

RESEARCH ARTICLE

STEM CELLS AND REGENERATION

Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant

Sergei Denisov^{1,*}, Helmut Hofemeister^{2,*}, Hendrik Marks^{1,*}, Andrea Kranz^{2,*}, Giovanni Ciotta², Sukhdeep Singh², Konstantinos Anastassiadis³, Hendrik G. Stunnenberg^{1,‡} and A. Francis Stewart^{2,‡}

ABSTRACT

Trimethylation of histone H3 lysine 4 (H3K4me3) at the promoters of actively transcribed genes is a universal epigenetic mark and a key product of Trithorax group action. Here, we show that Mll2, one of the six Set1/Trithorax-type H3K4 methyltransferases in mammals, is required for trimethylation of bivalent promoters in mouse embryonic stem cells. Mll2 is bound to bivalent promoters but also to most active promoters, which do not require Mll2 for H3K4me3 or mRNA expression. By contrast, the Set1 complex (Set1C) subunit Cxxc1 is primarily bound to active but not bivalent promoters. This indicates that bivalent promoters rely on Mll2 for H3K4me3 whereas active promoters have more than one bound H3K4 methyltransferase, including Set1C. Removal of Mll1, sister to Mll2, had almost no effect on any promoter unless Mll2 was also removed, indicating functional backup between these enzymes. Except for a subset, loss of H3K4me3 on bivalent promoters did not prevent responsiveness to retinoic acid, thereby arguing against a priming model for bivalency. In contrast, we propose that Mll2 is the pioneer trimethyltransferase for promoter definition in the naïve epigenome and that Polycomb group action on bivalent promoters blocks the premature establishment of active, Set1C-bound, promoters.

KEY WORDS: Epigenetics, Epigenome, Histone methylation, Bivalent promoters, Trithorax group, Polycomb group, Kmt2

INTRODUCTION

In eukaryotes, transcription is regulated not only by transcription factors that bind specific DNA sequences near the regulated gene, but also by post-translational modifications of the nucleosomes that surround and encompass these DNA sequences. The modifications include methylation, acetylation and mono-ubiquitylation of histone tails that project out from the core nucleosome and serve as binding sites for chromatin proteins and complexes (Bannister and Kouzarides, 2011; Sukanuma and Workman, 2011). In vertebrates, nucleosome modifications, together with cytosine methylation, influence transcriptional regulation during development, adult life

and disease (Albert and Helin, 2010; Butler et al., 2012; Reik, 2007). This epigenetic level of transcriptional regulation is crucial to the multiple ways in which a genome is interpreted in multicellular organisms (Goldberg et al., 2007).

Metazoan development is regulated by programmed transcriptional hierarchies acting in synergy with epigenetic mechanisms (Fisher and Fisher, 2011; Jaenisch and Bird, 2003; Magnúsdóttir et al., 2012). The first clues about how epigenetic mechanisms regulate gene expression were discovered in *Drosophila* through genetic screens that juxtaposed the repressive Polycomb group (PcG) and the activating Trithorax group (TrxG) (Brock and Fisher, 2005; Ringrose and Paro, 2004; Simon and Tamkun, 2002). These genetic interactions were given biochemical relevance when it was discovered that PcG and TrxG proteins confer mutually exclusive lysine methylations on the histone H3 tail. TrxG action was first linked to histone H3 lysine 4 methylation (H3K4me) when the yeast homologue of the *Drosophila* TrxG protein, Ash2, was found to be a subunit of the first H3K4 methyltransferase complex, the Set1 complex (Set1C) (Roguev et al., 2001). The SET domain of Set1 methylates H3K4 and is virtually identical to the SET domain in Trithorax itself, indicating that Trithorax is also an H3K4 methyltransferase, as subsequently shown for the mammalian orthologue Mll1 (Milne et al., 2002). PcG action was linked to H3K27 methylation when the SET domain protein Enhancer of Zeste homologue 2 (EZH2) was shown to methylate H3K27 and to be a subunit of the widely conserved Polycomb repressor complex 2 (PRC2) (Cao et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002).

Mammals have six Set1/Trithorax-type H3K4 methyltransferases (Glaser et al., 2006), which are all found in similar protein complexes based on the Set1C/Ash2 scaffold (Ruthenburg et al., 2007; Yokoyama et al., 2004). Hence, the core of the H3K4 methyltransferase complex is one of the most highly conserved components of the epigenetic machinery. This extreme evolutionary conservation reflects the fact that H3K4 trimethylation (H3K4me3) is the only universally conserved epigenetic modification.

The genetic opposition between PcG and TrxG in metazoan development relates to the differing consequences of H3K4 and H3K27 trimethylation for the nucleosomes surrounding promoters. H3K4me3 is bound by factors involved in transcriptional activity, such as ING proteins (Champagne and Kutateladze, 2009), TFIID/Taf3 (Vermeulen et al., 2007), JMJD2A (Huang et al., 2006), NURF/BPTF (Wysocka et al., 2006), CHD1 and U2snRNP components (Sims et al., 2007), PHF8 and Sgf29/SAGA (Vermeulen et al., 2010). This list also includes the Set1/Trithorax H3K4 methyltransferase complexes themselves (Eberl et al., 2013; Milne et al., 2010; Shi et al., 2007), suggesting that H3K4me3 ensures epigenetic maintenance through a feed-forward maintenance cycle. Similarly, H3K27me3 recruits PRC2 itself (Hansen et al., 2008; Margueron et al., 2009) and also PRC1, the companion PcG

¹Department of Molecular Biology, Faculty of Science, Nijmegen Centre for Molecular Life Sciences (NCMLS), Radboud University, PO Box 9101, 6500 HB Nijmegen, The Netherlands. ²Genomics, Biotechnology Center, Technische Universitaet Dresden, Tatzberg 47, 01307 Dresden, Germany. ³Stem Cell Engineering, Biotechnology Center and Center for Regenerative Therapy, Technische Universitaet Dresden, Tatzberg 47, 01307 Dresden, Germany. *These authors contributed equally to this work

[‡]Authors for correspondence (h.stunnenberg@ncmls.ru.nl; stewart@biotec.tu-dresden.de)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

complex in gene repression (Bernstein et al., 2006b; Fischle et al., 2003).

As well as recruiting proteins involved in gene expression, TrxG action also prevents PcG-mediated silencing (Klymenko and Müller, 2004), potentially through binding occlusion by H3K4me3, as found for exclusion of NuRD (Zegerman et al., 2002) or a DNA methylation complex (Ooi et al., 2007). Therefore, the finding that certain promoters, termed bivalent, feature both H3K4me3 and H3K27me3 (Azuara et al., 2006; Bernstein et al., 2006a) was unexpected and controversial (Akkers et al., 2009; Herz et al., 2009; Vastenhouw and Schier, 2012; Vastenhouw et al., 2010; Voigt et al., 2013). Recent progress with bivalency includes the findings that PRC2 cannot methylate H3K27 *in vitro* when H3K4 is trimethylated on the same tail (Schmitges et al., 2011; Voigt et al., 2012), that bivalent promoters have the least H3K4 trimethylation of any H3K4me3 promoters (Li et al., 2010; Xu et al., 2010) and that H3K4me3 is restricted to the +1 nucleosome whereas H3K27me3 is at least partially excluded from that nucleosome and found on flanking nucleosomes (Marks et al., 2012).

Bivalent promoters were discovered in mouse embryonic stem cells (mESCs), where they were mostly found in genes involved in lineage specification. In mESCs these genes are not, or are very poorly, expressed despite having an H3K4me3 peak near the transcription start site (TSS). Hence, a role for bivalency in keeping crucial gene expression off, but poised until the appropriate developmental stage, was proposed (Azuara et al., 2006; Bernstein et al., 2006a; Ku et al., 2008; Vastenhouw and Schier, 2012).

Bivalency is a component of the epigenetic blueprint that contributes to the orchestration of gene expression hierarchies in development. Epigenetic remodelling during germ cell maturation and early embryonic development resets the epigenome back to the totipotent ground state of the zygote and pluripotency of the epiblast (Magnúsdóttir et al., 2012; Marks et al., 2012; Wray et al., 2010). Towards the end of spermatogenesis and oogenesis, transcription from the germ cell genome is shut down in preparation for meiosis. Before transcriptional activation of the embryonic genome shortly after the first cell division, the paternal epigenome must be reassembled after its substantial stripping during spermatogenesis. How the epigenome is reprogrammed during early development is still unclear. Previously, we reported that the Trithorax orthologue Mll2 is a key player in resetting the epigenome. It is required during oogenesis and early cleavage stages after fertilisation, when it is the major H3K4 trimethyltransferase (Andreu-Vieyra et al., 2010). Later in development, it is again required after gastrulation, when it is no longer the major H3K4 trimethyltransferase. *Mll2*^{-/-} mouse embryos die at ~E10.5, displaying retarded growth and development as well as widespread apoptosis (Glaser et al., 2006; Glaser et al., 2009). Here we show that Mll2 is required for trimethylation of bivalent promoters in mESCs.

RESULTS

Mll2 regulates H3K4me3 at bivalent promoters

It has been shown that Mll2 is dispensable for the self-renewal of mESCs. Overall, fewer than 100 genes were downregulated more than 2-fold when *Mll2* was deleted, including just one gene, *Magohb*, that was entirely dependent on Mll2 for expression (Glaser et al., 2009; Ladopoulos et al., 2013). This modest impact on gene expression was obtained with two different loss-of-function alleles for *Mll2* (*Mll2*^{-/-} and *Mll2FC/FC*; supplementary material Fig. S1) in different genetic backgrounds (12901a [E14Tg2a=E14] and C57Bl/6, respectively).

