Histone demethylome map reveals combinatorial gene regulatory functions in embryonic stem cells

Yogesh Kumar, [1,12] **Pratibha Tripathi,** [1,12] **Majid Mehravar,** [1] **Michael J. Bullen,** [1] **Varun K. Pandey,** [1] **Dhaval Hathiwala,** [1] **Marc Kerenyi,** [2] **Andrew Woo,** [3,4] **Alireza Ghamari,** [5] **Alan B. Cantor,** [5] **Lee H. Wong,** [6] **Jonghwan Kim,** [7] **Kimberly Glass,** [8] **Guo-Cheng Yuan,** [9] **Luca Pinello,** [10,13, *] **Stuart H. Orkin,**[5,11,13 *] **Partha Pratim Das** [1,13,14 *]

[1]Department of Anatomy and Developmental Biology, Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Monash University, Wellington Road, Clayton, VIC 3800, Australia

[2]AstraZeneca, Landstraßer Hauptstraße 1a, 1030 Vienna, Austria

[3]Harry Perkins Institute of Medical Research, the University of Western Australia, Perth, Western Australia, 6009, Australia

[4]School of Medical and Health Sciences, Edith Cowan University, Joondalup, WA, 6027, Australia

[5]Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute (DFCI), Harvard Stem Cell Institute, Harvard Medical School, Boston, MA 02115, USA

[6]Department of Biochemistry and Molecular Biology, Cancer Program, Biomedicine Discovery Institute, Monash University, Wellington Road, Clayton, Victoria, 3800, Australia

[7]Department of Molecular Biosciences, Institute for Cellular and Molecular Biology, Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, TX, USA

[8]Channing Division of Network Medicine, Brigham and Women's Hospital, and Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

[9]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, MA 02115, USA

[10]Molecular Pathology & Cancer Center, Massachusetts General Hospital & Harvard Medical School, Boston, MA 02114, USA

[11]Howard Hughes Medical Institute, Boston, MA 02115, USA

[12]These authors contributed equally to this work
[13]Senior authors
[14]Lead contact
*Correspondence:partha.das@monash.edu(P.P.D);
orkin@bloodgroup.tch.harvard.edu (S.H.O.); lpinello@mgh.harvard.edu (L.P)

SUMMARY

Epigenetic regulators and transcription factors establish distinct regulatory networks for gene regulation to maintain the embryonic stem cells (ESC) state. Although much has been learned regarding individual epigenetic regulators, their combinatorial functions remain elusive. Here, we report combinatorial functions of histone demethylases (HDMs) in gene regulation of mouse ESCs that are currently unknown. We generated a histone demethylome (HDMome) map of 20 well-characterized HDMs based on their genome-wide binding. This revealed co-occupancy of HDMs in different combinations; predominantly, KDM1A-KDM4B-KDM6A and JARID2-KDM4A-KDM4C-KDM5B co-occupy at enhancers and promoters, respectively. Comprehensive mechanistic studies uncover that KDM1A-KDM6A combinatorially modulates P300/H3K27ac, H3K4me1, H3K4me2 deposition and OCT4 recruitment that eventually directs the OCT4/CORE regulatory network for target gene expression; while co-operative actions of JARID2-KDM4A-KDM4C-KDM5B control H2AK119ub1 and bivalent marks of polycomb-repressive complexes that facilitates the PRC regulatory network for target gene repression. Thus, combinatorial functions of HDMs impact gene expression programs to maintain the ESC state.

INTRODUCTION

Embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) have a tremendous capacity for regenerative medicine (Inoue and Yamanaka, 2011; Robinton and Daley, 2012). These stem cells are molecularly and functionally equivalent to each other. ESCs are derived from the inner cell mass of the blastocyst during embryonic development (Thomson, 1998), whereas iPSCs are derived from any cell types through the ectopic expression of specific factors (Takahashi and Yamanaka, 2006). Both ESCs and iPSCs have two unique properties: self-renewal (the ability to proliferate in the same state indefinitely) and pluripotency (the capacity to differentiate towards all lineages of the organism).

The self-renewal and pluripotency of the "ESC state" are established and maintained through several transcriptional regulatory networks (Young, 2011). Each of these regulatory networks encompasses several ESC-specific transcription factors (TFs), co-factors and chromatin/epigenetic regulators to control chromatin organisation and gene regulation (Orkin and Hochedlinger, 2011). Comprehensive studies using genome-wide binding of these ESC-TFs, co-factors and epigenetic regulators and their link to target gene expression revealed three main functionally distinct regulatory networks in mouse ESCs (mESCs) that maintain the ESC state: CORE (active), MYC (active) and polycomb/PRC (repressive) (Chen et al., 2008; Das et al., 2014; Kim et al., 2008; Seruggia et al., 2019). The CORE network comprises the core ESC-TFs, such as OCT4, SOX2 and NANOG (OSN); this OSN-trio recruits/binds to other ESC-TFs, co-activators, and chromatin/epigenetic regulators to activate ESC-specific genes and repress lineage-specific genes. The MYC network is c-MYC-centred, and co-occupies with its partners, mainly co-activators and epigenetic regulators, and predominantly regulates gene activation. In contrast, the polycomb

repressive complex (PRC) network is composed of both PRC1 and PRC2 components and mainly represses lineage-specific genes in mESCs (Das et al., 2014; Kim et al., 2010). Moreover, a handful of studies focused on numerous epigenetic/chromatin regulators demonstrated that multiple epigenetic regulators co-occupy in specific combinations to create modules that exhibit distinct chromatin/ histone modification states for gene regulation in different cell types (Filion et al., 2010; Ram et al., 2011; Venters et al., 2011). Nonetheless, these findings fail to provide mechanistic insights into how multiple epigenetic regulators control gene expression programs. Further, the majority of histone demethylases (HDMs) were omitted from these studies.

HDMs are a class of epigenetic regulators, which consist of 20 well-characterized individual members that "remove" methyl group(s) from specific lysine (K) and arginine (R) residue(s) on histone tails to modulate chromatin structure for the gene regulation (Kooistra and Helin, 2012). Notably, the demethylase activity of HDMs is more focused on lysine (K) than arginine (R) residues; thus, HDMs are often called lysine (K) demethylases – KDMs. HDMs are divided into two broad classes based on the mechanisms by which they demethylate their substrates/histone marks: LSD-HDMs and JMJC-HDMs (Kooistra and Helin, 2012; Mosammaparast and Shi, 2010). LSD1/KDM1A and LSD2/KDM1B are the two members of the LSD-HDM class. The remaining HDM members belong to the JMJC-HDM class and they are further sub-classified based on their sequence homology, domains and substrate specificity; namely, KDM2, KDM3, KDM4, KDM5 and KDM6 (Agger et al., 2008; Kooistra and Helin, 2012; Pedersen and Helin, 2010). Each of the HDMs plays critical roles in controlling numerous biological processes during development and diseases (Pedersen and Helin, 2010). Likewise, several individual HDMs play critical roles in controlling ESC state (Adamo et al., 2011; Das et al., 2014; He et al., 2013; Liang et

al., 2012; Pedersen et al., 2016; Welstead et al., 2012; Whyte et al., 2012; Xie et al., 2011; 2014). However, the “combinatorial” functions of HDMs have not been explored in a comprehensive manner, including an ESC context. Moreover, no datasets and mechanistic studies are currently available to support this hypothesis. Only a handful of previous studies demonstrated the combinatorial functions of a few selected HDMs in ESCs (Das et al., 2014; Pedersen et al., 2016). Therefore, here we have expanded the scope to interrogate combinatorial functions of “all” 20 well-characterized HDMs in gene regulation in mESCs.

In this study, we report genome-wide occupancy/binding data of 20 HDMs, a “histone demethylome (HDMome)” map for the first-time. This analysis revealed that “multiple” HDMs share the same binding sites at enhancers and promoters at the genome-wide level. Predominantly, KDM1A-KDM4B-KDM6A co-occupy enhancer regions and JARID2-KDM4A-KDM4C-KDM5B co-occupy promoter regions. Comprehensive computational and biochemical analyses demonstrate that KDM1A-KDM4B-KDM6A collaborates with ESC-TFs and belongs to the CORE network (active), whereas JARID2-KDM4A-KDM4C-KDM5B functions with the polycomb repressive complexes 1 and 2 (PRC1 and PRC2) and belongs to the PRC network (repressive). Furthermore, thorough molecular studies exhibit that KDM1A and KDM6A combinatorially modulate P300/H3K27ac, H3K4me1, H3K4me2 deposition and OCT4 recruitment at enhancers to direct the CORE regulatory network for target gene expression. In contrast, JARID2, KDM4A, KDM4C and KDM5B co-operatively regulate H2AK119ub1 (of PRC1) and bivalent (H3K27me3-H3K4me3) marks (of PRC2) at promoters and enables the PRC regulatory network for target genes repression. Hence, our findings reveal the combinatorial functions of HDMs, which coordinate proper gene expression programs for the maintenance of the mESC state.

