPIE binding. The BD and ePHD4 region was recently shown to be involved in binding the E3-ligase ASB2 that controls the half-life of MLL. B. The protein structure of PHD3-BD domain is shown (3LQH). PHD3 is shown in green, the linker region in pink, the BD domain in cyan blue. This domain is able to bind to trimethylated lysine-4 residues of histone H3 protein (3LQJ). When the RRM domain of the CYP33/PPIE protein binds to PHD3, the PHD domain becomes isomerized which allow efficient binding of the RRM domain (organge; 3MDF). Thus isomerization of PHD3 is a switch between histone H3 reading and CYP33 binding.C. The alternatively spliced products of PHD3. All splice events occur at the MLL exon 15/16 junctions and lead to variants of the PHD3 domain (0, –3, –11, –14).

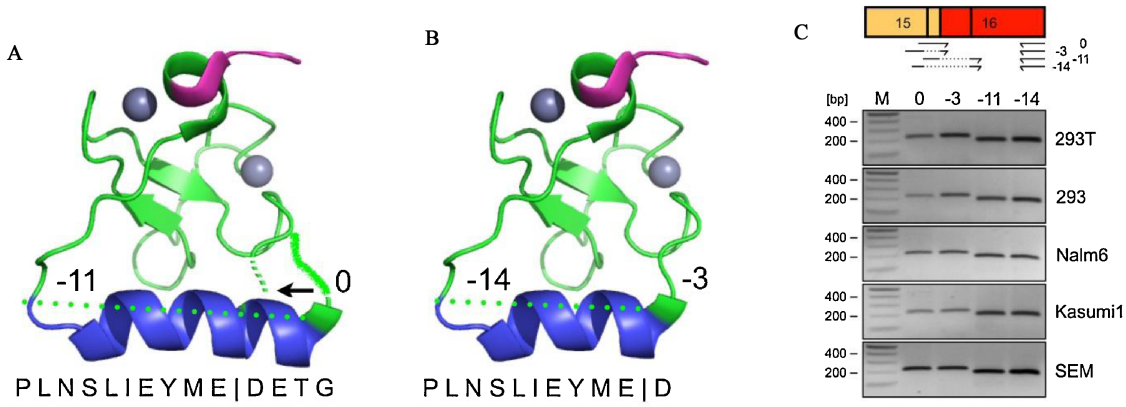
Therefore, 3 distinct histone H3 methylation signatures are basically decrypting the epigenetic information: H3K4 methylation is activating the transcriptional process, while H3K27 and H3K9 methylation leads to a transcriptional shut-down. The H3K9 methylation signature is recognized by *Heterochromatin protein* 1 (HP1/CBX5) that helps to form inactive heterochromatin by condensing the chromatin structure (Eskeland et al. 2007). HP1/CBX5 binds also directly to MBD1, MECP2 and DNMT3A/3L.

The MLL protein is encoded by a gene that localizes at chromosome 11q23. This gene exhibits 37 exons and has a length of about 90 kb (Nilson et al. 1996). The longest mature mRNA transcript of the *MLL* gene has a length of 16,693 nucleotides and exhibits an open reading frame of 12,015 nucleotides. Thus, a full-length MLL protein is encoded by 4,005 amino acids. However, about 75% of all *MLL* transcripts are missing exon 2 (Meyer et al. 2006) which has a length of only 99 nucleotides (Nilson et al. 1996). Thus, most public databases are displaying the *MLL* gene only with 36 exons, encoding an MLL protein with only 3,972 amino acids. Here, we are following the *MLL* exon/intron nomenclature (37 exons) that has been established by our own group already in 1996 (Nilson et al. 1996).

Relatively little is known about alternatively spliced transcripts deriving from the human *MLL* gene. An obvious alternative splice event affects *MLL* exon 11 (Mbangkollo et al. 1995) which was frequently observed by many groups, mainly because this exon lies within the breakpoint cluster region and could be skipped during transcription and the subsequent splice process when *MLL* is translocated to a translocation partner gene (TPG). Therefore, diagnostic RT-PCR analyses result frequently in two alternative mRNA species that span the breakpoint on the der(11) chromosome.

The first alternatively spliced mRNA ever described and not associated with chromosomal translocations was found for the third PHD finger domain of the MLL protein, where 9 nucleotides could be missing at the exon 15/16 junction (Ma et al. 1993; Mbangkollo et al. 1995). Later, a more extensive study was performed that identified additional splice variants of *MLL* (Nam et al. 1996), however, most of them displayed splice variations concerning *MLL* exon 4 and *MLL* exon 28, where shorter variants have been cloned and sequenced after RT-PCR amplification. Finally, it should be mentioned that *MLL* transcripts can be terminated within *MLL* intron 8 and intron 9 at cryptic poly-A sites (Kowarz et al. 2011), while a gene-internal promotor resides upstream of *MLL* exon 12 that allows the transcription of only *MLL* exons 12-37, coding for the MLL* protein (Scharf et al. 2007).