Consequently, we were surprised to discover that ~3000 promoters showed more than 2.5-fold loss of H3K4me3 in the E14 *Mll2*^{-/-} ESCs (Fig. 1A). To verify this observation, we employed tamoxifen-induced conditional mutagenesis in the C57Bl/6 ESCs to convert the *Mll2* allele from *F/F* to *FC/FC* (supplementary material Fig. S1), which completely eliminates Mll2 protein by 48 hours (Glaser et al., 2009). After culturing for a further 48 hours, over 900 promoters showed greater than 2.5-fold loss of H3K4me3. Again, this is far more than the 60 genes that were downregulated by more than 2-fold at the same 96 hour time point (Glaser et al., 2009), and most (75%) of these 900 affected promoters were also affected in constitutively ablated *Mll2*^{-/-} ESCs (Fig. 1A,B). This suggests that the promoters affected at 96 hours are predominantly the most sensitive direct targets of Mll2, whereas the additional changes found in *Mll2*^{-/-} ESCs include less sensitive promoters and secondary, adaptive changes.

By contrast, no H3K4me3 changes were found in E14 *Mll1*^{-/-} ESCs (Fig. 1C,D), despite the fact that Mll1 protein is expressed in wild-type ESCs (Testa et al., 2003) and is highly related to *Mll2*, having arisen by gene duplication (FitzGerald and Diaz, 1999; Glaser et al., 2006; Huntsman et al., 1999). Consistent with the unchanged H3K4me3 peaks, loss of Mll1 had very little effect on the mRNA expression profile (Fig. 1D). As with Mll2, fewer than 100 genes were downregulated more than 2-fold when Mll1 was ablated in ESCs.

To understand the discrepancy between the large effect on promoter H3K4me3 peaks and the small effect on mRNA levels in the absence of Mll2, we examined the relationship between promoter H3K4me3 levels and mRNA expression levels (Fig. 1D). This revealed that almost all of the affected promoters expressed very little, or no, mRNA in wild-type ESCs. These promoters also showed the least H3K4 trimethylation. Because small H3K4me3 peaks and low to absent mRNA expression are characteristics of bivalent promoters, we looked at the other signature of bivalency, H3K27me3.

Fig. 2A displays all 15,610 H3K4me3 promoters that we identified in ESCs (>30 reads above background in the 2 kb surrounding TSSs) sorted according to the ratio of H3K4me3 change comparing wild-type E14 and *Mll2*^{-/-} cells from most increased (Fig. 2A, lane 1, at the bottom) to most decreased (at the top). *Mll2FC/FC* results show a reduced, but similar, profile (Fig. 2A, lane 2). Remarkably, this criterion for sorting genes according to the impact on H3K4me3 of mutating Mll2 also approximately arranged: (1) promoters from the most to least H3K4me3 reads (Fig. 2A, lanes 3-7); (2) promoters from the least to most H3K27me3 reads (lanes 8-11); (3) genes from the most to least expressed (lanes 12, 13); (4) promoters from most to least RNA polymerase II (Pol II) occupancy (lane 14). Consequently, it also clustered the bivalent promoters at the top (Fig. 2A, lane 15). These data show that the bivalent promoters lose H3K4me3 upon loss of Mll2 and are the least H3K4 trimethylated (Fig. 2B). These points are also illustrated by typical examples and conventional ChIP of genes unaffected and affected by loss of Mll2 (Fig. 2C,D). Hence, the discrepancy between a small impact on expression levels and a much larger impact on H3K4me3 levels at thousands of promoters in *Mll2*^{-/-} ESCs is explained by the realisation that Mll2 is the primary H3K4 trimethyltransferase for bivalent promoters.

Characterisation of tagged BAC transgenes by rescue of Mll2 function

To identify Mll2 genomic binding sites in ESCs, we applied BAC transgenic protein tagging methods for ChIP-seq (Hofemeister et al.,

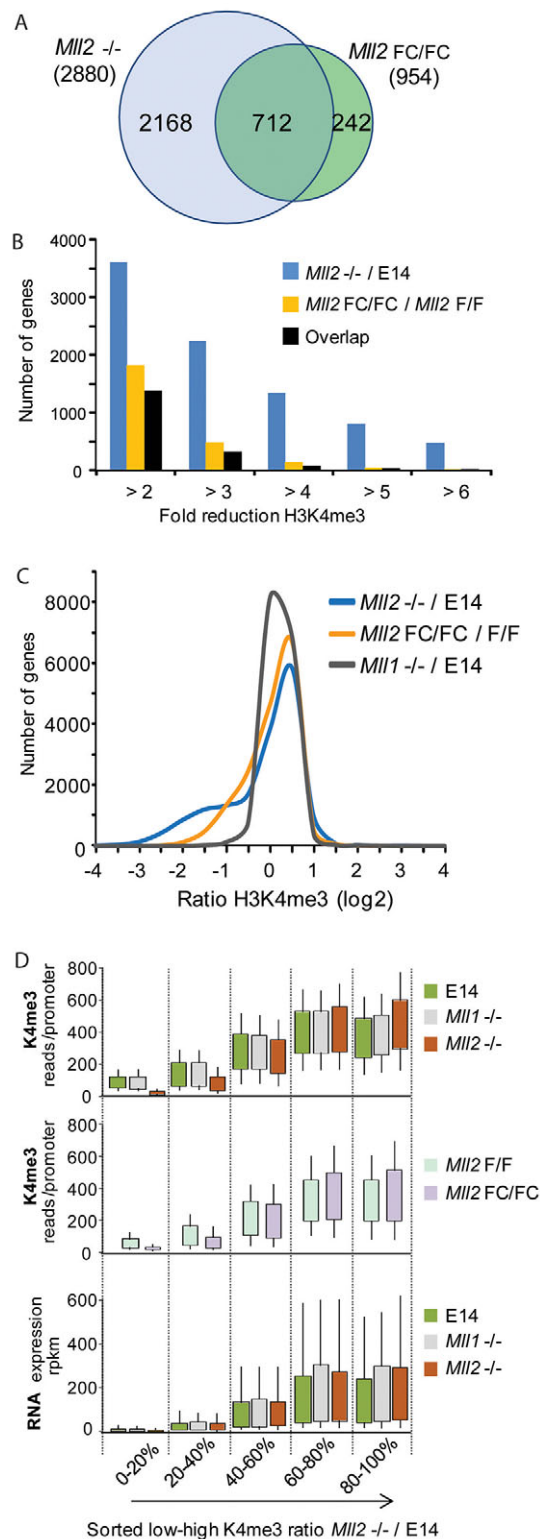


Fig. 1. MII2 regulates H3K4me3 at promoters of the least expressed genes. (A) Comparison of H3K4me3 peaks in mouse E14 ESCs and *MII2*^{-/-} ESCs revealed 2880 peaks that were reduced >2.5-fold in the absence of MII2. A similar comparison between *MII2F/F* cells and *MII2FC/FC* cells (that is, before and 96 hours after the start of 4-hydroxy tamoxifen induction of Cre recombination) revealed 954 reduced peaks with an overlap of 712 between the two analyses. (B) The same analyses as in A except divided into classes according to the extent of reduction. (C) Pairwise comparison of all 15,610 H3K4me3 peaks [>30 reads per transcription start site (TSS)] found in E14 ESCs with either *MII2*^{-/-} or *MII1*^{-/-} ESCs, as well as a comparison of *MII2FC/FC* with *MII2F/F*. (D) Data from ChIP-seq and RNA-seq profiling of E14, *MII1*^{-/-} and *MII2*^{-/-} ESCs were binned according to the H3K4me3 ratio between *MII2*^{-/-} and E14 (essentially the blue curve in C was sliced into five bins) and the results plotted for total number of H3K4me3 reads per TSS and reads per kilobase of exon per million mapped reads (rpkm), as indicated.

MII1 and MII2, like their fly orthologue *Drosophila* Trithorax, are cleaved by the asparaginyl protease Taspase at a conserved site in the middle of the protein. However, the N- and C-terminal sections remain associated due to an intramolecular interaction between the FYR-N and FYR-C domains (Hsieh et al., 2003a; Hsieh et al., 2003b). To tag MII2 internally, GFP was inserted next to the Taspase cleavage site, either on the N- or C-terminal side to generate MII2-FL N-GFP or MII2-FL C-GFP (FL, full-length; supplementary material Fig. S1) by recombineering a BAC transgene (Hofemeister et al., 2011). The tagged BAC transgenes were introduced into the conditional *MII2F/F* ESCs and tested for cleavage by Taspase and rescue of function after tamoxifen-induced elimination of endogenous MII2 (supplementary material Fig. S2). In ESCs, endogenous MII2 is almost completely cleaved (supplementary material Fig. S2A, lanes 1, 7). We observed physiological expression levels and Taspase cleavage of both full-length tagged MII2s in the presence and absence of endogenous MII2. Both the N-terminal-only (MII2 N-GFP) and C-terminal-only (MII2 C-GFP) tagged half-proteins were also expressed at physiological levels from the BAC transgenes. Loss or additional expression of MII2 did not change the expression of *Ash2l*, which is the core subunit for the H3K4 methyltransferases, or the pluripotency factor *Nanog* (supplementary material Fig. S2A).

To evaluate the function of the tagged forms of MII2, we examined their ability to rescue three defects in the *MII2*^{-/-} ESC phenotype. *MII2*^{-/-} ESCs show increased doubling time, loss of *Magohb* expression and delayed differentiation upon removal of LIF (Glaser et al., 2009; Lubitz et al., 2007). Both of the tagged full-length BAC transgenes (MII2-FL N-GFP and MII2-FL C-GFP) rescued all three phenotypic faults, whereas the N-terminal-only (MII2 N-GFPonly) and C-terminal-only (MII2 C-GFPonly) constructs did not (supplementary material Fig. S2B-D), except for partial rescue by the N-terminal-only construct in the differentiation assay (supplementary material Fig. S2D).

As well as demonstrating that the tagged BAC transgenes rescued function, we also demonstrated by generic native affinity purification using the tag followed by mass spectrometry (AP-MS) (Hofemeister et al., 2011; Hubner et al., 2010) that the tagged proteins retrieved all of the expected interaction partners (data not shown).