RESULTS

The HDMome signature reveals combinatorial co-occupancy of multiple HDMs

To determine whether HDMs act combinatorially in gene regulation, we conducted ChIP-seq (for antibodies specific to HDM/s) and/or *in vivo* biotinylation-mediated ChIP-seq (Bio-ChIP-seq) (Das et al., 2014; Kim et al., 2009) (for Flag-Biotin (FB)-tagged HDM mESC line/s) of all 20 well-characterized individual HDMs (Figures S1A; Table S1). Of note, most of these HDMs are expressed in mESCs (Das et al., 2014). The ChIP-seq and/or Bio-ChIP-seq generated genome-wide occupancy/binding data of all 20 HDMs in mESCs to create a "histone demethylome (HDMome)" map (Figures 1A, 1B, S1A; Table S1), which were further used to assess the combinatorial binding of multiple HDMs. To understand the significance of the combinatorial binding of multiple HDMs, we examined their "co-occupancy" particularly at the "enhancers" and "promoters", as these regulatory elements directly involved in gene regulation (Elkon and Agami, 2017; Whyte et al., 2013). Mouse ESC-specific 8,794 enhancers that were identified based on co-occupancy of ESC-specific TFs (OCT4, NANOG, SOX2, KLF4, ESRRB), mediators (MED1), enhancer histone marks (H3K4me1, H3K27ac) and DNase I hypersensitivity (Whyte et al., 2013); promoters that were defined as ±2kb of TSS (transcription start sites), used for our analyses. ChIP-seq peaks (the top 5,000 peaks based on intensities) of all HDMs were plotted within these enhancers and promoters to assess their co-binding. We observed co-occupancy of KDM1A-KDM4B-KDM6A mainly at enhancers (Figures 1A, 1C, S1B, S1E); whereas JARID2-KDM4A-KDM4C-KDM5B and KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2) largely co-occupied at promoters (Figures 1B, 1D, 1E, S1C). However, a very few overlapping regions were detected for KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-

PHF8 (without JARID2) at promoters (Figure S1D). Taken together, the HDMome data suggest that combinatorial co-occupancy of multiple HDMs occurs in mESCs.

HDMs combinatorially regulate gene expression

To determine the importance of the co-occupancy of KDM1A-KDM4B-KDM6A, JARID2-KDM4A-KDM4C-KDM5B and KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2), we correlated their co-occupancy with active (H3K27ac, H3K4me1, H3K4me2, H3K4me3, H3K36me3) and repressive (H2AK119ub1, H3K27me3, H3K9me3) histone marks (Zhou et al., 2010). As KDM1A-KDM4B-KDM6A frequently co-occupied enhancers, we found that KDM1A-KDM4B-KDM6A co-binding regions largely overlapped with active H3K27ac and H3K4me1 enhancer marks (Figure 2A). In contrast, JARID2-KDM4A-KDM4C-KDM5B co-binding regions overlapped repressive H2AK119ub1 and bivalent (H3K27me3 and H3K4me2, me3) histone promoter marks (Figure 2B). Moreover, KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2) co-occupied regions predominantly overlapping the active H3K4me3 promoter mark (Figure 2C).

In addition, gene expression analysis showed that the target genes (closest genes) that are co-occupied by KDM1A-KDM4B-KDM6A and KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2) were largely down-regulated upon differentiation; in contrast, JARID2-KDM4A-KDM4C-KDM5B co-occupied target genes were largely up-regulated upon differentiation (Figure 2D; Table S2). Collectively, gene expression data (differentiated state v. undifferentiated state of mESCs) and binding data of both HDMs and histone marks indicate that co-occupancy of KDM1A-KDM4B-KDM6A and KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8

(without JARID2) is related to gene activation, while co-occupancy of JARID2-KDM4A-KDM4C-KDM5B is related to gene repression.

To define the possible biological functions of KDM1A-KDM4B-KDM6A, JARID2-KDM4A-KDM4C-KDM5B and KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2), we used their target genes and performed GO (Gene Ontology) analysis. KDM1A-KDM4B-KDM6A co-occupied target genes enriched for transcription, regulation of cell proliferation, the stem cell population maintenance and cell differentiation (Figure 2E) processes related to transcriptional gene regulation for the maintenance of the mESC state. JARID2-KDM4A-KDM4C-KDM5B co-occupied target genes enriched for transcription, multicellular organism development, nervous system development, Wnt signalling pathway and circadian genes regulation (Figure 2F) processes related to repression of developmental genes that is also critical for maintaining the mESC state. Further, KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2) co-occupied target genes that did not exhibit biological functions apparently relevant to maintenance of the mESCs state. Perhaps, they relate to housekeeping functions of mESCs (Figure S2A). Taken together, these data imply that HDMs may act in different combinations to regulate either gene activation or repression in mESCs.

Combinatorial gene regulatory functions of KDM4 and KDM5 members

As the KDM4 and KDM5 are the largest sub-classes of HDMs (KDM4 sub-class: KDM4A, 4B, 4C, 4D; KDM5 sub-class: KDM5A, 5B, 5C, 5D), they can be used to investigate their combinatorial and overlapping functions. Our previous study demonstrated that KDM4B and KDM4C function combinatorially through different ESC regulatory networks to maintain the mouse ESC state (Das et al., 2014). Further, a

separate study showed that KDM4A and KDM4C have overlapping functions to maintain the self-renewal of mESCs (Pedersen et al., 2016). Here we examine their combinatorial gene regulatory functions in a more comprehensive fashion.

We first mapped all the binding sites of KDM4A, KDM4B and KDM4C at enhancers (mESC-specific 8,794 enhancers) and promoters (±2kb of TSS) that exhibited the unique binding sites of KDM4A, KDM4B and KDM4C, as well as their common binding sites in different combinations (Figures 3A-3C, S3A, S3B). Next, we compared these binding sites to several relevant active (H3K27ac, H3K4me1, H3K4me3) and repressive (H2AK119ub1, H3K27me3, H3K9me3) histone marks at both enhancers and promoters to examine the significance of the combinatorial binding of KDM4A, KDM4B and KDM4C. We excluded KDM4D, as it does not bind significantly at enhancers or promoters (Figures 1A, 1B). Enhancers encompasses largely KDM4B unique targets, which overlapped with active H3K27ac and H3K4me1 enhancer marks, and correlated with gene activation (Figures 3A, 3D, 3G, S3A) (Das et al., 2014). On the other hand, promoter regions included largely KDM4B unique, KDM4C unique, KDM4A-4C common and KDM4B-4C common targets. Specifically, KDM4C unique and KDM4A-4C common targets were predominant, co-occupied with repressive H2AK119ub1 and bivalent (H3K27me3-H3K4me3) marks, and correlated with gene repression (Figures 3B, 3E, 3H, 3J). Likewise, KDM4A unique targets were also co-occupied with repressive H2AK119ub1 and bivalent (H3K27me3-H3K4me3) marks and were associated to gene repression (Figures 3B, 3E). Also, we found that the majority of targets of KDM4A were also targets of KDM4C at promoters (Figure S3B). Collectively, these data hint that KDM4A and KDM4C function together in gene repression and have overlapping functions. In contrast, KDM4B unique and KDM4B-4C common targets largely overlapped with active H3K4me3 mark and were linked to

gene expression at different degrees (Figures 3B, 3E, S3C, S3D). Furthermore, KDM4A-4B-4C common targets overlapped with the active H3K4me3 mark, although these target genes did not show any significant link to either gene activation or repression (Figures 3B, 3E). Interestingly, we did not observe significant overlap between the KDM4A, KDM4B, and KDM4C targets and H3K9me3 (the substrate of KDM4 members). Taken together, these data suggest that KDM4A and KDM4C function together in gene repression, while KDM4B functions in gene activation; nonetheless, the presence of KDM4B de-represses the gene regulatory functions of KDM4A and KDM4C in mESCs.

Similar analyses were performed for KDM5 members, including KDM5A, KDM5B and KDM5C. KDM5D was excluded, as it does not bind significantly to enhancers or promoters (Figure 1A, 1B)). Since, very few binding sites of KDM5 members were found at enhancers (Figures 1A, S3E), we focused our analyses at promoters (Figures 1B, S3F) and found that promoters mainly included KDM5A unique, KDM5B unique and KDM5A-5B common targets (Figures 3C, S3F). KDM5A unique targets mostly overlapped with active H3K4me3 mark and were related to gene activation (Figures 3C, 3F, 3I), whereas KDM5B unique targets mainly overlapped with repressive H2AK119ub1 and bivalent (H3K27me3-H3K4me3) marks and were connected to gene repression (Figures 3C, 3F, 3K). Moreover, KDM5A-5B common targets predominantly co-occupied with active H3K4me3, and were linked to low-to-moderate levels of gene expression (Figures 3C, 3F, S3G). Collectively, these data signify that KDM5A and KDM5B have active and repressive gene regulatory functions respectively; however, when they act together, KDM5A suppresses the repressive gene regulatory functions of KDM5B in mESCs.