This paper is particularly focussing on the MLL PHD-domain and its splice variants. The amino acid sequence of the MLL PHD-domain predicts 3 consecutive PHD-domain that are dis-



D PHD3 x PHD3

	0	-3	-11	-14	RRM	PPIE
0	+++	+++	+++	+++	-	-
-3	+++	+	+	+	++	+++
-11	+++	+	++	++	-	-
-14	+++	+	++	++	-	-
RRM	-	++	-	-	-	-
PPIE	-	+++	-	-	-	-

E PHD1-3 x PHD3

	0	-3	-11	-14
PHD1-3	++	++	++	++
ΔPHD1-3	++	++	++	++

	RRM	PPIE	PHD1-3	ΔPHD1-3
PHD1-3	+	++	(+)	(+)
ΔPHD1-3	(+)	++	-	-

Fig. 2: The PHD3 domain variants. A. The PHD3 domain is depicted. The blue helical structure is changed by the alternative splice processes. This helical domain is important to form contacts to the BD-domain. Pink: linke region to the BD-domain. Grey: zinc ions. The -11 splice variant will cause a complete deletion of the blue helical structure. B. The -3 splice variant is shown. The -14 splice variant will also result in the deletion of the blue helical structure. C. Semiquantitative RT-PCR experiments with primer sets that exclusively detect only the 0, -3, -11 and -14 splice variants. As shown for 5 different cell lines, the transcription rate of all 4 splice variants is nearly equal in the investigated cell lines. D. Yeast-2-hybrid mating experiments for PHD3 and CYP33/PPIE. Growth of diploid yeast cells after mating is shown by either "-" or "+" to "+++". For CYP33/PPIE either the RRM domain alone of the full-length protein was used. E. Yeast-2-hybrid mating experiments for PHD1-3, ΔPHD1-3 and CYP33/PPIE. Growth of diploid yeast cells after mating is shown by either "-" or "+" to "+++". For CYP33/PPIE either the RRM domain alone of the full-length protein was used.

rupted by a normal zinc finger (ZF1). This PHD1-3-domain is followed by a Bromo-domain (BD) and another zinc finger (ZF2) and a final PHD4-domain. Both the ZF1/PHD3- and the ZF2/PHD4-domains are also termed 'ePHD3' and 'ePHD4', respectively. The Bromo-domain of MLL has no capacity for binding acetylated histone but is responsible for enhancing the functions of the PHD3-domain. The 2-dimensional structure of the complete domain, with its typical cysteine and histidine residues, is schematically depicted in Fig. 1A.

The MLL PHD1-3-domain was first described to confer oligomerization (Linder et al. 2000). Later, the PHD3-domain was described to bind to the CYP33/PPIE protein (Fair et al. 2001; Xia et al. 2003) and to recognize methylated lysine-4 residues of histone H3 (Chang et al. 2010). Binding of the PHD3-domain to $H3K4_{me2/3}$ peptides was described to be greatly enhanced by the adjacent Bromo-domain (Wang et al. 2010); only a tiny linker of 8 amino acids (N-TERHPAEW-C; depicted in pink in Fig. 1B) separates both domains. Binding of CYP33/PPIE has a dramatic effect on the biological activity of the PHD3-Bromo-domain, as CYP33/PPIE represents a prolyle peptidyl isomerase (PPIase) that changes the conformation of co-bound proteins. CYP33/PPIE performs a *cis-trans* isomerization of the proline-1665 residue that localizes in the tiny linker region between the PHD3- (aa 1602-1660) and the Bromo-domain (aa 1669-1803). This *cis-to-trans* conversion now allows efficient binding of CYP33/PPIE to PHD3-Bromo-domain, which is mutual exclusive with $H3K4_{me2/3}$ binding by the PHD3-domain. The reason became clear from the structural analysis: the RRM-domain of CYP33/PPIE (RNA recognition motif; depicted in orange in Fig. 1B) binds to the PHD3-domain in a way, that binding to histone H3 is inhibited (reviewed in Grow and Wysocka 2010). This way, the human MLL protein switches from a transcriptional activator/maintenance factor to a transcriptional repressor.

Noteworthy, the same Bromo-domain - enhanced by the ePHD4-domain - triggers a specific degradation process that was identified recently (Wang et al. 2012). Binding of ASB2 to the Bromo-domain causes a rapid degradation of the MLL protein. Thus, different functions and the steady-state stability of MLL are controlled by this particular domain.

Here, we present data addressing the splice junction between *MLL* exon 15 and 16. This splice junction is important for the structure and function of the PHD3-domain of the MLL protein which turned out to be an important molecular switch that enables the MLL complex to be either a "transcriptional activator" or "repressor".

2. Investigations and results

During our studies of the functional properties of the MLL PHD1-3-domain (amino acids 1467-1660) we identified 4 different mRNA species that differ only by a few nucleotides at the splice junction between MLL exon 15 and 16 (Fig. 1C). Besides the normal transcript (0) and the splice variant that misses 9 nucleotides from the end of *MLL* exon 15 (-3 variant), an additional splice variant was identified that misses either the first 33 nucleotides of *MLL* exon 16 (-11 variant), or 9 and 33 nucleotides are missing at both exon junctions (-14 variant). Consequently, all these splice variants change the conformation of the PHD3-domain, in particular the alpha-helix of the PHD3-domain (depicted in blue in Fig. 2A, B). This alpha helix is known to be important for the protein-protein interaction with the adjacent Bromo-domain (see Fig. 1B, the Bromo-domain is depicted in cyan blue), and thus, influences