Characterisation of MII2 binding sites in the ESC genome

Tagged MII2 was used for ChIP-seq. Peak identification aligned almost all of the 11,000 peaks to TSSs (Fig. 3A), which were H3K4 trimethylated as expected (Fig. 3B). Fifteen percent also showed H3K27me3. A preference for CpG islands was observed (84%, versus 70% if randomly distributed) (Mikkelsen et al., 2007)

2011). For H3K4 methyltransferases, the stop codon is a highly conserved and integral part of the post-SET domain. Previously, we found that C-terminal tagging of yeast Set1 inactivated the methyltransferase activity (Roguev et al., 2001), whereas N-terminal tagging of MII2 unexpectedly generated a hypomorphic allele that caused sterility in homozygous females (Andreu-Vieyra et al., 2010). Hence, we explored a third strategy based on internal tagging.

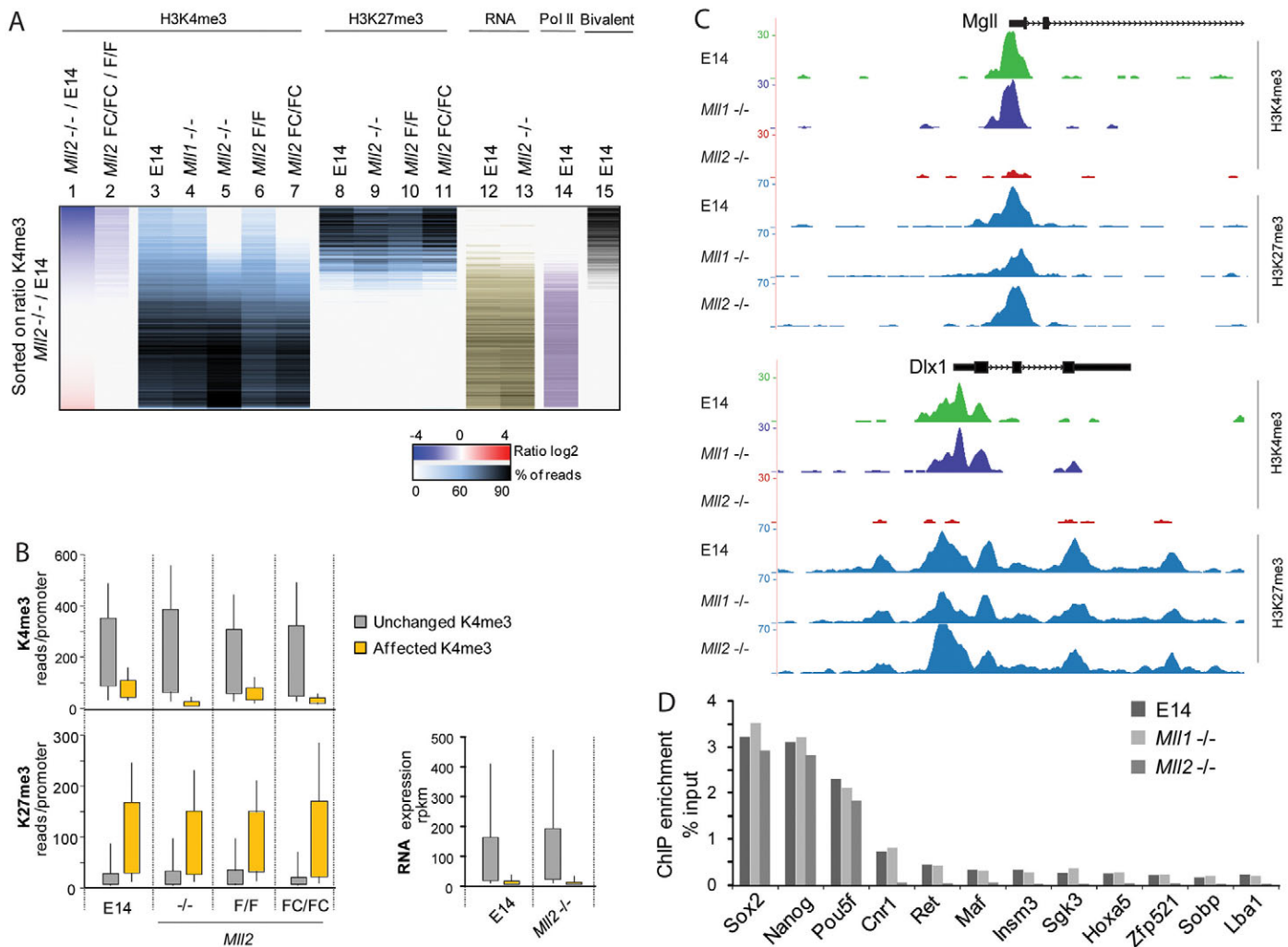


Fig. 2. *MII2* regulates H3K4me3 on bivalent promoters. (A) All 15,610 H3K4me3 TSSs were ordered from most increased (at bottom) to most decreased (at top) H3K4me3 from the comparison of *MII2*^{-/-} with E14 ESCs, which is plotted in the column on the far left (column 1). All data in columns 2-15 are stacked according to this order. Column 2 shows the comparison of *MII2*FC/FC with *MII2*F/F followed by the intensities of H3K4me3 and H3K27me3 peaks from the various cell lines indicated (columns 3-11). RNA intensities of the corresponding mRNAs were estimated from RNA-seq (columns 12, 13) and Pol II promoter occupancy from ChIP-seq analysis of E14 cells (column 14) (Marks et al., 2012). Bivalency (column 15) scores the co-occurrence of H3K4me3 and H3K27me3. (B) Box plots of genes unaffected or affected (>2.5-fold lower H3K4me3) in *MII2*^{-/-} cells compared with E14 ESCs. (C) Two typical bivalent promoters from H3K4me3 and H3K27me3 ChIP-seq analyses performed with E14, *MII1*^{-/-} and *MII2^{-/-} ESCs. (D) Conventional ChIP qPCR analysis of three active gene promoters (*Sox2*, *Nanog*, *Pou5f*) and nine bivalent promoters from E14, *MII1*^{-/-} and *MII2^{-/-} ESCs.**

(Fig. 3B). We also tagged the common H3K4 methyltransferase subunit Ash2l and the Set1C-specific subunit Cxxc1 (supplementary material Fig. S3). In contrast to a recent report on Ash2l in ESCs (Wan et al., 2013), ~10,000 TSS promoters enriched for Ash2l were observed. These strongly correlated with *MII2* sites (88%) and CpG islands (84%), indicating that Ash2l was also localised at most *MII2*-occupied TSSs. Interestingly, this correlation was less evident for the bivalent promoters (907/1652=55%). Fewer TSS binding sites were observed for Cxxc1 than for *MII2* (8271 versus 10936); however, most Cxxc1 sites (90%) were co-occupied by *MII2*. Conversely, only 68% of *MII2* sites were co-occupied by Cxxc1 and only 8% of Cxxc1-bound TSSs also showed H3K27me3 (Fig. 3B). Whereas the tagged C-terminal half of *MII2* did not yield peaks or interpretable data in a ChIP-seq experiment (not shown), the N-terminal half frequently colocalised to the same sites as full-length *MII2*, indicating that *MII2* predominantly binds via its N-terminus (Fig. 3B).

To evaluate these data in more detail, we included peak quantifications in an intensity plot. *MII2* was bound to the promoters that lost H3K4me3 when *MII2* was mutated, reinforcing the conclusion that *MII2* is the main H3K4 trimethyltransferase for bivalent promoters. However, *MII2* was also bound to most H3K4me3 promoters (Fig. 3C,D). To challenge the veracity of *MII2* genomic binding sites, we employed a different method for ChIP using a double-crosslinking protocol (Fig. 3C) (Nowak et al., 2005; van den Berg et al., 2008). This protocol achieved a very similar binding profile as found for conventionally crosslinked full-length and N-terminal *MII2* (Fig. 3C,D), albeit with a greater difference between the specific signals and the mock ChIP (Fig. 3C).

Ash2l was also bound to most H3K4me3-marked promoters; however, it was less frequently bound to bivalent promoters and this trend was even more pronounced for Cxxc1 (Fig. 3D), suggesting that Set1C was primarily bound to active but not bivalent promoters. These conclusions are also illustrated by representative examples