Overall, these findings imply that KDM4 and KDM5 members execute combinatorial functions for gene regulation.

HDM modules co-operate with key ESC regulatory networks

Epigenetic regulators function with ESC-TFs to establish ESC regulatory networks for the maintenance of the ESC state (Orkin and Hochedlinger, 2011; Young, 2011). To comprehend how HDMs collaborate with ESC-TFs and other epigenetic regulators for the gene regulation, we integrated the HDMome dataset along with binding datasets of ESC-TFs and other epigenetic regulators and mapped their binding sites at enhancers (mESC-specific 8,794 enhancers) and promoters (±2kb of TSS). The co-occupancy of multiple HDMs, ESC-TFs and other epigenetic regulators was defined as a module. We found three distinct modules at the enhancers (named eModule – I, II, III) and promoters (named pModule – I, II, III) in an unbiased manner (Figures 4A, 4B). Among these, "eModule – I" exhibited significant co-occupancy of KDM1A-KDM4B-KDM6A HDMs, ESC-TFs (OCT4, NANOG, SOX2), mediators (MED1,12), a cohesin complex (SMC1, NIPBL) and co-activator P300 at enhancers (Figures 4A, 4C). Mediators and the cohesin complex facilitate enhancer-promoter looping at the active gene loci to maintain the mESC state (Kagey et al., 2010). P300 is also involved in enhancer-promoter looping for ESC-specific genes expression in mESCs (Fang et al., 2014). Conversely, "pModule – II" displayed significant co-occupancy between JARID2-KDM4A-KDM4C-KDM5B HDMs, PRC1 (RING1B, CBX7) and PRC2 (EZH2, SUZ12, JARID2) components (Figures 4B, 4D). Collectively, these data hint at the possible mechanisms by which KDM1A-KDM4B-KDM6A and JARID2-KDM4A-KDM4C-KDM5B may control gene regulation.

We also explored the *trans*-regulatory factors (including TFs) that may be involved in KDM1A-KDM4B-KDM6A and JARID2-KDM4A-KDM4C-KDM5B mediated gene regulation. Motif analysis revealed enrichment of several ESC-specific TF motifs, including OCT4::SOX2, OCT4 (ESC master regulator), KLF4, KLF5 and ESSRB at the KDM1A-KDM4B-KDM6A co-occupied enhancers (Figure S4A). These observations are consistent with binding data, which demonstrated that certainly KDM1A-KDM4B-KDM6A co-occupied with the OCT4 and SOX2 TFs (Figure 4A); indicating a strong connection between KDM1A-KDM4B-KDM6A and ESC-TFs for gene activation in mESCs. Moreover, we identified a list of enriched motifs recognised by the ZBTB33, ZBTB7A, HINFP, TFAP2A and TFAP2C TFs at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters (Figure S4B). These TFs are linked to transcriptional repression of the developmental genes (Liu et al., 2013; Mitra et al., 2007; Pastor et al., 2018; Rothstein and Simoes-Costa, 2020) and perhaps act directly or indirectly through this set of HDMs and PRC complexes for gene repression in mESCs.

Furthermore, we examined whether HDMs interact with existing ESC regulatory networks (Das et al., 2014; Kim et al., 2010; Seruggia et al., 2019) to control the gene expression. We assembled the HDMome, ESC regulatory networks and gene expression (undifferentiated v. differentiated mESCs) datasets and found that KDM1A, KDM6A and KDM4B were mainly connected to the CORE (active) network at enhancers (Figure 4E). The majority of the KDM1A, KDM6A, KDM4B bound CORE network genes were highly expressed in the undifferentiated state, but were down-regulated upon differentiation (Figures 4E). In contrast, JARID2, KDM4A, KDM4C and KDM5B were predominantly connected to the PRC (repressive) network at the promoters (Figures 4F). The PRC target genes bound by these HDMs were mostly

repressed in the undifferentiated state but were up-regulated or de-repressed upon differentiation (Figures 4F). Nonetheless, we observed binding of multiple individual HDMs to the different ESC regulatory networks at the enhancers and promoters (Figures 4E, 4F), which indicate that individual HDMs may also act through distinct gene regulatory networks in mESCs.

Taken together, these data suggest that different combinations of HDMs, along with ESC-TFs and other epigenetic regulators, create distinct modules that co-operate with different key ESC regulatory networks for gene regulation.

A specific HDM set containing multiple HDMs physically interacts with the critical components of a specific ESC regulatory network

Based on the strong connection between KDM1A-KDM4B-KDM6A and the CORE regulatory network (Figures 4A, 4E), we anticipate that ESC-TFs (from the CORE network) may physically interact and recruit KDM1A, KDM4B and KDM6A HDMs at their targets for gene activation. To test this hypothesis, we examined physical interactions between the key components of the CORE network and KDM1A, KDM4B and KDM6A. Co-immunoprecipitation (co-IP) of KDM1A and KDM6A revealed an association with OCT4, SOX2, KLF4 and KLF5 ESC-specific TFs (Figure 4G). The physical interactions of KDM4B with OCT4 and SOX2 of the CORE network were observed in our previous study (Das et al., 2014). Moreover, KDM1A and KDM6A were also present with OCT4, SOX2 and KLF4 in the same fractions, indicating that they potentially form a complex (Figure S4C). In addition, we observed interactions among the KDM1A, KDM4B and KDM6A HDMs to different degrees (Figures 4G).

Our data also demonstrated a strong correlation between JARID2-KDM4A-KDM4C-KDM5B and the PRC regulatory network (Figures 4B, 4F), which hints at

possible physical association between this set of HDMs and PRC complexes. As JARID2 is a component of the PRC2 complex (Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009), co-IP of JARID2 exhibited physical interactions with EZH2 (a catalytic component of PRC2) and SUZ12 of the PRC2 complex, as expected (Figure 4H). JARID2 also physically interacted with the components of the PRC1 complex, such as RING1B (a catalytic constituent of PRC1 complex) (Figure 4H). Moreover, physical associations were observed between JARID2 and KDM4A, KDM5B HDMs to different extents (Figure 4H); the interaction between PRC2 and KDM4C was demonstrated previously (Das et al., 2014).

These data reveal physical interactions among the HDMs in a specific HDM set, as well as interactions between HDMs (from an HDM set) and the key components of a specific network. Hence, these findings support functional relationships between KDM1A-KDM4B-KDM6A and the CORE network, and between JARID2-KDM4A-KDM4C-KDM5B and the PRC regulatory network in mESCs.

KDM1A and KDM6A combinatorially modulate P300/H3K27ac, H3K4me1, H3K4me2 deposition and OCT4 recruitment at the enhancers for target gene expression

Although KDM1A-KDM4B-KDM6A co-occupies at enhancers, the co-occupancy is predominant between KDM1A and KDM6A (Figures 1A, S1B). In addition, compared to KDM4B targets, the KDM1A and KDM6A targets largely overlap with the CORE network (Figure 4E). Based on these criteria, KDM1A and KDM6A were the focus of further mechanistic studies. KDM1A was identified as a histone demethylase for H3K4me1 and H3K4me2 (Pedersen and Helin, 2010; Whyte et al., 2012), while KDM6A demethylates H3K27me2/me3 (Pedersen and Helin, 2010; Welstead et al.,

2012). Since KDM6A functions related to MLL4 and P300 at enhancers in mESCs (Wang et al., 2017), we included these two factors in the analyses. MLL4 is a histone methyltransferase for H3K4me1 (Cao et al., 2017) and P300 acts as a histone acetyltransferase (HAT) for H3K27ac (Fang et al., 2014). To examine how these players potentially co-operate, we analyzed their co-occupancy and found significant co-occupancy of KDM1A, KDM6A, MLL4 and P300 with H3K4me1, H3K4me2 and H3K27ac enhancer marks (Figure S5A). Similarly, both KDM1A and KDM6A overlapped with the OCT4, NANOG and SOX2 ESC-TFs (Figure S5A), as previously mentioned in this study (Figures 4A, 4C). Nevertheless, H3K27me3 (a known substrate of KDM6A demethylase) did not overlap with KDM6A binding sites (Figure S5A), which hints at a non-catalytic role of KDM6A/UTX in mESCs (Wang et al., 2017).