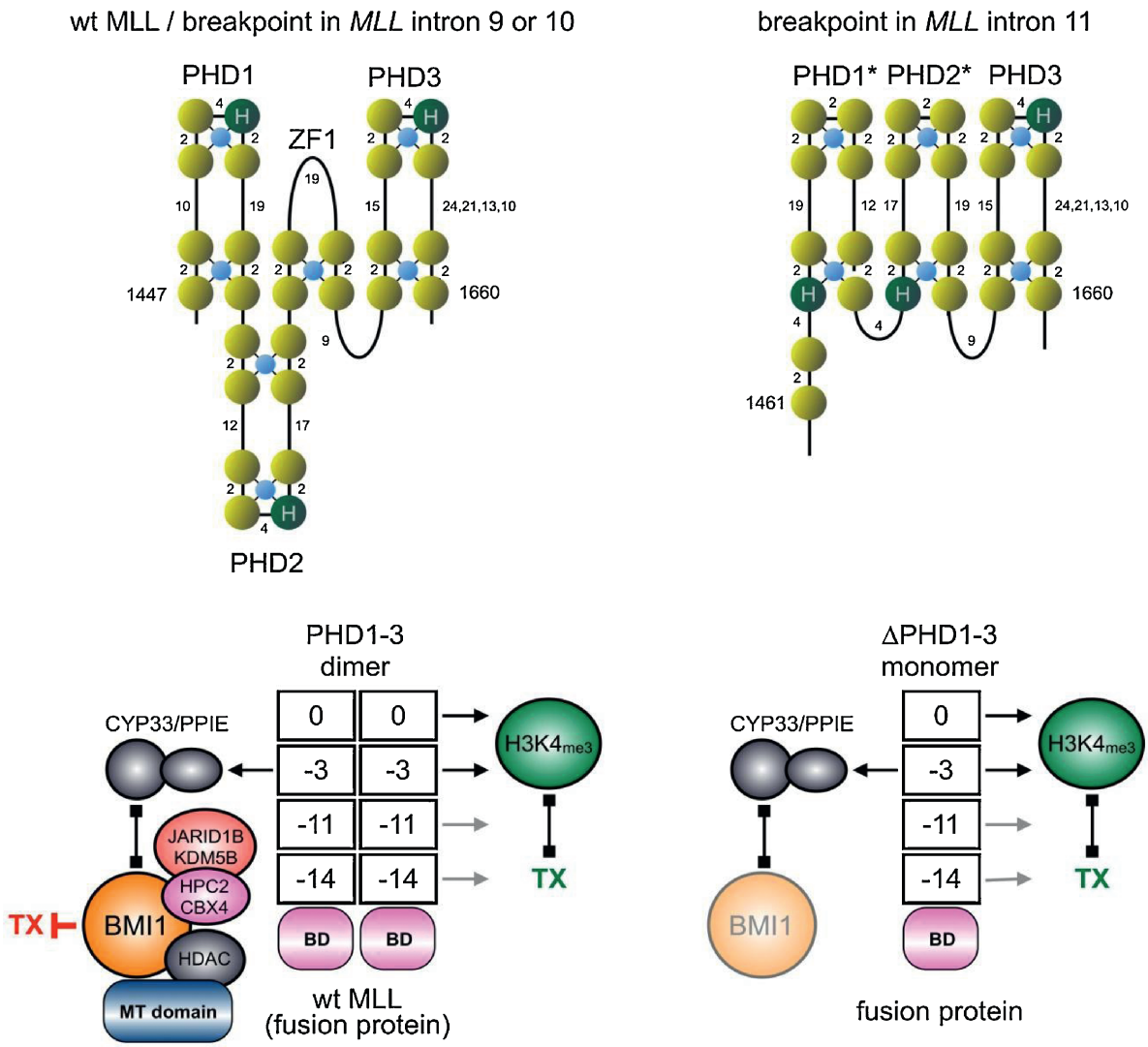


Fig. 3: Breakpoints within the *MLL* gene influence the PHD domain structure. A. The PHD1-3 domain of the MLL protein, or, a reciprocal MLL fusion protein with a chromosomal breakpoint within MLL intron 9 and 10. B. The ΔPHD1-3 domain of a reciprocal MLL fusion protein with a chromosomal breakpoint within MLL intron 11. Due to the alternative folding, the properties of this domain may be compromised. Only PHD3 seems to be intact, but ZF2 is lost. C. According to our mating data, all 4 PHD3 splice variants are able to dimerize with other PHD3 domain variants (10 combinations possible due to the symmetry displayed in Table 2D). While the 0 and -3 variant of the PHD3-domain should be able to bind directly to $H3K4_{me3}$ signatures of nucleosomes, only the -3 splice variant is capable of binding to the RRM domain of CYP33/PPIE protein. This allows binding to BMI1, HPC2/CBX4, JARID1B/KDM5B and HDAC1 which dock to the MT domain and repress transcription (TX). D. The dimerization of ΔPHD1-3 is compromised, while binding to CYP33/PPIE is not affected. Whether BMI1 is still able to bind is unclear, because of the missing MT-domain in reciprocal MLL fusion proteins.