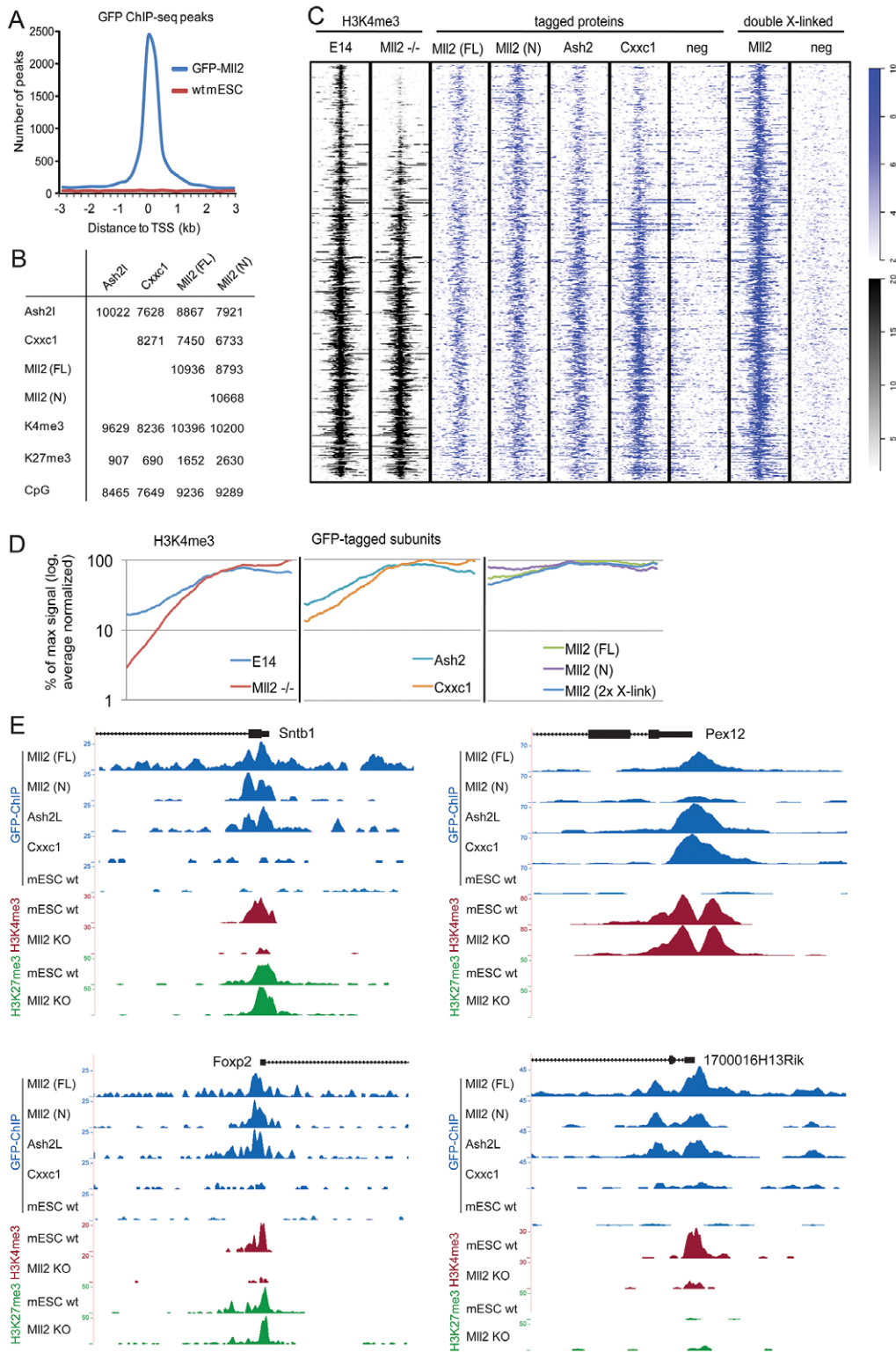


Fig. 3. MII2 binds to most H3K4me3 promoters, including bivalent promoters. (A) The distribution of all significant MII2 peaks with respect to TSSs shows that most peaks are found at TSSs. The negative control based on GFP antibody ChIP using untagged E14 ESCs is also shown. (B) Quantification and pairwise comparisons of ChIP-seq peaks mapped from tagged Ash2, Cxyc1, MII2 (FL) (full-length MII2; single-crosslinked sample) and MII2 (N) (the N-terminal region of MII2 to the Taspase cleavage site) as well as H3K4me3, H3K27me3 and CpG islands. (C) Intensity plots of the same data sets in B plus MII2 (FL) after double crosslinking as indicated, stacked in the same order as in Fig. 2A. Each column is orientated according to the TSS and encompasses 3 kb upstream and downstream. Lanes 1 and 2 present the same data as in Fig. 2A, lanes 3 and 5. Neg refers to ChIP-seq results from wild-type E14 untagged ESCs crosslinked using either standard or double-crosslinking protocols as indicated, followed by immunoprecipitation using the anti-GFP antibody. (D) Line plots of moving averages (50 TSSs per bin) based on the profiles shown in the heatmaps of C. The percentage of maximal signal is displayed on log plots from the top to the bottom of the heatmaps. (E) Examples of two bivalent (*Sntb1*, *Foxp2*) and two non-bivalent (*Pex12*, *1700016H13Rik*) promoters displaying tagged protein, H3K4me3 and H3K27me3 ChIP-seq profiles.

(Fig. 3E). *Sntb1* and *Foxp2* are bivalent genes that rely on and are bound by Mll2 and not Cxxc1. *Pex12* is a non-bivalent gene that does not rely on Mll2 and is bound by both Mll2 and Cxxc1. *1700016H13Rik* is an example of a rare, non-bivalent gene that relies on Mll2 in ESCs; notably, it is bound by Mll2 but not by Cxxc1.

ATRA differentiation in *Mll2*^{-/-} ESCs

Self-renewal of mESCs does not require Mll2 (Lubitz et al., 2007) or, as now revealed, H3K4me3 enrichment on bivalent promoters. Because bivalency is important for lineage commitment (Fisher and Fisher, 2011), we treated E14 and *Mll2*^{-/-} ESCs with all-trans retinoic acid (ATRA) for 4 days to examine early events in differentiation.

About half of the 2880 TSSs (1536) that had reduced H3K4me3 in undifferentiated *Mll2*^{-/-} ESCs compared with wild type (Fig. 1A) were still reduced in *Mll2*^{-/-} cells after 4 days of ATRA treatment (Fig. 4A). However, the other half (1344) recovered H3K4me3 levels to within 2.5-fold of wild-type levels, despite the lack of Mll2, indicating that another H3K4 methyltransferase was acting on these bivalent promoters to restore H3K4me3 levels.

In wild-type E14 cells, we identified 136 bivalent TSSs at which H3K4me3 levels increased more than 2.5-fold after 4 days of ATRA treatment compared with undifferentiated ESCs (Fig. 4B). Of these, 120 were also elevated in *Mll2*^{-/-} ESCs, suggesting that only a few bivalent promoters (16) were impaired for ATRA responsiveness by the absence of Mll2. In addition, a further 442 TSSs that are bivalent in wild-type E14 ESCs showed increased H3K4me3 in *Mll2*^{-/-} ESCs (Fig. 4B,C), reflecting a ‘catch-up’ during ATRA differentiation to near wild-type H3K4me3 levels. Of the 16 bivalent promoters that were impaired, we selected six for closer examination (Fig. 4D). In several cases, H3K4me3 levels were induced by ATRA but remained less than 2.5-fold of the H3K4me3 levels at the equivalent wild-type promoter, whereas in other cases an ATRA response was absent. Because *Mll2*^{-/-} ESCs showed delayed differentiation (Lubitz et al., 2007), we reasoned that the reduced ATRA response at these promoters could be due to a general delay. Consequently, we induced for 7 days with ATRA, and found a very similar profile including a similar minority of retarded promoters after 7 days (supplementary material Fig. S5), indicating that the problem was not simply due to timing and delay. In conclusion, the absence of Mll2 or H3K4me3 peaks on bivalent TSSs did not prevent ATRA stimulation in the majority of cases. In a minority of cases, the ATRA response was apparently diminished beyond a point of recovery.

ATRA induction in the absence of Mll1

As observed above, in *Mll2*^{-/-} ESCs (Fig. 4A) another H3K4 methyltransferase(s) must trimethylate the responsive bivalent TSSs during ATRA differentiation. To explore this issue, we knocked out both alleles of the *Mll2* sister gene *Mll1* by consecutive rounds of targeting in E14 ESCs (supplementary material Fig. S3). We also generated double-knockout ESCs lacking both Mll1 and Mll2 by crossing mouse lines carrying conditional alleles for both *Mll1* and *Mll2* with *Rosa26CreERT2* (Seibler et al., 2003) followed by *de novo* establishment of a male *Mll1F/F; Mll2F/F; R26CreERT2/+* ESC line. This line was induced with tamoxifen to simultaneously knock out both *Mll1* and *Mll2* to generate *Mll1FC/FC; Mll2FC/FC* cells, which were harvested 144 hours after tamoxifen was administered.

Concordant with the absence of changes in undifferentiated *Mll1*^{-/-} ESCs (Fig. 1C,D), very few H3K4me3 TSS peaks (84) were

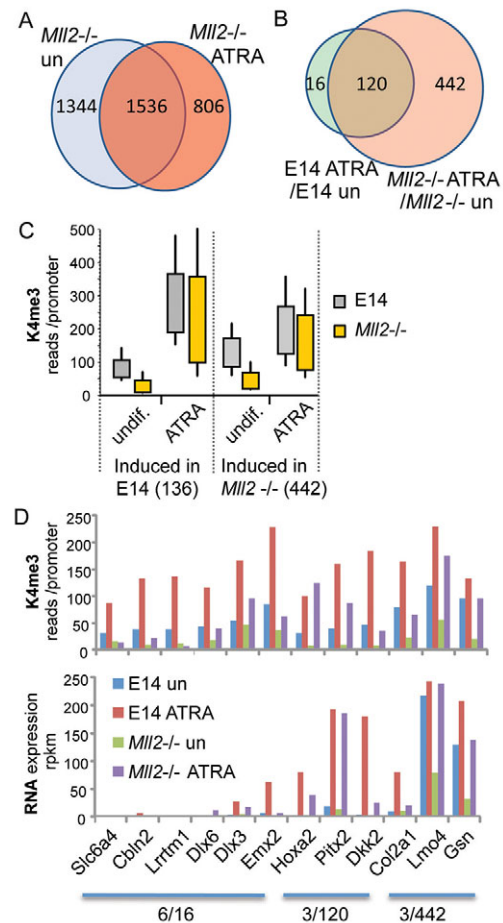


Fig. 4. Retinoic acid induction in wild-type and *Mll2*^{-/-} ESCs. (A) After 4 days of all-trans retinoic acid (ATRA) induction, H3K4me3 levels at TSSs were compared between *Mll2*^{-/-} and E14 ESCs to identify the peaks that were more than 2.5-fold reduced in the absence of Mll2 (2342; orange circle). These were compared with the 2880 peaks identified in undifferentiated ESCs (Fig. 1A, blue circle). (B) Putative direct bivalent targets of ATRA were identified by comparing H3K4me3 peaks before and after induction. In E14 ESCs, 136 bivalent TSSs (green circle) increased more than 2.5-fold after 4 days ATRA treatment, whereas in *Mll2*^{-/-} ESCs most of the same TSSs (120) plus an additional 442 bivalent TSSs (pink circle) increased more than 2.5-fold. (C) Box plot averages of the two categories of bivalent TSSs presented in B showed that ATRA induction in both cases resulted in increased H3K4me3 in the absence of Mll2. (D) Six TSSs were selected from the 16 genes identified in B that still had >2.5-fold reduced H3K4me3 in *Mll2*^{-/-} compared with E14 ESCs after 4 days of ATRA induction. These six TSSs are compared with three TSSs from the other two categories (containing 120 and 442 genes, respectively), as indicated. The plots show data extracted from the H3K4me3 ChIP-seq and RNA expression profile datasets.