To investigate the potential mechanisms of KDM1A and KDM6A in gene expression, we created individual *knockout (KO)* mESC lines of *Kdm1a* and *Kdm6a* (Figures S5B, S5C; Table S3). ChIP-seq data of H3K4me1, H3K4me2, H3K27ac and H3K27me3 histone marks were generated from wild-type, *Kdm1a KO* and *Kdm6a KO* mESCs (Figures 5B, 5C, S5E-S5F, S5H-S5I, S5M-S5N). To capture these histone marks changes at the KDM1A and KDM6A binding sites within the enhancers, we catalogued high-confidence KDM1A and KDM6A binding sites based on overlapping binding/co-binding sites of – KDM1A (antibody-based ChIP-seq) and KDM1A-FB (Bio-ChIP-seq), as well as between KDM6A and KDM6A-FB within the enhancers. A total 5,651 KDM1A and KDM1A-FB co-occupied regions; 3,321 KDM6A and KDM6A-FB co-occupied regions was obtained (Figure S5D; Table S4) within the mESC-specific “all” enhancers (8,794) (“all” enhancers were listed based on (Whyte et al., 2013)). We also extracted KDM1A and KDM1A-FB co-occupied regions (2,021), KDM6A and KDM6A-FB co-occupied regions (833) within the “active” enhancers (4,346) (Figure

S5G; Table S4). The “active” enhancers contained H3K27ac and H3K4me1 enhancer marks, but were devoid of H3K4me3 promoter mark and ± 2kb of TSS as promoter regions (i.e. without ± 2kb TSS, NO H3K4me3, +H3K27ac, +H3K4me1) (Figure S5G; Table S4). Hence, these “active” enhancers (4,346) represent as a sub-set of “all” enhancers (8,794). The analyses revealed significant elevations of H3K27ac and H3K4me2, marginal increases of H3K4me1 and no changes in H3K27me3 at the KDM1A and KDM1A-FB, KDM6A and KDM6A-FB co-occupied regions within the “all” and “active” enhancers, in their respective *KOs* compared to wild-type (Figures S5E–S5F, S5H–S5I). Furthermore, to understand the “combined” functions of KDM1A and KDM6A at the enhancers, we also acquired high-confidence “KDM1A-KDM6A co-occupied regions” within the “all” and “active” enhancers. This yielded 3,147 and 747 KDM1A-KDM6A co-occupied regions within the “all” and “active” enhancers respectively (Figures 5A, S5L; Table S4), which displayed marked increases of H3K27ac and H3K4me2, as well as minor increases of H3K4me1 in both *Kdm1a* and *Kdm6a KOs* compared to wild-type (Figures 5B-5C, 5E-5F, S5M-S5N). However, the H3K27me3 mark remained unchanged at KDM1A/KDM6A occupied enhancers in *Kdm1a and Kdm6a KOs* (Figures 5B-5C, 5E-5F, S5E-S5F, S5H-S5I, S5M-S5N). Moreover, the occupancy of P300 (that deposits H3K27ac) and OCT4 (a key ESC-TF, and a central player in the CORE network) were significantly increased at KDM1A-KDM6A co-occupied enhancers in *Kdm1a* and *Kdm6a KOs* (Figures 5D-5F, S5O). These comprehensive analyses indicate that KDM1A/KDM6A modulates H3K27ac through P300, while only KDM1A modulates H3K4me1 and H3K4me2 (as KDM1A is a demethylase for H3K4me1 and H3K4me2) at their targeted enhancers.

Next, we examined the gene expression changes of KDM1A and KDM6A target genes in the *Kdm1a KO* and *Kdm6a KO,* respectively. Both KDM1A and KDM6A target

genes were significantly upregulated in their respective *KOs* compared to wild-type (Figures 5G-5H, S5J-S5K). The upregulation of these target genes was positively correlated with elevated levels of H3K27ac, H3K4me1 and H3K4me2 in the *Kdm1a* and *Kdm6a KOs* (Figures 5B-5C, 5G-5H, S5E-S5F, S5H-S5K, S5M-S5N), which suggests that KDM1A/KDM6A represses enhancer functions to some degree for gene regulation in mESCs. This agrees with the previous finding, which showed that KDM1A/LSD1 is partially involved in the repression of active genes through enhancer binding in mESCs (Cao et al., 2018).

These data imply that KDM1A and KDM6A combinatorially modulates P300/H3K27ac, H3K4me1 and H3K4me2 deposition that recruits OCT4 at the KDM1A/KDM6A occupied enhancers for target gene expression; further, KDM6A shows “non-catalytic” activity for H3K27me3 in mESCs.

JARID2, KDM4A, KDM4C and KDM5B co-operatively control H2AK119ub1 (of PRC1) and bivalent marks (of PRC2) at promoters for target gene repression

We have described a strong correlation between JARID2-KDM4A-KDM4C-KDM5B HDMs and the PRC network (Figure 4). As JARID2-KDM4A-KDM4C-KDM5B HDMs co-occupy with PRC1/H2AK119ub1 and PRC2/bivalent marks (Figure 4), these HDMs co-occupied targets serve as polycomb targets. To interrogate the potential mechanisms of this set of HDMs through PRC-mediated gene repression, we used *knockout (KO)* or knockdown (KD) HDM mESC lines (Figures S6A-S6C) and assessed changes in their related histone marks at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters.

In *Jarid2 knockout (KO)* mESCs, significant reduction of H2AK119ub1 of PRC1 and alteration of bivalent marks (reduction of H3K27me3 and elevation of H3K4me3) of PRC2 were observed at the JARID2-KDM4A-KDM4C-KDM5B co-occupied

promoters compared to wild-type (Figures 6A-6B, S6D). This observation is consistent with the known function of JARID2 (of the PRC2 complex) that recognises H2AK119ub1 of PRC1, and mediates H3K27me3 repressive activity of PRC2. Hence, JARID2 connects the PRC1 and PRC2 complexes for gene repression (Cooper et al., 2016a; Kalb et al., 2014).

Similarly, knockdown (KD) of Kdm4a and Kdm4c (known H3K9me2/me3 and H3K36me2/me3 demethylases) mESCs exhibited altered bivalent marks (reduced level of H3K27me3 and increased level of H3K4me3) at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters (Figures 6C-6F, S6E-S6F). However, we observed a nominal increase of H3K9me3 only in the Kdm4c KD (Figures 6D, 6F, S6F).

Moreover, Kdm5b (a known H3K4me2/me3 demethylase) knockdown (KD) mESCs showed marked reductions of H3K4me2 and H3K4me3, and a minimal reduction of H3K27me3 at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters (Figure 6G-6H, S6G).

Furthermore, gene expression analysis revealed up-regulation/de-repression of the JARID2-KDM4A-KDM4C-KDM5B co-occupied target genes in the absence of *Jarid2, Kdm4a*, *Kdm4c* and *Kdm5b* (Figure 6I). Taken together, these data suggest that JARID2, KDM4A, KDM4C and KDM5B co-operatively regulate H2AK119ub1 (of PRC1) and bivalent marks (of PRC2) at promoters for repression of their target genes.

DISCUSSION

The HDMome atlas and its significance in ESCs

A handful of studies have revealed combinatorial functions of a few selected HDMs (Das et al., 2014; Pedersen et al., 2016). To gain a more comprehensive view of this class of epigenetic regulators and their combinational actions in gene regulation, we

have generated an innovative HDMome map of all 20 well-characterized HDMs based on their genome-wide binding (Figure 1). This uncovered specific combinations of multiple HDMs co-occupy at enhancers (ENs) and promoters in mESCs (Figure 1). Additionally, a specific set of multiple HDMs co-occupy with distinct ESC-TFs and epigenetic regulators to form modules, and the modular activity of each specific set of HDMs is linked to discrete active or repressive ESC regulatory networks (Figure 4).

Our data also demonstrate that individual HDM acts through different combinations with other HDMs to regulate gene expression. For example, KDM4A, KDM4C and KDM5B co-occupy at the promoters in two different combinations for either gene activation or repression (Figures 1, 2, 4). Likewise, KDM4B acts at both enhancers and promoters through different combinations to regulate gene activation (Figures 1, 2). However, KDM4B functions are predominant at the enhancers (Figures 1, 2, 3, 4) (Das et al., 2014).

Multiple HDMs within the same sub-class tend to demethylate the same histone mark(s); hence, they appear to have redundant or overlapping functions (Kooistra and Helin, 2012; Pedersen et al., 2016). However, this is not necessarily always correct. Our comprehensive analyses provide insights into how distinct combinations of KDM4 and KDM5 members are linked to different active and repressive histone marks, and the corresponding gene expression (Figure 3). Thus, not all HDMs of the same class act similarly, even though they share the same “catalytic Jumonji domain”. This may happen due to several factors: i) the presence of other protein domains in individual HDMs; ii) variations in 3D conformation of the catalytic domain; and iii) alterations of a histone mark that affect other histone marks on the same histone tail (Kooistra and Helin, 2012). Therefore, the HDMome atlas/map is highly relevant in defining the targets of individual HDMs, as well as the shared targets of multiple HDMs (from the

same sub-class and/or different sub-classes), along with ESC-TFs, other epigenetic regulators, and histone marks in the genome. Our study provides comprehensive and systematic analyses that delineates potential functions of multiple HDMs (beyond their sub-class criteria) in controlling the turnover of histone marks *in situ* for fine-tuning gene expression programs in mESCs.