also CYP33/PPIE binding (see Fig. 1B, orange domain: RRM-domain of CYP33/PPIE) and binding to histone tails containing the epigenetic $H3K4_{me2/3}$ modification (see Fig. 1B**,** blue structure: $H3K4_{me3}$). This alpha-helix has either a length of 12 amino acids (0 variant), 9 amino acids (-3 variant) or is completely absent in the -11 and -14 variants (Fig. 2A,B: green dotted line). Therefore, we first performed RT-PCR experiments (30 cycles) to ascertain a semi-quantitative analysis. However, all 4 *MLL* transcript variants (0: 213 bp; -3: 218 bp; -11: 182 bp; -14: 183 bp) are nearly equally transcribed in all investigated cell lines (HEK293T, HEK293, Nalm-6, Kasumi-1, SEM; Fig. 2C). Thus, their relative abundance should be quite equal in different tissues.

We next cloned the PHD3-domain in all four variations into the yeast expression vectors pGADT7 and pGBKT7 that can be used for yeast-2-hybrid mating experiments. In addition, we also cloned the RRM-domain of the CYP33/PPIE protein (aa 2-82) as well as the full-length CYP33/PPIE protein (aa 1-301). Stably transfected yeast clones were mated into diploid cells, and protein interactions (= cell growth) were studied on corresponding selective drop-out media. We have tested all combinations to investigate (a) dimerization capacity and (b) binding to the RRM-domain of full-length protein of CYP33/PPIE. As shown in Fig. 2D, the 4 different PHD3 variants exert quite different functions.

The most important result of these experiments was that only the -3 variant of the PHD3-domain was capable of binding to the RRM-domain or to the full-length CYP33/PPIE protein. Noteworthy, the -3 PHD3 variant was exactly the protein that has been co-crystalized with the BD-domain, the RRM-domain or the modified Histone H3 lysine-4 residue ($H3K4_{me3}$; Wang et al. 2010; Chang et al. 2010). All other PHD3 variants that have been tested were unable to directly interact with the RRM-domain of the CYP33/PPIE protein. Moreover, the -3 PHD3 variant displayed a weaker homodimerization capacity, suggesting that it may preferentially interact with the other 3 possible PHD3 variants that display no RRM-domain binding capacity. This may indicate that MLL dimerization is only favored when the PHD3

domains are alternatively spliced. In principle there are 10 different combinations possible of which only 4 are able to directly bind to the CYP33/PPIE protein (marked in light red in Fig. 2D). Next, we investigated the complete PHD1-3-domain (-3 variant; aa 1467-1660) by analyzing its interaction capacity with all 4 PHD3-domain variants. This analysis revealed no measurable differences (Fig. 2E, upper panel). We also tested the -3 variant of a ΔPHD1-3-domain that misses the first 11 amino acids (aa 1478-1660). This shorter protein mimics a chromosomal translocation that had occured within intron 11 of the *MLL* gene. This is the predominant recombination site diagnosed in infant acute leukemia patients that bear a t(4;11) translocation. We assumed that such a ΔPHD1-3-domain - without the first 2 cysteine residues of PHD1 - may result in folding deficiencies of the PHD1 and PHD2 structure, while PHD3 should not be affected. As expected, the partial deletion of PHD1 amino acids sequences had no effect on binding the four PHD3-domain variants (Fig. 2E, upper panel). However, we saw a reduced capability of binding to the RRM-domain of CYP33/PPIE protein (Fig. 2E, lower panel), and more importantly, an inability to bind either to another PHD1-3 domain or another ΔPHD1-3 domain. Thus, the deletion of 11 amino acids deriving from the PHD1 domain was sufficient to severely compromize the dimerization function of the PHD1-3-domain.

3. Discussion

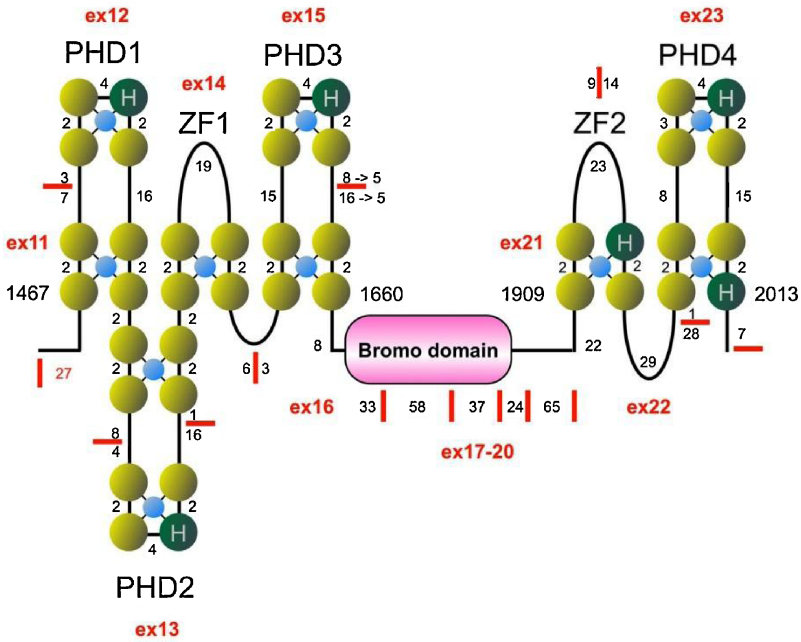
Here we present for the first time experiments that aimed to unravel the functional consequences of an alternative splice process that create different variants of the PHD3-domain of the human MLL protein. The splice events all occur at the *MLL* exon 15/16 junction and lead to 4 protein variants that strongly affect the conformation of this domain, as those protein variants eliminate an alpha-helix, depicted in blue in Fig. 2A and 2B. This helix was shown to be important, because it represents the docking site of the adjacent BD-domain. We termed the investigated PHD3 splice variants according to their missing amino acids either the "0", "–3", "–11" or "–14" variant. Only one of these 4 splice variants, the -3 variant, result in a PHD3-domain that was capable of binding to the CYP33/PPIE protein or the RRM-domain thereof.