changed more than 2.5-fold in ATRA-induced *Mll1*^{-/-} ESCs compared with ATRA-induced wild type (Fig. 5A,B), indicating again that Mll1 is not required for ESC self-renewal. Notably, most of the affected TSSs in ATRA-induced *Mll1*^{-/-} or *Mll2*^{-/-} ESCs were also affected in ATRA-induced double-knockout *Mll1FC/FC; Mll2FC/FC* ESCs (Fig. 5C; 75/84=89% and 2060/2342=88%, respectively) thereby indicating high reproducibility of these data despite the differences in experimental origin and ESC genetic background. In addition to these >2000 TSSs, a further 1717 H3K4me3-reduced TSSs were observed in the doubly mutated cells, indicating that either Mll1 or Mll2 catalyzes H3K4 trimethylation on these promoters (Fig. 5C).

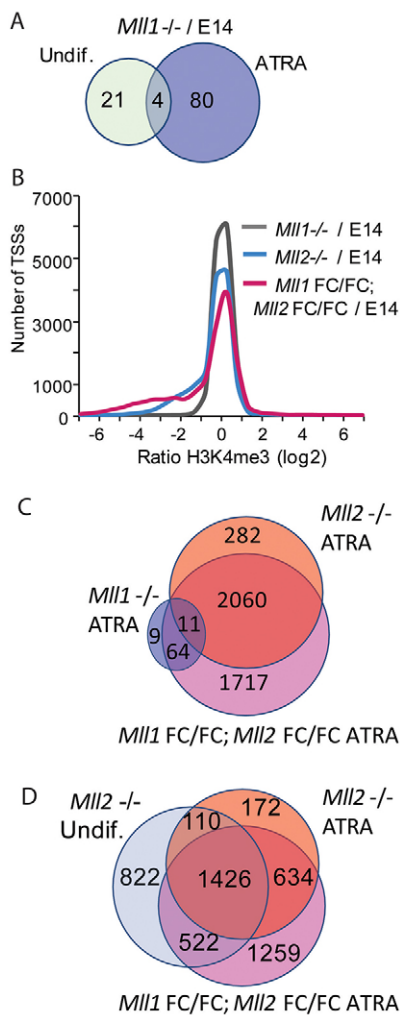


Fig. 5. Functional relationship between MII2 and MII1. (A) E14 ESCs knocked out for *MII1* (*MII1*^{-/-}) show very few H3K4me3 differences (of >2.5-fold) to wild-type E14 ESCs before (green circle, 25 genes) or after (blue circle, 84 genes) 4 days of ATRA induction. (B) All H3K4me3 TSS peaks were compared after 4 days of ATRA induction between *MII1*^{-/-} and E14 (grey line), *MII2*^{-/-} and E14 (blue line), or *MII1FC/FC*; *MII2FC/FC* and E14 (red line) ESCs. (C) Comparisons of H3K4me3 differences after 4 days of ATRA differentiation between *MII2*^{-/-} and E14 (orange circle, same as Fig. 4A), *MII1*^{-/-} and E14 (blue circle, same as A) and *MII1FC/FC*; *MII2FC/FC* and E14 (pink circle) ESCs. After ATRA, most (2060/2342) TSSs that were reduced >2.5-fold in H3K4me3 in the absence of MII2 compared with wild type were also reduced in the double-mutant *MII1FC/FC*; *MII2FC/FC* ESCs. Similarly, most (75/84) TSSs that were reduced >2.5-fold in H3K4me3 in the absence of MII1 compared with wild type were also reduced in the double-mutant *MII1FC/FC*; *MII2FC/FC* ESCs. A further 1717 cases in which H3K4me3 levels were reduced >2.5-fold were also observed in the double-mutant cells. (D) The same orange and pink datasets in C were compared with the set of H3K4me3-reduced TSSs observed in undifferentiated ESCs (light blue, same as in Fig. 1A and Fig. 4A). Notably, many TSSs that showed restored levels of H3K4me3 after ATRA differentiation of *MII2*^{-/-} cells (522/1344) were reduced in the double-mutant cells, indicating that MII1 had methylated these promoters in the absence of MII2.

As noted above (Fig. 4A), nearly half (1344) of the H3K4me3-reduced TSSs in *MII2*^{-/-} undifferentiated ESCs were restored to near wild-type H3K4me3 levels after ATRA differentiation. Of these, more than a third (522/1344=39%; Fig. 5D) were not restored in the double-mutant cells. Hence, MII1 is the trimethyltransferase that acted on

these TSSs in the absence of MII2. The remaining ~60% (822) indicate that other methyltransferases were involved. As expected, most of the 522 TSSs (401/522=77%) were bivalent in undifferentiated E14 ESCs. By contrast, only one-third of the extra 1259 TSSs additionally decreased in the double-mutant ESCs were bivalent (429/1259=34%). However, a GO term analysis of these 1259 TSSs revealed similar properties to bivalent promoters (i.e. mainly differentiation and lineage commitment terms, although the regulation of apoptosis might be a notable exception; data not shown).

The overlapping relationships between MII1, MII2 and other H3K4 methyltransferases are well illustrated by selected ChIP-seq profiles (Fig. 6). The selection includes active genes in ESCs (*Pou5f1*, *Nanog*); genes induced during ATRA differentiation (*Gata6*, *Rec8*); bivalent genes that rely on MII2 (*Dkk2*, *Mef2c*); a bivalent gene that relies on either MII1 or MII2 (*Pcdh18*) and a non-bivalent gene that relies on MII2 (*Coll1a1*). Notably, this unusual non-bivalent case, like *1700016H13Rik* (Fig. 3E), also has bound MII2 and a very small signal for Cxxc1.

H3K4me3 profiles on the Hox complexes exemplify the functional overlaps between the H3K4 methyltransferases (Fig. 6C) and include one of the few direct targets that we could identify for MII1. Before ATRA induction, few H3K4me3 peaks are visible on the Hox complexes, as expected in undifferentiated ESCs. After ATRA induction, the acquisition of some H3K4me3 peaks required MII2 (e.g. *Hoxd1*), some peaks required MII1 (e.g. *Hoxb5*) and some required both MII1 and MII2 (*Hoxa3*, *b2*, *c4*). However, many H3K4me3 peaks did not require either MII2 or MII1, again indicating the contribution of another methyltransferase.

DISCUSSION

Here, we report an unexpected specialisation among the six mammalian Set1/Trithorax-type methyltransferases. MII2 is mainly responsible for H3K4 trimethylation at ESC bivalent promoters. This was seen both shortly after removal of the protein, when ~1000 bivalent promoters had reduced H3K4me3, and many passages after the loss of MII2, when almost all ~3000 bivalent promoters showed little H3K4me3. Besides at bivalent promoters, few other changes in H3K4me3 were found. Furthermore, removal of its sister protein, MII1, had little effect on any H3K4me3 TSS peaks, unless MII2 itself was also removed. Concordantly, removal of either MII1 or MII2 or both together had no observable impact on bulk H3K4 methylation levels (supplementary material Fig. S4).

In agreement with the loss of H3K4me3 on bivalent promoters, we found that MII2 was bound to bivalent promoters in wild-type ESCs. However, it was also bound to most active promoters as well, without apparent effect on H3K4me3 levels or mRNA production. By contrast, the Set1C-specific subunit, Cxxc1, was bound to virtually every active promoter but much less frequently to bivalent promoters. This matches recent H3K4me3 ChIP-seq data from *Cxxc1*^{-/-} ESCs, which revealed that H3K4me3 levels are greatly diminished near active TSSs but not at bivalent sites (Clouaire et al., 2012). Similar connections between Cxxc1 and active promoters have been made using RNAi knockdown in fibroblasts or with a *Cxxc1* (*Cfp1*) mutant fly line (Ardehali et al., 2011; Thomson et al., 2010). Hence, it appears likely that the Set1 complexes account for most H3K4 trimethylation at active TSSs in many cell types, whereas the Trithorax orthologues (MII1, MII2) play more specialised roles.