The combinatorial functions of KDM1A and KDM6A HDMs with P300 and OCT4 establish a KDM1A/KDM6A-P300-OCT4 system that directs the CORE regulatory network for target gene expression

KDM1A/LSD1 is the first reported HDM, which mainly catalyzes the demethylation of H3K4me1 and H3K4me2 (Mosammaparast and Shi, 2010). In mESCs, KDM1A/LSD1 form the KDM1A-NuRD (HDAC1/2, MBD3, Mi-2ß) repressive complex; also, KDM1A co-occupies with components of the NuRD complex, as well as with ESC-TFs (OCT4, NANOG, SOX2), co-activators (mediators) and active P300 (HAT). This mainly occurs at enhancers of active ESC genes, but the H3K4me1, me2– demethylase activity of KDM1A is suppressed due to the presence of acetylated histones (H3K27ac)/HAT P300 that ultimately drives the active ESC enhancers functions (Lee et al., 2006; Whyte et al., 2012). During the differentiation of mESCs, the levels of OCT4, NANOG, P300 HAT/H3K27ac are decreased which allows activation of the H3K4me1, me2– demethylase activity of KDM1A (of the KDM1A-NuRD complex) for silencing the active ESC enhancers (Whyte et al., 2012). Nonetheless, this model has not been validated fully. Our findings agree with the “suppression” of H3K4me1, me2– demethylase activity of KDM1A/LSD1 at the active enhancers in mESCs (Figures 5, S5). Additionally, we show that KDM1A regulates H3K27ac through P300 at these enhancers. The regulation of H3K27ac is more complex, since its level controlled

through both KDM1A and KDM6A HDMs, as well as by P300 HAT (Figures 5, S5). Altogether, our data implying that maintaining the fine-balance between acetylated (H3K27ac) and methylated (H3K4me1, me2) histones is crucial to preserve active enhancers functions in the steady-state of mESCs.

KDM6A/UTX demethylates H3K27me2, me3 repressive marks and thereby antagonizes the repressive function of PRC2, mainly at promoters (Kooistra and Helin, 2012; Laugesen and Helin, 2014). However, our data show that KDM6A predominantly binds at enhancers, and it has "non-catalytic activity of H3K27me3" in mESCs (Figures 1, 2, 4, 5), as observed in a prior study (Wang et al., 2017). The same study exhibited reciprocal regulation between KDM6A-P300 mediated H3K27ac and KDM6A-MLL4 mediated H3K4me1 at enhancers, which established a KDM6A-P300-MLL4 network at active enhancers in mESCs (Wang et al., 2017). We found that KDM6A regulates H3K27ac via P300 (same as the KDM1A); nonetheless, KDM6A-targeted H3K4me1, me2 regulation is primarily controlled through KDM1A at enhancers (Figures 5, S5). Collectively, our data reveal that KDM1A/KDM6A-P300 mediate H3K27ac, and only KDM1A mediate H3K4me1, me2 at enhancers that mutually recruit OCT4 at the KDM1A/KDM6A occupied enhancers for target gene expression (Figures 5, S5). Hence, we propose a "KDM1A/KDM6A-P300-OCT4" system at the enhancers that ultimately directs the CORE regulatory network for target gene expression to maintain the mESC state (Figure 7). However, we cannot exclude that MLLs (mainly MLL1, 2 and 4) and KDM1A act together to balance H3K4me1, me2 marks at murine ESC enhancers through establishing a complex "KDM1A/KDM6A-MLL(s)-P300-OCT4" system for KDM1A/KDM6A target gene expression.

The co-operative actions of JARID2-KDM4A-KDM4C-KDM5B and PRCs facilitates the PRC regulatory network for target gene repression

The polycomb repressive complexes, PRC1 and PRC2, play a central role in gene repression and thereby regulate cellular identity, development, and diseases (Margueron and Reinberg, 2011). PRC1 complexes exist in multiple forms. The canonical PRC1 contains CBX proteins and catalytic RING1A/B that catalyses H2AK119ub1 (Gao et al., 2012). While PRC2 complexes are present in two forms: PRC2.1 and PRC2.2; both have three core subunits – EZH1/2, SUZ12, EED – and distinct accessory factors, such as PCL1-3, EPOP/PAL1 (for PRC2.1) and JARID2, AEBP2 (for PRC2.2) (Healy et al., 2019; Højfeldt et al., 2019). JARID2 interacts with H2AK119ub1 (of canonical PRC1) and facilitates JARID2-PRC2.2 mediated H3K27me3 deposition and repressive activity at the targeted genes (Cooper et al., 2016b; Fursova et al., 2019; Healy et al., 2019; Højfeldt et al., 2019). JARID2 is also an HDM, but lacks demethylase activity (Kooistra and Helin, 2012; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009).

This study demonstrates the co-occupancy of JARID2-KDM4A-KDM4C-KDM5B with components of PRC1/H2AK119ub1 and PRC2/bivalent marks, which link to the PRC gene regulatory network (Figures 2, 4). In addition, our mechanistic data provide evidence that JARID2 is critical for the modulation of H2AK119ub1 of PRC1, as well as for the bivalent marks (H3K27me3-H3K4me3) of PRC2 at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters (Figure 6). Recent studies support these conclusions and acknowledge the importance of JARID2 in the PRC1-H2AK119ub1-JARID2-PRC2.2 pathway, where JARID2 acts a missing link between the PRC1 and PRC2 complexes (Di Croce and Helin, 2013; Healy et al., 2019; Højfeldt et al., 2019). In addition, our work shows that both KDM4A and KDM4C mainly

regulate the bivalent marks of PRC2 at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters (Figure 6), suggesting a functional connection between KDM4A, KDM4C and PRCs. This finding is consistent with the known repressive function of KDM4C together with PRC2 in mESCs (Das et al., 2014). Furthermore, KDM5B exhibits regulation of H3K4me2, me3 at the bivalent promoters of JARID2-KDM4A-KDM4C-KDM5B co-occupied targets (Figure 6). This supports the known functions of KDM5B, which mainly controls H3K4me2, me3 marks at the bivalent promoters in mESCs (Xie et al., 2011). The prevalent regulation of H2AK119ub1 (of PRC1) and bivalent marks (of PRC2) by JARID2, KDM4A, KDM4C and KDM5B at their co-targets certainly implies that these HDMs co-operatively regulate their target genes through PRC-mediated gene repression (Figure 7). We assume that KDM4A, KDM4C, KDM5B assist JARID2, which is the central player of the "PRC1-H2AK119ub1-JARDI2-PRC2 pathway" for target gene repression. Nonetheless, further studies will be required to test this hypothesis.

Taken together, our data provide an HDMome atlas of mESCs, which reveals the combinatorial functions of HDMs to orchestrate gene expression programs for the maintenance of the mESC state.

ACKNOWLEDGMENTS

We thank Dr Wang and Prof Roeder for helpful discussions. We also thank Genewiz high-throughput sequencing facility. This work was supported by National Health and Medical Research Council (NHMRC) of Australia (APP1159461 and APP1182804) to P.P.D. L.P. is supported by a National Human Genome Research Institute (NHGRI) Career Development Award (R00HG008399), and Genomic Innovator Award (R35HG010717). S.H.O. is an Investigator of the Howard Hughes Medical Institute

(HHMI).

AUTHOR CONTRIBUTIONS

P.P.D conceptualised the study. Y.K., P.T., M.M., M.J.B., V.K.P., D.H., M.K., A.W., A.G., A.B.C and L.W. performed the experiments and analysed the data. Y.K., P.T., M.M., L.W., J.K., L.P., S.H.O. and P.P.D interpreted the data. Y.K., K.G., G.C.Y. and L.P. performed bioinformatics analyses. Y.K., P.T., L.P., S.H.O. and P.P.D. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

MAIN FIGURE TITLES AND LEGENDS

Figure 1. The HDMome signature reveals combinatorial co-occupancy of multiple HDMs

(A) Heat map representation of co-occupancy of HDMs at the mESC-specific enhancers (8,794). Plots are centred on the region midpoint ± 2kb. Relative ChIP-seq peak intensities are indicated. The cartoon displays co-occupancy of KDM1A-KDM4B-KDM6A at the enhancers.

(B) Heat map representation of co-occupancy of HDMs at the promoters. Plots are centred on the region midpoint ± 2kb. Relative ChIP-seq peak intensities are indicated. The cartoon shows the co-occupancy of JARID2-KDM4A-KDM4C-KDM5B (with JARID2) and KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2) at the promoters.

(C-E) Genome browser representation of ChIP-seq normalised reads for multiple HDMs at the *Pou5f1* (C), *HoxA* (D) and *Dpy 30* (E) gene loci. RNA-seq tracks demonstrate expression of these genes in wild-type mESCs.

Related to Figure S1.

Figure 2. Multiple HDMs combinatorially regulate gene expression

(A-C) Heat map representation of co-occupancy of multiple HDMs with histone marks. Plots are centred on the region midpoint ± 2kb. Relative ChIP-seq peak intensities are indicated.