Based on our yeast-2-hybrid mating studies, all 4 splice variants display an equal capacity for dimerization when analyzing the PHD1-3-domain *in toto*, however, there are slight differences when focussing only on a single PHD3-domain. Our interaction studies revealed that the CYP33/PPIE binding -3 variant of PHD3 has the lowest capacity of dimerizing to the other PHD3 variants (0, -11 and -14). This may indicate that the PHD1- and PHD2-domain are functionally stabilizing the dimerization process mediated by PHD3. We also observed that the full-length CYP33/PPIE protein binds much stronger to the PHD1-3-domain than the RRM-domain alone. This may be another hint that there are additional interactions between other protein domains of CYP33/PPIE and PHD1 or PHD1-2.

A striking result of our studies is the functional consequence of disrupting the *MLL* gene within intron 11. This kind of genetic rearrangement is frequently identified in infant Acute Lymphoblastic Leukemia (ALL), childhood ALL and therapy-induced Acute Myeloid Leukemia (t-AML), while adult t(4;11) ALL patients display a tendency for chromosomal translocations preferentially to occur within *MLL* introns 9 and 10 (Reichel et al. 2001). All these patients loose the first 11 amino acids of their PHD1 domain (encoded by *MLL* exon 11; see Fig. 4). This may cause either a non-functional PHD1 domain, or alternatively, cause an alternative folding process that inflicts the whole PHD1-3-domain structure. As outlined in Fig. 3A vs. Fig. 3B, such an alternative fold may result in similar fold of PHD1 and PHD2, but will eliminate ZF1. Such a misfolding could in principle result either in (1) a different protein structure, (2) different protein stability (higher turn-over), or (3) a functional impairment of the PHD1-3-domain. We observed in our ΔPHD1-3 experiments a reduced affinity to the RRM-domain of CYP33/PPIE, and moreover, to a complete inability to form protein dimers. This could be of functional relevance because this disables the reciprocal AF4-MLL fusion protein, shown to initiate the development of proB ALL in our recently established mouse model (Bursen et al. 2010), to form protein dimers. Thus, this particular reciprocal MLL fusion protein will be unable to form functional dimers with another AF4-MLL fusion protein or with a wildtype MLL protein. Depending on which splice variant is actually produced in a leukemic cell, binding to CYP33/PPIE will be still possible, but there are are only 4 instead of 10 functional combinations remaining (compare Fig. 3C with Fig. 3D). Binding of CYP33/PPIE has an important consequence for the wildtype MLL protein. The isomerization of the PHD3-domain leads to the binding of the RRM-domain of the CYP33/PPIE protein. This converts the CYP33/PPIE protein from an RNA-binding protein to an mediator for protein interactions. This allows the MLL protein (or MLL fusion protein) to interact with BMI1 and other proteins related to the "Polycomb maintenance complex" (CtBP, HPC2/CBX4 and HDAC1; Xia et al. 2003). HPC2/CBX4 represents a SUMO E3-ligase that sumoylates several target proteins (e.g. CtBP; Agrawal et al. 2008), and was shown to bind directly to JARID1B/KDM5B ($H3K4_{me1-3}$ demethylase; Zhou et al. 2009). The latter protein inactivates gene loci by erasing their necessary histone $H3K4_{me2/3}$ signature. All this explains how the MLL protein is being converted from a transcriptional maintenance factor into an inhibitory protein for gene transcription. This particular and switchable feature is important and explains why gene expression profiling of homozygous *Mll* k.o. vs. wildtype cells not only caused a decrease in gene transcription, but displayed twice as many genes that displayed an increase of gene transcription (Schraets et al. 2003). Thus, MLL has a bivalent role in transcriptional maintenance: activation and repression. This bivalent role can be explained by the alternative splice process that was investigated here, because only the -3 variant of PHD3 confers the possibility to bind to CYP33/PPIE, and thus, to the inhibitory proteins BMI1, HPC2/CBX4, JARID1B/KDM5B and HDAC1 (Fig. 3C).

The functional consequences for reciprocal MLL fusion proteins have to be investigated in the future. However, reciprocal MLL fusion protein - exhibiting either a normal or altered PHD1-3 domain - exhibit *per se* no MT-domain which was shown to be important for binding BMI1 and other cooperating proteins (Fig. 3D). Thus, the question remains whether reciprocal MLL fusion proteins still have the property to bind to these Polycomb repressor proteins, or not. If not, it would explain the oncogenic effects deriving from the AF4-MLL (. . .a prototypic reciprocal MLL fusion protein) which have been experimentally shown *in vitro* and *in vivo* (Bursen et al. 2004; Gaussmann et al. 2007; Bursen et al. 2010).