Of the six mammalian H3K4 methyltransferases, only the localisation of MII1 and MII4 has been investigated by ChIP-seq. Two studies in somatic cells found that MII1 is bound to nearly all active promoters (Guenther et al., 2005; Scacheri et al., 2006), which matches our results for MII2 and Cxxc1. However, few TSSs

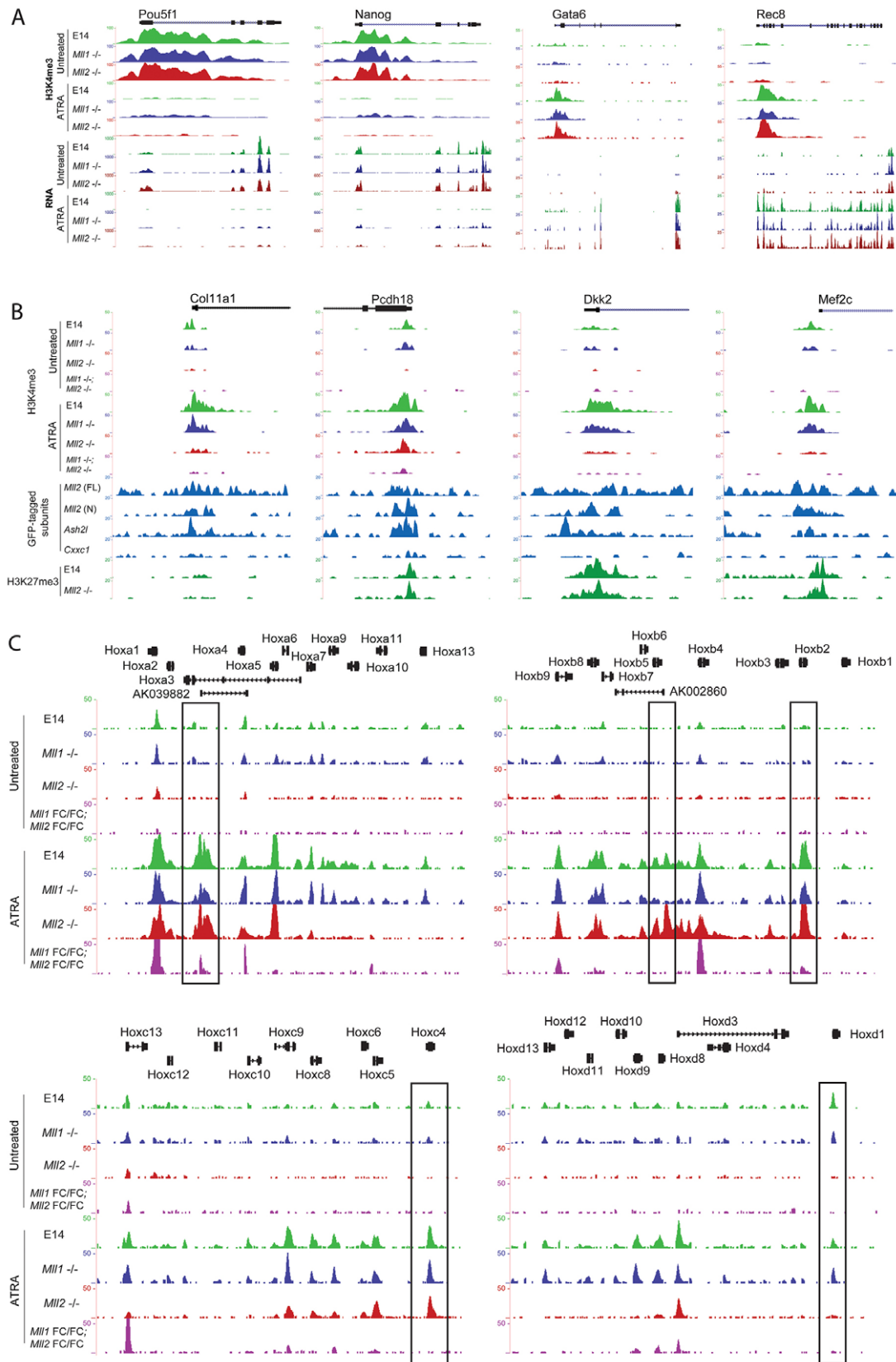


Fig. 6. Gene profiles of selected genes and Hox complexes. (A) Examples of non-bivalent genes downregulated (*Pou5f1*, *Nanog*) or upregulated (*Gata6*, *Rec8*) before and after 4 days of ATRA induction in E14, *MI1*^{-/-} and *MI2*^{-/-} ESCs evaluated by H3K4me3 ChIP-seq and RNA-seq. (B) Examples of three bivalent genes (*Pcdh18*, *Dkk2*, *Mef2c*) and one non-bivalent gene that relies on MI2 (*Col11a1*) before and after 4 days of ATRA induction in E14, *MI1*^{-/-}, *MI2*^{-/-} and *MI1FC/FC*; *MI2FC/FC* ESCs evaluated by H3K4me3 and H3K27me3 ChIP-seq as compared with tagged protein ChIP-seq from wild-type E14 ESCs for MI2 (FL, single crosslink), MI2 (N), Ash2l and Cxxc1. (C) H3K4me3 ChIP-seq profiles of the four Hox complexes before and after 4 days of ATRA induction in E14, *MI1*^{-/-}, *MI2*^{-/-} and *MI1FC/FC*; *MI2FC/FC* ESCs.

showed reduced H3K4me3 in *Mll1*^{-/-} mESCs (Figs 1, 6) or mouse embryonic fibroblasts (Wang et al., 2009). Together with data from *Mll4* (Guo et al., 2012), the emerging picture indicates that most active promoters are bound by more than one H3K4 methyltransferase. However, binding of *Mll1* or *Mll2* does not indicate function. Despite localisation to most TSSs, *Mll2* contributes to the expression of only a few genes. In mESCs, expression of *Magohb*, but not its sister gene *Magoha*, requires *Mll2*. Upon *Mll2* removal, the *Magohb* promoter, which is a CpG island, loses H3K4me3, acquires H3K27me3 and increased CpG methylation (Glaser et al., 2009; Ladopoulos et al., 2013). In macrophages, *Mll2* is required for the expression of only a few mRNAs, including *Pigp*, an essential component in glycosylphosphatidylinositol (GPI) anchor synthesis (Austena et al., 2012). Here, we identify two additional non-bivalent genes that are reliant on *Mll2* in ESCs: *Coll1a1* and *1700016H13Rik*. Notably, these genes lack *Cxxc1* bound at their promoters, suggesting that they are dependent on *Mll2* because *Set1C* is not available.

Because deletion of *Mll2* ablated H3K4me3 at bivalent promoters but left active promoters unaffected, we were presented with the unique opportunity to test the proposition that bivalent promoters are primed for responsiveness. Unexpectedly, we found only limited evidence to support this proposition because the ATRA response was largely unaffected, even from H3K4me3-denuded bivalent promoters. However, a minority of bivalent promoters appeared to be impaired for the ATRA response. Potentially, the impaired response of even a few genes involved in a complex lineage commitment programme could derail development and thereby explain the developmental and growth retardations observed in *Mll2*^{-/-} embryos, which die at ~E10.5 (Glaser et al., 2006). Because ATRA induction of ESCs is not physiological or a recapitulation of embryonic events, further work is required to determine if our *in vitro* observations can be extrapolated to the embryo. However, we previously reported that *Mll2*^{-/-} ESCs are impaired for neural differentiation (Lubitz et al., 2007) and further work has revealed that the expression of only a few genes is significantly perturbed (K. Neumann, K.A. and A.F.S., unpublished).

Because many bivalent promoters showed normal ATRA responsiveness in the absence of H3K4me3, rather than being primed for lineage-specific activation, we suggest a different explanation for bivalency. The idea that CpG islands can attract epigenetic regulators without assistance from sequence-specific DNA-binding proteins (i.e. transcription factors) has recently been proposed (Blackledge and Klose, 2011; Blackledge et al., 2010; Glaser et al., 2009; Ku et al., 2008; Mendenhall et al., 2010; Thomson et al., 2010). This dogma-breaking concept has been most concisely embodied in the suggestion that *Cxxc1*, via its CxxC domain, which binds to unmethylated CpG dinucleotides (Lee et al., 2001), positions *Set1C* on CpG islands (Clouaire et al., 2012; Thomson et al., 2010). Our evidence suggests a modified version of this idea. Rather than *Cxxc1/Set1C*, which we find to be concentrated on active but not bivalent promoters, *Mll2* with its CxxC domain appears to be a pioneer in the epigenetic definition of CpG islands as TSSs. With respect to bivalent promoters, we suggest that H3K27me3 and PcG repression serves to prevent the maturation of CpG islands into *Set1C*-bound active promoters (Fig. 7).

The proposition that *Mll2* plays an early role in defining the epigenetic status of ESCs is supported by its requirement during oogenesis and the early cleavage stages after fertilisation. During oogenesis, *Mll2* is the major H3K4 tri-, but not mono-, methyltransferase and expression from the paternal allele lessens the

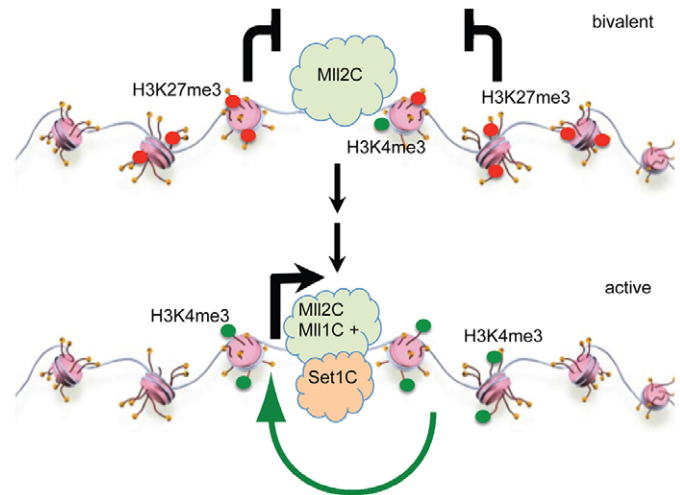


Fig. 7. Model for the relationships between *Mll* and *Set1* complexes on bivalent and active promoters. The *Mll2* complex (*Mll2C*) deposits H3K4me3 (green dots) on a TSS nucleosome but is prevented from further action by PcG, which deposits H3K27me3 (red dots) on surrounding histone H3 tails. Maturation of the promoter to active transcription involves the removal of H3K27me3 and the recruitment of other H3K4 methyltransferase complexes, notably *Set1C*, and increased H3K4me3. The green arrow represents feed-forward epigenetic maintenance.

deleterious impact of a mutant maternal allele before blastocyst formation (Andreu-Vieyra et al., 2010). As a CxxC domain protein and the major H3K4 trimethyltransferase during late oogenesis and early development, *Mll2* is the prime candidate to define CpG islands as potential promoters in the naïve epigenome. At some point in early development, presumably before the blastocyst, *Mll2* stops being the major H3K4 trimethyltransferase and is not again required until after gastrulation, despite continuous expression (Glaser et al., 2006).