(D) Gene expression changes of three different target genes sets of multiple HDMs in the differentiated state compared to the undifferentiated state of mESCs.

(E-F) Gene ontology (GO) term analyses (biological processes) of KDM1A-KDM4B-KDM6A and JARID2-KDM4A-KDM4C-KDM5B co-occupied target genes.

Related to Figure S2.

Figure 3. Combinatorial gene regulatory functions of KDM4 and KDM5 members

(A-B) Heat maps show the bindings of KDM4A, KDM4B and KDM4C along with histone marks at the ± 2kb regions around the "unique" and different combinations of "common" KDM4A, KDM4B and KDM4C binding sites within the enhancers (A) and promoters (B). Relative ChIP-seq peak intensities are indicated.

(C) A heat map displays the bindings of KDM5A, KDM5B and KDM5C along with histone marks at the ± 2kb regions around the "unique" and different combinations of "common" KDM5A, KDM5B and KDM5C binding sites within the promoters. Relative ChIP-seq peak intensities are indicated.

(D-E) Target gene expression changes of KDM4 members ("unique" and different combinations of "common" target genes of KDM4 members) in the undifferentiated state compared to the differentiated state of mESCs at the enhancers (D) and promoters (E).

(F) Target gene expression changes of KDM5 members ("unique" and different combinations of "common" target genes of KDM5 members) in the undifferentiated state compared to the differentiated state of mESCs at the promoters.

(G-K) Genomic tracks represent ChIP-seq normalised reads for KDM4 and KDM5 members, as well as for histone marks at the different gene loci. RNA-seq tracks demonstrate expression of these genes in wild-type mESCs.

Related to Figure S3.

Figure 4. HDM modules cooperate with key ESC regulatory networks for gene regulation

(A-B) Modules are presented based on the co-occupancy of multiple HDMs, ESC-TFs and epigenetic regulators at the enhancers (A) and promoters (B) in an unbiased manner. Each row represents one module; enhancer modules (eModule – I, II, III) (A), and promoter modules (pModule – I, II, III) (B) are shown. The relative binding intensities of each factor are indicated. A schematic diagram of "eModule – I" illustrates prominent co-occupancy of KDM1A-KDM4B-KDM6A HDMs along with ESC-TFs (OCT4, NANOG, SOX2), mediators (MED1,12), a cohesin complex (SMC1, NIPBL) and co-activator P300 at the enhancers (A); whereas a schematic diagram of "pModule – II" illustrates significant co-occupancy of JARID2-KDM4A-KDM4C-KDM5B HDMs, PRC1 (RING1B, CBX7) and PRC2 (EZH2, SUZ12, JARID2) components at the promoters (B).

(C) The genome browser displays significant co-occupancy of KDM1A, KDM4B, KDM6A, ESC-TFs (OCT4, NANOG, SOX2), mediators (MED1, 12), the cohesin complex (SMC1/3, NIPBL), P300 and enhancer histone marks (H3K4me1, me2 and H3K27ac) at the *Pou5f1/Oct4* gene locus. ATAC-seq track is shown; both ATAC-seq and H3K27ac tracks represent open chromatin regions. The RNA-seq track exhibits expression of the *Pou5f1/Oct4* in wild-type mESCs. Highlighted regions are enhancers.

(D) The genome browser presents significant co-occupancy of JARID2, KDM4A, KDM4C, KDM5B, PRC1 (CBX7, RING1B, RYBP) and PRC2 (EZH2, SUZ12, JARID2) components, as well as H2AK119ub1 (of PRC1) and H3K27me3–H3K4me2, me3 bivalent (of PRC2) histone marks at the *HoxA* gene cluster. The RNA-seq track shows repression of *HoxA* cluster genes in wild-type mESCs.

(E-F) HDMome–ESC regulatory networks at the enhancers (E) and promoters (F). The inner ring contains all 20 HDMs; while the outer ring represents the CORE (active),

MYC (active) and PRC (repressive) networks. The lines extending from these HDMs to the networks represent the binding sites of HDMs to the networks. Ticks along the exterior of the outer ring shows differentially expressed HDM bound network genes in the differentiated state compared to undifferentiated state of mESCs. Magenta tick marks indicate increased expression of HDM bound network genes upon differentiation, while green tick marks indicate decreased expression of HDM bound network genes upon differentiation.

(G) Co-immunoprecipitation using streptavidin beads against biotinylated KDM1A-FLAG-Biotin(FB) and KDM6A-FB tagged proteins from the nuclear extract of mESCs, followed by Western blot analysis to detect their interacting partners.

(H) Co-immunoprecipitation using streptavidin beads against biotinylated JARID2-FB tagged protein from the nuclear extract of mESCs, followed by Western blot analysis to detect its interacting partners.

Related to Figure S4.

Figure 5. KDM1A and KDM6A combinatorially modulate P300/H3K27ac, H3K4me1, H3K4me2 deposition and OCT4 recruitment at the enhancers for target gene expression

(A) Venn diagram represents high-confidence KDM1A-KDM6A co-occupied sites (3,147) at “all” the mESC-specific enhancers (ENs) (8,794). KDM1A and KDM1A-FB overlapping sites (5,651); KDM6A and KDM6A-FB overlapping sites (3,321) at “all” the enhancers were used to determine the KDM1A-KDM6A co-occupied sites (see Figure S5D).

(B-C) Heat maps exhibit differential binding of H3K7ac, H3K27me3, H3K4me1 and H3K4me2 at the KDM1A-KDM6A co-occupied enhancers (3,147) in the *Kdm1a* (B)

and *Kdm6a* (C) *KOs* compared to wild-type. Two individual *Kdm6a KOs* were used to assess the data reproducibility.

(D) Heat maps display differential binding of P300 and OCT4 at the KDM1A-KDM6A co-occupied enhancers (3,147) in the *Kdm1a* and *Kdm6a KOs* compared to wild-type.

(E-F) Genomic tracks demonstrate the co-occupancy of KDM1A and KDM6A. Additionally, they show differential binding of histone marks (H3K7ac, H3K4me2, H3K4me1 and H3K27me3), P300 and OCT4 at the KDM1A-KDM6A co-occupied enhancers of the *Nanog* and *Prdm14* gene loci in the *Kdm1a* and *Kdm6a KOs* compared to wild-type. The highlighted regions represent KDM1A-KDM6A co-occupied enhancers.

(G-H) Gene expression changes of KDM1A-KDM6A co-occupied target genes (3,147 and 747) in the *Kdm1a* (G) and *Kdm6a* (H) *KOs* compared to wild-type.

Related to Figure S5.

Figure 6. JARID2, KDM4A, KDM4C and KDM5B co-operatively control H2AK119ub1 (of PRC1) and bivalent marks (of PRC2) at promoters for target gene repression

(A) Heat map represents differential binding of H2AK119ub1 and H3K27me3-H3K4me3 (bivalent marks) at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters in the *Jarid2 KO* compared to wild-type.

(B) Genomic tracks demonstrate the co-occupancy of JARID2-KDM4A-KDM4C-KDM5B with PRC1 and PRC2 components; also, they show differential binding of histone marks (H2AK119ub1, H3K27me3 and H3K4me3) at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters of the *Hoxa1-13* gene cluster in the *Jarid2 KO* compared to wild-type.

(C-D) Heat maps exhibit differential binding of H3K4me3, H3K27me3, H3K9me3 and H3K36me3 at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters in the Kdm4a (C) and Kdm4c (D) KDs compared to the anti-GFP KD (control).

(E-F) Genomic tracks display differential binding of H3K4me3, H3K27me3, H3K9me3 and H3K36me3 at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters of *Cbx2, 8* gene loci in the Kdm4a (E) and Kdm4c (F) KDs compared to the anti-GFP KD (control). The highlighted regions indicate differential binding of histone marks (KD v. control) at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters.

(G) Heat map represents differential binding of H3K4me2, H3K4me3 and H3K27me3 at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters in the Kdm5b KD compared to wild-type.

(H) Genomic tracks represent differential binding of H3K4me2, H3K4me3 and H3K27me3 at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters of *Cdkn2a, 2b* gene loci in the Kdm5b KD compared to the anti-GFP KD (control). Highlighted regions show differential binding of histone marks (KD v. control) at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters.

(I) Gene expression changes of JARID2-KDM4A-KDM4C-KDM5B co-occupied target genes in the *Jarid2 KO*, Kdm4a KD, Kdm4c KD and Kdm5b KD compared to wild-type. Related to Figure S6.