Also of importance is the fact that the recently identified gene-internal promotor within the MLL protein (Scharf et al. 2007), necessary to drive expression of the MLL C-terminus in those leukemia patients that do not bear a functional reciprocal fusion (Meyer et al. 2009), will always result in an MLL* protein that do neither exhibit the PHD1-3- nor the BD-domain. This N-terminal truncated MLL protein was shown to start with a *bona fide* AUG start codon within *MLL* exon 18, and thus, encodes only the ZF2 and PHD4 domain. This novel N-terminal region of the MLL* protein may have lost its ability to bind to the E3 ligase ASB2. Thus, important biological properties of MLL*,



MLL length in…

exon	nt	aa
11	114	38
12	147	49
13	96	32
14	121	40,33
15	123	41
16	185	61,67
17	174	58
18	111	37
19	74	24,67
20	194	64,67
21	107	35,67
22	138	46
23	159	53
∑	**1743**	**581**

Fig. 4: MLL exons encoding the PHD/zinc finger domains. Again, the complete structure of the PHD domain of the MLL protein is depicted. Yellow and green spheres represent cysteine- and histidine-residues, respectively. The smaller blue circles represent Zn^{2+} ions that are coordinating the structure of the PHD or zinc finger domains. Red lines are now indicating the exon/exon boundaries. Numbers indicate the amino acids spacing either cysteine/histidine residues and the amount of amino acids encoded by the single exons. Between the first three and the fourth PHD domain, the BD-domain is localized and depicted in pink (aa 1669-1803). The table on the right side is summarizing all relevant information about *MLL* exons 11-23 (length of each exon in nucleotides (nt) and coding capacity in amino acids (aa)).

like e.g. protein stability, should be investigated in the future to shed light on the putative functions of this not yet investigated variant of the MLL protein.

In summary, we provide important new information about a novel regulatory mechanism of the MLL protein that has not been analyzed in such detail before. An alternative splice process seems to convert the MLL protein from a positive modulator of gene transcription into a transcriptional repressor complex. This new knowledge will allow new research activities in order to unravel the last mysteries of the human MLL protein.

4. Experimental

4.1. Cell culture

The Nalm-6, Kasumi-1 and SEM cell lines were cultured in RPMI 1640 containing 10% fetal calf serum. HEK293T and HEK293 cells were maintained in DMEM supplemented with 10% fetal calf serum. All media were supplemented with 1% L-glutamine and 1% Pen/Strep.

4.2. Total RNA extraction and reverse transcription

RNA was prepared from all cell lines by using the Qiagen RNeasy Mini Kit (Qiagen, Germany). One µg RNA was reverse transcribed using hexamer primers in a total volume of 25 µl. Final cDNA synthesis was diluted to 50 µl using sterile water. In addition, all isolated RNAs were directly tested in PCR reactions to exclude any contamination with genomic DNA (data not shown). Five µl of each cDNA was used as template in 50 µl PCR reactions throughout all experiments.

4.3. PCR experiments

RT-PCR experiments were performed with Taq-Polymerase (Promega) using the forward primers 5'-CGCGGATCCCAGGTACAGAAGAT GAGATGTATGAGAT-3' for the 0, 5'-CGCGGATCCTGAGAATCTTTC AGATGAGATGTATGAGAT-3' for the -3, 5'-CGCGGATCCAGGT ACAGAAGAAAGTGTGGCCTAC-3' for the -11 and 5'-CGCGGATCCAGAATCTTTCAGAAAGTGTGGCCTAC-3' for the -14 PHD3 variant. The reverse primer for all reactions was 5'-CGGCTCGAGTTAGCCGGTAGCGTAGCAAATGG-3'.

4.4. Yeast-2-hybrid mating experiments

All yeast-2-hybrid screens of this study were conducted using the MATCHMAKER III Two-Hybrid System (Clontech) according to the manufacturers recommendations. The system provides the two vectors pGBKT7 (bait) and pGADT7 (prey).

All described constructs (PHD3, PHD1-3, ΔPHD1-3, RRM-domain and CYP33/PPIE) were cloned into both vector backbone. pGBKT7 plasmids were transformed into the MATa yeast strain AH109 whereas the MATα strain Y187 was host strain for pGADT7 plasmids using an LiAc transformation procedure. Transformed cells were plated on appropriate SD agar plates (–Trp for pGBKT7 and –Leu for pGADT7) to select for transformed yeast clones. Incubation time was usually 4 to 5 days at 30 °C. Yeast mating was performed with corresponding clones of both transformed yeast strains (Mata with MATα) by using microtiter plates. Mated cultures were put on Leu/Trp drop-out SD agar plates (2x drop-out medium) to select for diploid cells that bear both plasmids. Growth of diploid cells on agar dishes was usually 2 days at 30 °C. Selected colonies were picked, resuspended in sterile water, and incubated on Leu/Trp/His/Ade drop-out SD agar plates (4x drop-out medium) and incubated for 5 days at 30 °C. Cell growth was visually inspected and categorized into "no growth" ("-"), very weak growth ("+") sufficient growth ("++") and very good growth ("+++"). Importantly, control experiments were performed by mating vector-containing yeast clones with corresponding yeast clones bearing empty vector controls, and by switching the vector backbones of tested protein interactions. Thus all mentioned protein interactions have been double-cross-checked.