Although *Mll1* is not required in development until definitive haematopoiesis (Ernst et al., 2004; Yagi et al., 1998), our analyses indicate that it contributes to the H3K4 methylation system in mESCs. By examining mESCs doubly mutated for *Mll1* and *Mll2*, a further depth to the complexities of H3K4 methylation and interplay between the six *Set1*/Trithorax-type methyltransferases was revealed. What is the purpose of this complexity? Potentially, it establishes a multiply redundant system. If so, the redundancy appears to be imperfect and the causes of the imperfections are not obvious. Alternatively, a multiply backed-up system could be mutually reinforcing and achieve epigenetic momentum to ensure that, once a promoter becomes activated, it will remain active without reliance on the continuous reiteration of regulatory signalling. We previously proposed this mechanism for housekeeping gene promoters as well as for the maintenance of activation after lineage-specific decisions (Glaser et al., 2009). In this scenario, housekeeping gene promoters, which are invariably CpG islands, remain active by epigenetic maintenance, as do certain lineage-specific promoters that become recruited to the same status.

MATERIALS AND METHODS

General methods

Recombineering, gene targeting, BAC transgenesis, chromatin immunoprecipitation (ChIP) and generic affinity purification using the GFP tag were performed as previously described (Fu et al., 2010; Hofmeister et al., 2011).

ESC methods

The *Mil2*^{-/-} (Lubitz et al., 2007) and *Mil1*^{-/-} ESCs were derived from E14TG2a by consecutive rounds of gene targeting using the same targeting construct for the second allele after exchanging the neomycin resistance gene for hygromycin by recombineering. The *Mil2F/F;CreERT2*⁺, *Mil1F/F;CreERT2*⁺ and *Mil1F/F;Mil2F/F;CreERT2*⁺ ESCs were established *de novo* from blastocysts after germline transmission of targeted R1 ESCs and crossing to C57Bl/6 at least five times, including a cross to a C57Bl/6 *Rosa26CreERT2* mouse line (Seibler et al., 2003). All ESCs were cultured with fetal calf serum (FCS)-based medium [DMEM + GlutaMAX (Invitrogen), 15% FCS (PAA), 2 mM L-glutamine (Invitrogen), 1× non-essential amino acids (Invitrogen), 1 mM sodium pyruvate (Invitrogen), 0.1 mM β-mercaptoethanol] in the presence of LIF. For ATRA inductions, the LIF was replaced with 0.1 μM all-trans retinoic acid (Sigma).

RNA-Seq and ChIP-Seq

Total RNA was isolated from 10⁷ cells using Trizol reagent, poly(A) fractions were purified by double selection on oligo(dT) beads and cDNA prepared as described previously (Marks et al., 2012). Sequencing libraries for cDNA and ChIP samples were prepared using Illumina kits and processed on Illumina GA2 and High-Seq instruments. Single 36 nt reads were aligned to mm9 (*Mus musculus* 2007) genome assembly and analysed by SAMtools (samtools.sourceforge.net) and BAMtools (sourceforge.net/projects/bamtools). Sequencing data have been deposited in NCBI GEO SuperSeries with accession GSE52071.

Peaks on the ChIP-seq data were defined using the FindPeaks tool from the Vancouver Short Read Analysis Package (vancouvershortr.sourceforge.net). Levels of H3K4me3 and H3K27me3 at promoters were calculated as number of reads per 3 kb promoter regions centred at the TSS. Expression and histone modifications were normalised to total read counts. Bivalency was assigned to promoters covered by >30 and >45 reads of H3K4me3 and H3K27me3, respectively. Low levels (<30 reads per promoter) of H3K4me3 were excluded from the analysis of *Mil2*^{-/-} affected genes. Gene expression was counted as the number of reads per 1 kb mRNA. Gene ontology (GO) and pathway analyses were performed by DAVID (david.abcc.ncifcrf.gov).

To compensate for differences in sequencing depth and mapping efficiency, the total number of unique reads of each sample was uniformly equalised, allowing quantitative comparisons, as for the heatmaps (Fig. 3C). The number of tags per 80 bp bin was calculated, aligned on the maximum signal in the H3K4me3 profile of E14 ESCs, and plotted essentially as described (Marks et al., 2012). For quantification of the heatmaps using line plots (Fig. 3D), we performed a moving average (50 sites per bin) on the profiles shown in the heatmap (Fig. 3C). For the GFP-tagged subunits, the negative control based on GFP antibody ChIP using untagged E14 ESCs was subtracted. Data were normalised to compensate for differences in ChIP efficiencies for the various profiles and the maximum value was set to 100%.

Antibodies

Antibodies used for western blot analysis were: goat anti-GFP (obtained from MPI-CBG, Dresden); anti-Ash2l (A300-112A-1, Bethyl Laboratories); anti-Nanog (sc1000, Calbiochem); anti-β-actin (A5441, Sigma); anti-Cfp1/CXXC1 (H-120, Santa Cruz); anti-Mil1 (A300-086A, Bethyl Laboratories); anti-H3K4me3 (ab8580, Abcam); anti-H3K4me2 (ab32356, Abcam); anti-H3K4me1 (ab8895, Abcam); anti-H3pan (07-690, Millipore); and anti-Mil2 [as described previously (Glaser et al., 2006)]. For westerns, primary antibodies were detected using secondary HRP-coupled antibodies (Pierce or ThermoFisher).

Note on nomenclature

All published references to *Mil2* before 2006 refer to the gene that is called *Mil2* in this article, which in the mouse is also called *Wbp7*. This gene is on human chromosome 19 and mouse chromosome 7. Since 2006, *Mil2* has also been used in publications for the gene first named *ALR* and sometimes called *Mil4*. This gene is on human chromosome 15 and mouse chromosome 12. An attempt to establish a new nomenclature (Allis et al., 2007) has not resolved the problem because both genes have been called either *Kmt2b* or *Kmt2d*.

Acknowledgements

We thank Michelle Meredyth, Davi Torres Coe, Ashish Gupta and Katrin Neumann for discussions.

Competing interests

The authors declare no competing financial interests.

Author contributions

S.D. and H.M. conceived and performed the experiments and analysed the data; H.H. generated the cell lines, conceived and performed the experiments; A.K., G.C. and K.A. generated the cell lines; S.S. analysed the data; H.G.S. conceived the experiments and analysed the data; A.F.S. conceived the experiments, analysed the data and wrote the paper.

Funding

This work was supported by funding from the EU 6th Framework Program HEROIC to H.G.S. and A.F.S.; the EU 7th Framework Program SyBoSS to A.F.S.; the Deutsche Forschungsgemeinschaft (DFG) SPP1356 Program Pluripotency and Cellular Reprogramming to K.A. and A.F.S.; and a DFG grant [KR2154/3-1] to A.K. and A.F.S. Deposited in PMC for immediate release.

Supplementary material

Supplementary material available online at <http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.102681/-/DC1>

References

- Akkers, R. C., van Heeringen, S. J., Jacobi, U. G., Janssen-Megens, E. M., François, K.-J., Stunnenberg, H. G. and Veenstra, G. J. C. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in *Xenopus* embryos. *Dev. Cell* **17**, 425-434.
- Albert, M. and Helin, K. (2010). Histone methyltransferases in cancer. *Semin. Cell Dev. Biol.* **21**, 209-220.
- Allis, C. D., Berger, S. L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhata, R. et al. (2007). New nomenclature for chromatin-modifying enzymes. *Cell* **131**, 633-636.
- Andreu-Vieyra, C. V., Chen, R., Agno, J. E., Glaser, S., Anastassiadis, K., Stewart, A. F. and Matzuk, M. M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS Biol.* **8**, doi: 10.1371/journal.pbio.1000453.
- Ardehali, M. B., Mei, A., Zobeck, K. L., Caron, M., Lis, J. T. and Kusch, T. (2011). *Drosophila* Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *EMBO J.* **30**, 2817-2828.
- Austena, L., Barozzi, I., Chronowska, A., Termanini, A., Ostuni, R., Prosperini, E., Stewart, A. F., Testa, G. and Natoli, G. (2012). The histone methyltransferase Wbp7 controls macrophage function through GPI glycolipid anchor synthesis. *Immunity* **36**, 572-585.
- Azura, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merckenschlager, M. et al. (2006). Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8**, 532-538.
- Bannister, A. J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell. Res.* **21**, 381-395.
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K. et al. (2006a). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326.
- Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E. and Allis, C. D. (2006b). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.* **26**, 2560-2569.
- Blackledge, N. P. and Klose, R. (2011). CpG island chromatin: a platform for gene regulation. *Epigenetics* **6**, 147-152.
- Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J. and Klose, R. J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell* **38**, 179-190.
- Brock, H. W. and Fisher, C. L. (2005). Maintenance of gene expression patterns. *Dev. Dyn.* **232**, 633-655.
- Butler, J. S., Koutelou, E., Schibler, A. C. and Dent, S. Y. R. (2012). Histone-modifying enzymes: regulators of developmental decisions and drivers of human disease. *Epigenomics* **4**, 163-177.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043.
- Champagne, K. S. and Kutateladze, T. G. (2009). Structural insight into histone recognition by the ING PHD fingers. *Curr. Drug Targets* **10**, 432-441.
- Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J. H., Skalnik, D. and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev.* **26**, 1714-1728.
- Eberl, H. C., Spruijt, C. G., Kelstrup, C. D., Vermeulen, M. and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. *Mol. Cell* **49**, 368-378.