Figure 7. The model represents the combinatorial gene regulatory functions of HDMs in mESCs

The proposed model demonstrates that multiple HDMs function together in different combinations at either enhancers or promoters for gene regulation. KDM1A and KDM6A combinatorially modulate P300/H3K27ac, H3K4me1, H3K4me2 deposition

and recruit OCT4 at the KDM1A/KDM6A occupied enhancers to activate the CORE regulatory network for target genes expression. Conversely, JARID2, KDM4A, KDM4C and KDM5B co-operatively regulate H2AK119ub1 (of PRC1) and bivalent (H3K27me3-H3K4me3/me2) marks (of PRC2) at the promoters that facilitates the PRC regulatory network for target genes repression. Hence, different combinations of HDMs regulate distinct ESC gene regulatory networks for the maintenance of the mESC state.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact:

Partha Das (partha.das@monash.edu)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse embryonic stem cells (mESCs)

Mouse ESCs (mESCs) were cultured in mouse ESC media that contains DMEM (Dulbecco's modified Eagle's medium) (Thermo Fisher Scientific) supplemented with 15% fetal calf serum (FCS) (Thermo Fisher Scientific), 0.1mM b-mercaptoethanol (Sigma-Aldrich), 2mM L-glutamine (Thermo Fisher Scientific), 0.1mM nonessential amino acid (Thermo Fisher Scientific), 1% of nucleoside mix (Merck Millipore), 1000 U/ml recombinant leukemia inhibitory factor (LIF/ESGRO) (Merck Millipore), and 50U/ml Penicillin/Streptomycin (Thermo Fisher Scientific). mESCs were cultured at 37°C, 5% CO2.

Human embryonic kidney cells (HEK293T cells)

HEK293T cells were cultured with DMEM supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific) and 2% penicillin-streptomycin (Thermo Fisher Scientific). These cells were cultured at 37°C, 5% CO2. HEK293T cells were used only for lentiviral production.

METHOD DETAILS

ChIP-seq

Chromatin Immunoprecipitation (ChIP) was performed as described previously (Das et al., 2014; Kim et al., 2009). For bioChIP reactions, streptavidin beads (Dynabeads MyOne Streptavidin T1- Thermo Fisher Scientific) were used for the precipitation of chromatin, and 2% SDS was applied for the first washing step. All other steps were the same as conventional ChIP protocol. BirA expressing J1 mESCs were used as a control.

Conventional ChIP reactions were performed as described previously (Das et al., 2014; 2015). Input genomic DNA was used for the reference sample. Briefly, cells were trypsinized from 15cm dishes, washed twice with 1XPBS, and cross-linked with 37% formaldehyde solution (Calbiochem) to a final concentration of 1% for 8 min at room temperature with gentle shaking. The reaction was quenched by adding 2.5M glycine to a final concentration of 0.125M. Cells were washed twice with 1XPBS, and the cell pellet was resuspended in SDS-ChIP buffer (20 mM Tris-HCl pH 8.0, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100 and protease inhibitor), and chromatin was sonicated to around 200-500 bp. Sonicated chromatin was incubated with 5~10µg of antibody overnight at 4^0C. After overnight incubation, protein A/G Dynabeads magnetic beads (Thermo Fisher Scientific), were added to the ChIP reactions and incubated for 2-3 hours at 4^0C to immunoprecipitate chromatin. Subsequently, beads were washed twice with 1 ml of low salt wash buffer (50mM HEPES pH 7.5, 150mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate), once with 1ml of high salt wash buffer (50mM HEPES pH 7.5, 500mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate), once with 1ml of LiCl wash buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 0.5% sodium deoxycholate, 0.5% NP-40, 250 mM LiCl), and twice

with 1ml of TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, pH 8.0). The chromatin was eluted and reverse-crosslinked simultaneously in 300µl of SDS elution buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.0) at 65^0C overnight. The next day, an equal volume of TE was added (300µl). ChIP DNA was treated with 1µl of RNaseA (10mg/ml) for 1hr, and with 3µl of proteinase K (20mg/ml) for 3hrs at 37ºC, and purified using phenol-chloroform extraction, followed by QIAquick PCR purification spin columns (Qiagen). Finally, ChIP-DNA was eluted from the column with 40µl of water. For several factors, we used multiple ChIPs. In the end, all eluted ChIP-DNA samples were pooled and precipitated to enrich the ChIP-DNA material to make the libraries for high-throughput sequencing. Input ChIP samples were reserved before adding the antibodies, and these input samples were processed from reverse cross-link step until the end, same as other ChIP samples.

2-10 ng of purified ChIP DNA was used to prepare sequencing libraries, using NEBNext ChIP-seq Library Prep Master Mix Set for Illumina (NEB) and NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) according to the manufacturer's instructions. All libraries were checked through a Bio-analyser for quality control purpose. ChIP-sequencing (50bp SE reads or 150bp PE reads) was performed using Hiseq-2500 or Hiseq-4000 (Illumina).

Generation of Flag-Biotin (FB) tagged mESC lines

Open reading frames (ORFs) of HDM genes of interest were synthesized (IDT), and cloned into BamHI-digested pEF1a-Flagbio(FB)-puromycin vector, using Gibson Assembly Master Mix (NEB). Positive FB constructs were analysed by Sanger sequencing. 10µg of each of the HDM-FB construct was electroporated into $5x10^6$ J1 wild-type mESCs, which constitutively express BirA ligase (neomycin). The

electroporated cells were plated on a 15cm dish, with mESC media. After ~24hrs, media was replaced with fresh mESC media containing 1µg/ml of puromycin (Sigma-Aldrich) and 1µg/ml of neomycin (Sigma-Aldrich), and cells were selected for 4-5days. Individual mESC colonies were picked, expanded, and tested by western blot using either streptavidin-HRP antibody (GE healthcare) (dilution 1:2000 in 5% BSA) or specific antibodies against HDM to detect the HDM-FB lines.

Co-Immunoprecipitation

For each immunoprecipitation (IP), HDM-FB mESCs were harvested from 15 cm dish and washed twice with ice-cold PBS. The cell pellet was allowed to swell in twice the volume of Hypotonic solution (10mM HEPES- pH 7.3, 1.5mM MgCl2, 10mM KCl, 1mM DTT, 1mM PMSF and protease inhibitors), and passed through $26^{1/2}$-gauge needle 5 times, followed by quick centrifuge at 14000 rpm for 20-30 sec. The cloudy supernatant cytoplasmic fraction was removed, and the cell pellet was resuspended in the same volume of High salt buffer (20mM HEPES- pH 7.3, 1.5mM MgCl2, 420mM KCl, 0.2mM EDTA, 30% Glycerol, 1mM DTT, 1mM PMSF and protease inhibitors) and rotated for 1-2 hour at cold room. Then, Neutralizing buffer “without salt” (20mM HEPES- pH 7.3, 0.2mM EDTA, 20% Glycerol, 1mM DTT, 1mM PMSF and protease inhibitors) was added to the nuclear extract (NE) to bring the salt concentration to around 150mM. NE was centrifuged at 14000 rpm for 20 min at 4ºC, and the supernatant was collected. The volume of supernatant was increased up to 1ml using IP buffer (combining High salt buffer and Neutralizing buffer to make final conc. of 150mM, with 1mM DTT, 1mM PMSF and protease inhibitors). Percentage of supernatant was collected as an “Input”, and rest of the supernatant was incubated with streptavidin Dynabeads (Thermo Fisher Scientific) for 2hrs at 4ºC. Subsequently,

immunoprecipitated protein-beads were washed 3 times with IP buffer, each for 5 minutes at 4ºC. IP-ed protein and its interacting partners were eluted from the beads in 2X XT buffer (Bio-Rad) by boiling for 10 minutes; resolved on a 4-12% gradient Bis-Tris gel (Bio-Rad) and analyzed by western blot using specific antibodies.

Western Blot

Protein extract was mixed with 2X XT buffer (Bio-Rad), boiled for 10 minutes, and resolved on a 4-12% gradient Bis-Tris gel (Bio-Rad). Proteins are on the gel transferred to the PVDF membrane, and specific antibodies were used to detect proteins of interest.

RNA-seq

Total RNA (DNA free) was isolated using RNeasy Mini Kit (Qiagen), and ribosomal RNAs were depleted using ribosomal RNA depletion kit (rRNA Depletion Kit, NEB). Ribosomal depleted RNA was used to make the RNAseq libraries using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB). All the libraries were checked through a Bio-analyzer for quality control purpose. Paired-End (PE) 150bp reads were generated using a HiSeq-4000 sequencer (Illumina). RNAseq were performed in triplicates.

Generation of mESC *KO* lines using CRISPR-Cas9

Paired sgRNAs were designed to delete coding exons (exons 1-2) of *Kdm1a* and *Kdm6a.* These created 17kb and 2kb genomic deletions of *Kdm1a* and *Kdm6a.* The sgRNAs were cloned into lentiguide-Puro (Addgene ID: 52963) plasmid, using Golden Gate Cloning approach as mentioned previously. Wild-type (J1) mESCs were

transduced with Cas9-blast virus (generated from pLentiCas9-Blast, Addgene ID: 52962) and selected with 10µg/ml blasticidin (InvivoGen) to generate the J1 mESC cell line with stable expression of Cas9 (mESC+Cas9). 50,000 of mESC+Cas9 cells were transfected with 500ng of each sgRNAs (5 and 3' sgRNAs) using Lipofectamine 2000 (Thermo Fisher Scientific), and cells were selected with 1µg/ml puromycin (Sigma-Aldrich) for 3-4 days. Next, puromycin-resistant cells were re-plated on a 15cm plate as single cells to grow individual clones. The individual clones were picked, expanded, and genotype PCR was performed to screen the homozygous/biallelic deletion or *knockout (KO)* mESC clones. Furthermore, Sanger sequencing and Western blot analysis were performed to validate *Kdm1a* and *Kdm6a KOs*.