4.5. In silico work

The MLL protein structure of 4005 amino acids was used according to the original publication by Nilson et al. (1996). The structural analyis was performed with the PDB files 2KYU (Park et al. 2010), 3LQJ (Wang et al. 2010) and 3LQH and 3MDF (Hom et al. 2010) and retrieved from the PDB database (http://www.rcsb.org/pdb/). For analysis and modelling the open source program MacPyMol was used.

References

Agrawal N, Banerjee R (2008) Human polycomb 2 protein Is a SUMO E3 ligase and alleviates substrate-induced inhibition of cystathionine β-synthase sumoylation. PLoS ONE 3: e4032.

Bouyer D, Roudier F, Heese M, Andersen ED, Gey D, Nowack MK, Goodrich J, Renou JP, Grini PE, Colot V, Schnittger A (2011) Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. PLoS Genet 7: e1002014.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA,

Jaenisch R, Young RA (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122: 947–56.

Bursen A, Moritz S, Gaussmann A, Moritz S, Dingermann T, Marschalek R (2004) Interaction of AF4 wild-type and AF4-MLL fusion protein with SIAH proteins: indication for t(4;11) pathobiology? Oncogene 23: 6237–6249.

Bursen A, Schwabe K, Rüster B, Henschler R, Ruthardt M, Dingermann T, Marschalek R (2010) The AF4-MLL fusion protein is capable of inducing ALL in mice without requirement of MLL-AF4. Blood 115: 3570–3579.

Caldas C, So CW, MacGregor A, Ford AM, McDonald B, Chan LC, Wiedemann LM (1998) Exon scrambling of MLL transcripts occur commonly and mimic partial genomic duplication of the gene. Gene 208: 167–176.

Chang PY, Hom RA, Musselman CA, Zhu L, Kuo A, Gozani O, Kutateladze TG, Cleary ML (2010) Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription. J Mol Biol 400: 137–144.

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell 4: 80–93.

Eskeland R, Eberharter A, Imhof A (2007) HP1 binding to chromatin methylated at H3K9 is enhanced by auxiliary factors. Mol Cell Biol 27: 453–465.

Fair K, Anderson M, Bulanova E, Mi H, Tropschug M, Diaz MO (2001) Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. Mol Cell Biol 21: 3589–3597.

Gaussmann A, Wenger T, Eberle I, Bursen A, Bracharz S, Herr I, Dingermann T, Marschalek R (2007) Combined effects of the two reciprocal t(4;11) fusion proteins MLL-AF4 and AF4-MLL confer resistance to apoptosis, cell cycling capacity and growth transformation. Oncogene 26: 3352–3363.

Grow EJ, Wysocka J (2010) Flipping MLL1's switch one proline at a time. Cell 141: 1108–1010.

Hom RA, Chang PY, Roy S, Musselman CA, Glass KC, Selezneva AI, Gozani O, Ismagilov RF, Cleary ML, Kutateladze TG (2010) Molecular mechanism of MLL PHD3 and RNA recognition by the Cyp33 RRM domain. J Mol Biol 400: 145–154.

Jin S, Zhao H, Yi Y, Nakata Y, Kalota A, Gewirtz AM (2010) c-Myb binds MLL through menin in human leukemia cells and is an important driver of MLL-associated leukemogenesis. J Clin Invest 120: 593–606.

Kato Y, Koseki H, Vidal M, Nakauchi H, Iwama A (2007) Unique composition of polycomb repressive complex 1 in hematopoietic stem cells. Int J Hematol 85: 179–181.

Kowarz E, Merkens J, Karas M, Dingermann T, Marschalek R (2011) Premature transcript termination, trans-splicing and DNA repair: a vicious path to cancer. Am J Blood Res 1: 1–12.

Kuzmichev A, Margueron R, Vaquero A, Preissner TS, Scher M, Kirmizis A, Ouyang X, Brockdorff N, Abate-Shen C, Farnham P, Reinberg D (2005) Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. Proc Natl Acad Sci USA 102: 1859–1864.

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125: 301–313.

Linder B, Newman R, Jones LK, Debernardi S, Young BD, Freemont P, Verrijzer CP, Saha V (2000) Biochemical analyses of the AF10 protein: the extended LAP/PHD-finger mediates oligomerisation. J Mol Biol 299: 369-378.

Ma Q, Alder H, Nelson KK, Chatterjee D, Gu Y, Nakamura T, Canaani E, Croce CM, Siracusa LD, Buchberg AM. (1993) Analysis of the murine All-1 gene reveals conserved domains with human ALL-1 and identifies a motif shared with DNA methyltransferases. Proc Natl Acad Sci U S A.: 6350–6354.

Mbangkollo D, Burnett R, McCabe N, Thirman M, Gill H, Yu H, Rowley JD, Diaz MO (1995) The human MLL gene: nucleotide sequence, homology to the Drosophila trx zinc-finger domain, and alternative splicing. DNA Cell Biol 14: 475–483.