- Ernst, P., Fisher, J. K., Avery, W., Wade, S., Foy, D. and Korsmeyer, S. J. (2004). Definitive hematopoiesis requires the mixed-lineage leukemia gene. *Dev. Cell* **6**, 437-443.
- Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D. and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870-1881.
- Fisher, C. L. and Fisher, A. G. (2011). Chromatin states in pluripotent, differentiated, and reprogrammed cells. *Curr. Opin. Genet. Dev.* **21**, 140-146.
- FitzGerald, K. T. and Diaz, M. O. (1999). MLL2: A new mammalian member of the trx/MLL family of genes. *Genomics* **59**, 187-192.
- Fu, J., Teucher, M., Anastassiadis, K., Skarnes, W. and Stewart, A. F. (2010). A recombinering pipeline to make conditional targeting constructs. *Methods Enzymol.* **477**, 125-144.
- Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tuftefeld, K. R., Aasland, R., Anastassiadis, K., Ang, S.-L. and Stewart, A. F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133**, 1423-1432.
- Glaser, S., Lubitz, S., Loveland, K. L., Ohho, K., Robb, L., Schwenk, F., Seibler, J., Roellig, D., Kranz, A., Anastassiadis, K. et al. (2009). The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics Chromatin* **2**, 5.
- Goldberg, A. D., Allis, C. D. and Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* **128**, 635-638.
- Guenther, M. G., Jenner, R. G., Chevalier, B., Nakamura, T., Croce, C. M., Canaani, E. and Young, R. A. (2005). Global and Hox-specific roles for the MLL1 methyltransferase. *Proc. Natl. Acad. Sci. USA* **102**, 8603-8608.
- Guo, C., Chang, C.-C., Wortham, M., Chen, L. H., Kernagis, D. N., Qin, X., Cho, Y.-W., Chi, J.-T., Grant, G. A., McLendon, R. E. et al. (2012). Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. *Proc. Natl. Acad. Sci. USA* **109**, 17603-17608.
- Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., Rappsilber, J., Lerdrup, M. and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**, 1291-1300.
- Herz, H.-M., Nakanishi, S. and Shilatifard, A. (2009). The curious case of bivalent marks. *Dev. Cell* **17**, 301-303.
- Hofemeister, H., Ciotta, G., Fu, J., Seibert, P. M., Schulz, A., Maresca, M., Sarov, M., Anastassiadis, K. and Stewart, A. F. (2011). Recombinering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. *Methods* **53**, 437-452.
- Hsieh, J. J. D., Ernst, P., Erdjument-Bromage, H., Tempst, P. and Korsmeyer, S. J. (2003a). Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol. Cell. Biol.* **23**, 186-194.
- Hsieh, J. J.-D., Cheng, E. H.-Y. and Korsmeyer, S. J. (2003b). Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* **115**, 293-303.
- Huang, Y., Fang, J., Bedford, M. T., Zhang, Y. and Xu, R. M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**, 748-751.
- Hubner, N. C., Bird, A. W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A. and Mann, M. (2010). Quantitative proteomics combined with BAC Transgenomics reveals in vivo protein interactions. *J. Cell Biol.* **189**, 739-754.
- Huntsman, D. G., Chin, S. F., Muleris, M., Batley, S. J., Collins, V. P., Wiedemann, L. M., Aparicio, S. and Caldas, C. (1999). MLL2, the second human homolog of the *Drosophila* trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines. *Oncogene* **18**, 7975-7984.
- Jaenisch, R. and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33**, 245-254.
- Klymenko, T. and Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep.* **5**, 373-377.
- Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S. et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242.
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16**, 2893-2905.
- Ladopoulos, V., Hofemeister, H., Hoogenkamp, M., Riggs, A. D., Stewart, A. F. and Bonifer, C. (2013). The histone methyltransferase KMT2B is required for RNA polymerase II association and protection from DNA methylation at the MagohB CpG island promoter. *Mol. Cell. Biol.* **33**, 1383-1393.
- Lee, J. H., Voo, K. S. and Skalinik, D. G. (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. *J. Biol. Chem.* **276**, 44669-44676.
- Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B. E. and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* **24**, 368-380.
- Lubitz, S., Glaser, S., Schaft, J., Stewart, A. F. and Anastassiadis, K. (2007). Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Mol. Biol. Cell* **18**, 2356-2366.
- Magnúsdóttir, E., Gillich, A., Grabole, N. and Surani, M. A. (2012). Combinatorial control of cell fate and reprogramming in the mammalian germline. *Curr. Opin. Genet. Dev.* **22**, 466-474.
- Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury, W. J., 3rd, Voigt, P., Martin, S. R., Taylor, W. R., De Marco, V. et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762-767.
- Marks, H., Kalkan, T., Menafrá, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A. F., Smith, A. et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590-604.
- Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., Ku, M. and Bernstein, B. E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* **6**, e1001244.
- Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R. P. et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560.
- Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D. and Hess, J. L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol. Cell* **10**, 1107-1117.
- Milne, T. A., Kim, J., Wang, G. G., Stadler, S. C., Basrur, V., Whitcomb, S. J., Wang, Z., Ruthenburg, A. J., Elenitoba-Johnson, K. S. J., Roeder, R. G. et al. (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. *Mol. Cell* **38**, 853-863.
- Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E. and Simon, J. A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197-208.
- Nowak, D. E., Tian, B. and Brasier, A. R. (2005). Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *Biotechniques* **39**, 715-725.
- Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C. D. et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-717.
- Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425-432.
- Ringrose, L. and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38**, 413-443.
- Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R. and Stewart, A. F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.* **20**, 7137-7148.
- Ruthenburg, A. J., Allis, C. D. and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25**, 15-30.
- Scacheri, P. C., Davis, S., Odom, D. T., Crawford, G. E., Perkins, S., Halawi, M. J., Agarwal, S. K., Marx, S. J., Spiegel, A. M., Meltzer, P. S. et al. (2006). Genomewide analysis of menin binding provides insights into MEN1 tumorigenesis. *PLoS Genet.* **2**, e51.
- Schmitges, F. W., Prusty, A. B., Faty, M., Stützer, A., Lingaraju, G. M., Aiwanian, J., Sack, R., Hess, D., Li, L., Zhou, S. et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* **42**, 330-341.
- Seibler, J., Zevnik, B., Küter-Luks, B., Andreas, S., Kern, H., Hennek, T., Rode, A., Heimann, C., Faust, N., Kauselmann, G. et al. (2003). Rapid generation of inducible mouse mutants. *Nucleic Acids Res.* **31**, e12.
- Shi, X., Kachirskaia, I., Walter, K. L., Kuo, J.-H. A., Lake, A., Davrazou, F., Chan, S. M., Martin, D. G. E., Fingerman, I. M., Briggs, S. D. et al. (2007). Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J. Biol. Chem.* **282**, 2450-2455.
- Simon, J. A. and Tamkun, J. W. (2002). Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. *Curr. Opin. Genet. Dev.* **12**, 210-218.
- Sims, R. J., 3rd, Millhouse, S., Chen, C.-F., Lewis, B. A., Erdjument-Bromage, H., Tempst, P., Manley, J. L. and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol. Cell* **28**, 665-676.
- Suganuma, T. and Workman, J. L. (2011). Signals and combinatorial functions of histone modifications. *Annu. Rev. Biochem.* **80**, 473-499.
- Testa, G., Zhang, Y., Vintersten, K., Benes, V., Pijnappel, W. W. M. P., Chambers, I., Smith, A. J. H., Smith, A. G. and Stewart, A. F. (2003). Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nat. Biotechnol.* **21**, 443-447.
- Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R. W., Deaton, A., Andrews, R., James, K. D. et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082-1086.
- van den Berg, D. L. C., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I. and Poot, R. A. (2008). Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. *Mol. Cell. Biol.* **28**, 5986-5995.
- Vastenhouw, N. L. and Schier, A. F. (2012). Bivalent histone modifications in early embryogenesis. *Curr. Opin. Cell Biol.* **24**, 374-386.
- Vastenhouw, N. L., Zhang, Y., Woods, I. G., Imam, F., Regev, A., Liu, X. S., Rinn, J. and Schier, A. F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464**, 922-926.
- Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W. M. P., van Schaik, F. M. A., Varier, R. A., Baltissen, M. P. A., Stunnenberg, H. G., Mann, M. and Timmers, H. T. M. (2007). Selective anchoring of TFIIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58-69.

- Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G. et al. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967-980.
- Voigt, P., LeRoy, G., Drury, W. J., 3rd, Zee, B. M., Son, J., Beck, D. B., Young, N. L., Garcia, B. A. and Reinberg, D. (2012). Asymmetrically modified nucleosomes. *Cell* **151**, 181-193.
- Voigt, P., Tee, W. W. and Reinberg, D. (2013). A double take on bivalent promoters. *Genes Dev.* **27**, 1318-1338.
- Wan, M., Liang, J., Xiong, Y., Shi, F., Zhang, Y., Lu, W., He, Q., Yang, D., Chen, R., Liu, D. et al. (2013). The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells. *J. Biol. Chem.* **288**, 5039-5048.
- Wang, P., Lin, C., Smith, E. R., Guo, H., Sanderson, B. W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L. M. et al. (2009). Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol. Cell. Biol.* **29**, 6074-6085.
- Wray, J., Kalkan, T. and Smith, A. G. (2010). The ground state of pluripotency. *Biochem. Soc. Trans.* **38**, 1027-1032.
- Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P. et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90.
- Xu, H., Handoko, L., Wei, X., Ye, C., Sheng, J., Wei, C. L., Lin, F. and Sung, W. K. (2010). A signal-noise model for significance analysis of ChIP-seq with negative control. *Bioinformatics* **26**, 1199-1204.
- Yagi, H., Deguchi, K., Aono, A., Tani, Y., Kishimoto, T. and Komori, T. (1998). Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. *Blood* **92**, 108-117.
- Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D. J., Kitabayashi, I., Herr, W. and Cleary, M. L. (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.
- Zegerman, P., Canas, B., Pappin, D. and Kouzarides, T. (2002). Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *J. Biol. Chem.* **277**, 11621-11624.