Meyer C, Kowarz E, Schneider B, Oehm C, Klingebiel T, Dingermann T, Marschalek R (2006) Genomic DNA of leukemic patients: target for clinical diagnosis of MLL rearrangements. Biotechnol J 1: 656–663.

Meyer C, Kowarz E, Hofmann J, Renneville A, Zuna J, Trka J, Ben Abdelali R, Macintyre E, De Braekeleer E, De Braekeleer M, Delabesse E, de Oliveira MP, Cavé H, Clappier E, van Dongen JJ, Balgobind BV, van den Heuvel-Eibrink MM, Beverloo HB, Panzer-Grümayer R, Teigler-Schlegel A, Harbott J, Kjeldsen E, Schnittger S, Koehl U, Gruhn B, Heidenreich O, Chan LC, Yip SF, Krzywinski M, Eckert C, Möricke A, Schrappe M, Alonso CN, Schäfer BW, Krauter J, Lee DA, Zur Stadt U, Te Kronnie G, Sutton R, Izraeli S, Trakhtenbrot L, Lo Nigro L, Tsaur G, Fechina L, Szczepanski T, Strehl S, Ilencikova D, Molkentin M, Burmeister T, Dingermann T, Klingebiel T, Marschalek R (2009) New insights to the MLL recombinome of acute leukemias. Leukemia 23: 1490–1499.

Müller J, Verrijzer P (2009) Biochemical mechanisms of gene regulation by polycomb group protein complexes. Curr Opin Genet Dev 19: 150–158.

Nakamura T, Mori T, Tada S, Krajewski W, Rozovskaia T, Wassell R, Dubois G, Mazo A, Croce CM, Canaani E (2002) ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol Cell 10: 1119–1128.

Nakata Y, Brignier AC, Jin S, Shen Y, Rudnick SI, Sugita M, Gewirtz AM (2010) c-Myb, Menin, GATA-3, and MLL form a dynamic transcription complex that plays a pivotal role in human T helper type 2 cell development. Blood 116: 1280–1290.

Nam DK, Honoki K, Yu M, Yunis JJ (1996) Alternative RNA splicing of the MLL gene in normal and malignant cells. Gene 178: 169–175.

Nilson I, Löchner K, Siegler G, Greil J, Beck JD, Fey GH, Marschalek R (1996) Exon/intron structure of the human ALL-1 (MLL) gene involved in translocations to chromosomal region 11q23 and acute leukaemias. Br J Haematol 93: 966–972.

Park S, Osmers U, Raman G, Schwantes RH, Diaz MO, Bushweller JH (2010) The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression. Biochemistry 49: 6576–6586.

Reichel M, Gillert E, Angermüller S, Hensel JP, Heidel F, Lode M, Leis T, Biondi A, Haas OA, Strehl S, Panzer-Grümayer ER, Griesinger F, Beck JD, Greil J, Fey GH, Uckun FM, Marschalek R (2001) Biased distribution of chromosomal breakpoints involving the MLL gene in infants versus children and adults with t(4;11) ALL. Oncogene 20: 2900–2907.

Scharf S, Zech J, Bursen A, Schraets D, Oliver PL, Kliem S, Pfitzner E, Gillert E, Dingermann T, Marschalek R (2007) Transcription linked to recombination: a gene-internal promotor coincides with the recombination hot spot II of the human MLL gene. Oncogene 26: 1361–1371.

Schraets D, Lehmann T, Dingermann T, Marschalek R (2003) MLL-mediated transcriptional gene regulation investigated by gene expression profiling. Oncogene 22: 3655–3668.

Schuettengruber B, Cavalli G (2009) Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development 136: 3531–3542.

Turner BM (2002) Cellular memory and the histone code. Cell 111: 285–291.

Wang Z, Song J, Milne TA, Wang GG, Li H, Allis CD, Patel DJ (2010) Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. Cell 141: 1183–1194.

Weishaupt H, Sigvardsson M, Attema JL (2010) Epigenetic chromatin states uniquely define the developmental plasticity of murine hematopoietic stem cells. Blood 115: 247–256.

Xia ZB, Anderson M, Diaz MO, Zeleznik-Le NJ (2003) MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. Proc Natl Acad Sci USA 100: 8342–8347.

Yokoyama A, Wang Z, Wysocka J, Sanyal M, Aufiero DJ, Kitabayashi I, Herr W, Cleary ML (2004) Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. Mol Cell Biol 24: 5639–5649.

Yokoyama A, Cleary ML (2008) Menin critically links MLL proteins with LEDGF on cancer-associated target genes. Cancer Cell 14: 36–46.

Zhang Z, Jones A, Sun CW, Li C, Chang CW, Joo HY, Dai Q, Mysliwiec MR, Wu 38, Guo Y, Yang W, Liu K, Pawlik KM, Erdjument-Bromage H, Tempst P, Lee Y, Min J, Townes TM, Wang H (2011) PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. Stem Cells 29: 229–240.

Zhou W, Chen H, Zhang L (2009). The PcG protein hPc2 interacts with the N-terminus of histone demethylase JARID1B and acts as a transcriptional co-repressor. BMB Rep 42: 154–159.