QUANTIFICATION AND STATISTICAL ANALYSIS

ChIP-seq analysis

The 50bp Single-End (SE) or 150bp Paired-End (PE) reads were mapped to the mm9 mouse genome assembly using Bowtie 2 (v2.3.2) (Langmead and Salzberg, 2012). BigWig tracks were generated using the 'bamCoverage' program from the DeepTools software suite (v2.5.3) (Ramírez et al., 2016), with reads extended to 300bp and tracks normalized using the RPKM (Reads Per Kilobase per Million mapped reads) algorithm. Peak calling was done with the MACS2 program (v2.1.1.20160309) (Zhang et al., 2008). For most ChIP-seq datasets default MACS2 settings were used. Artefact peaks were filtered out based on the mouse ENCODE project (Mouse ENCODE Consortium et al., 2012); they were removed using "bedtools intersect" from the BEDTools software suite (v2.26.0) (Quinlan and Hall, 2010). Mapping of peaks to the nearest gene was done using "bedtools closest". Mouse ESC-specific enhancers taken from the literature (Whyte et al., 2013). Promoters defined as a +/-2kb of TSS.

Heat-map

ChIP-seq datasets were generated using the “computeMatrix” and “plotHeatmap” programs of the DeepTools software suite (v.2.5.3) (Ramírez et al., 2016) for ChIP-seq heatmaps and profile plots of histone marks and transcription factors binding. Regions from -2kb to +2kb around the centre of the peaks were used, split into 10 or 50bp bins, with each bin receiving the mean of the BigWig signal scores across it.

GO (Gene Ontology)

Gene Ontology (Ashburner et al., 2000) analysis was performed using the NIH DAVID website (v6.8) (Dennis et al., 2003); default settings and the whole genome were used as background. Genes were identified using their Ensembl Gene IDs for the DAVID analysis, and their associated gene names were obtained from the GENCODE vM1 mouse transcriptome annotation and added to the resulting GO term enrichment tables.

MOTIF analysis

Enriched motifs of transcription factors (TFs) were identified at the ChIP peaks target regions using the “Haystack” bioinformatics pipeline, as described previously (Pinello et al., 2018). Each enriched motif assigned with motif logos, p-values (calculated with the Fisher’s exact test) and q-values, the central enrichment score, the average profile in the target regions containing the motif, and the closest genes for each region.

Module analysis

The most significant combinations were detected performing principal component

analysis on the signal intensity of the different tracks and considering the loadings of the first 3 eigenvectors with the largest eigenvalues.

Network analysis

Binding sites of each HDMs were mapped to the closest regions of defining networks (PRC, CORE, and MYC), as well as to the closest genes. We drew two diagrams, one representing all the mappings between HDMs and network regions associated with promoters; another one representing all the mappings between HDMs and network regions associated with enhancers. Edges were coloured based on the designation of the mapped networks (PRC = red, CORE = blue, MYC = orange). Tick marks next to network regions indicate that differential expression of HDM bound network genes in the differentiated state compared to the undifferentiated mESCs ($p<0.05$ and absolute log2-fold-change>1).

RNA-seq analysis

Reads were mapped to the mm9 mouse genome assembly and the GENCODE version M1 mouse transcriptome (Mudge and Harrow, 2015) using the STAR RNA-seq read alignment program (v2.5.3a) (Dobin et al., 2013) with default parameters. Bigwig tracks were generated using STAR to generate RPM-normalized wiggle files. Subsequently, the wigToBigWig tool (v4) from the UCSC genome browser Kent tools (Kent et al., 2010) was used to convert the wiggle files to BigWig format. Gene expression quantification, both raw read counts and TPM (Transcripts Per Million) values were obtained using htseq-count, part of the HTSeq package. Differential gene expression analysis was done using the DESeq2 (v1.14.1) (Love et al., 2014) R statistical programming language (v3.3.2) software package from the Bioconductor

project (v3.4) (Huber et al., 2015) using the GENCODE version M1 mouse transcriptome. DESeq2 was also used to calculate the normalized counts and the FPKM (Fragments Per Kilobase per Million mapped reads) expression values. MA plots were created employing the ggplot2 R package (v2.2.1) that used log2 fold changes and mean normalized counts calculated by the DESeq2 R package.

Violin Plots/ differential expression

Log2 fold change for closest genes was plotted as violin plots using the **ggpbur R package (v0.4.0),** and the statistical significance of differential expression was calculated using the Wilcoxon rank-sum test with continuity correction, as implemented in the R statistical program (v3.3.2)

Statistical analysis

All statistical analyses for the RNA-seq and ChIP-seq data were performed using the R statistical program (v3.3.2).

DATA AND SOFTWARE AVAILABILITY

All the NGS data from this study have been deposited in GEO under the accession number GSE156107.

KEY RESOURCES (a separate document to be uploaded)

Antibodies

Anti-KDM1A: Abcam ab17721 (ChIP); Santa Cruz Biotechnology sc-271720 (WB)

Anti-KDM2A: Novus Biologicals NB100-74602

Anti-KDM2B: Merck #17-10264

Anti-KDM3A: Bethyl laboratories A301-538A

Anti-KDM3B: Merck #07-1535

Anti-KDM4A: Cell Signaling #3393

Anti-KDM4B: Bethyl laboratories A301-478A

Anti-KDM4C: Abcam ab85454 (WB); Novus Biologicals NBP1-49600 (WB)

Anti-KDM4D: Santa Cruz Biotechnology sc-393750

Anti-KDM5A: Abcam ab70892

Anti-KDM5B: Bethyl laboratories A301-813A

Anti-KDM5C: Bethyl laboratories A301-034A

Anti-KDM5D: Merck #ABE203

Anti-KDM6A: Bethyl laboratories A302-374A

Anti-UTY: Abcam ab91236

Anti-KDM6B: Abcam ab95113

Anti-PHF2: Cell Signaling #3497

Anti-PHF8: Bethyl laboratories A301-772A

Anti-JMJD6: Santa Cruz Biotechnology sc-28348

Anti-JARID2: Cell Signaling #13594; Novus Biologicals NB100-2214 – (ChIP, WB)

Anti-EZH2: Cell Signaling #5246

Anti-SUZ12: Active motif # 39357; Abcam ab12073

Anti-RING1B: Bethyl laboratories A302-869A

Anti-KLF4: Abcam ab106629

Anti-KLF5: Abcam ab137676

Anti-SOX2: Abcam ab59776

Anti-OCT4: Santa Cruz Biotechnology sc-5279 (ChIP); Abcam ab19857 (WB)

Anti-P300: Abcam ab54984 (ChIP); Thermo scientific #33-7600 (ChIP)

Anti-H3K4me1: Abcam ab8895

Anti-H3K4me2: Abcam ab32356

Anti-H3K4me3: Merck #07-473

Anti-H3K9me3: Abcam ab8898

Anti-H3K27me3: Abcam ab6002

Anti-H3K27ac: Abcam ab4729

Anti-H3K36me3: Abcam ab9050

Anti-H2AK119ub1: Cell Signaling #8240

Cell lines

J1 mESCs (wild-type), *Kdm1a KO, Kdm6a KO* mESCs.

Jarid2 KO mESCs (Das et al., 2015; Shen et al., 2009).

Kdm4a KD, Kdm5b KD mESCs; Kdm4c KD cells (Das et al., 2014).

BirA, JARID2-FB, KDM1A-FB, and KDM6A-FB mESC lines.

HEK293T cells.

Critical commercial assays

RNeasy Mini Kit (74106, Qiagen)

Ribosomal RNA depletion kit (rRNA Depletion Kit, E6310L, NEB)

NEBNext Ultra Directional RNA Library Prep Kit (E7420L, NEB)

NEBNext ChIP-seq Library Prep Master Mix Set for Illumina (E6240L, NEB)

NEBNext Ultra DNA Library Prep Kit for Illumina (E7370L, NEB)

Dynabeads MyOne Streptavidin T1 beads (65601, Thermo Fisher Scientific)

Deposited data

All the high-throughput sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO) under accession number of GSE…

ChIP-seq data and RNA-seq data were used from Das et al., 2014. GSE43231; Seruggia et al., 2019. GSE113335.

ChIP-seq data for modules and networks analyses were used from Das et al., 2014. GSE43231; Seruggia et al., 2019. GSE113335.

ChIP-seq and RNA-seq data of Kdm4c KD were used from Das et al., 2014. GSE43231.

ChIP-seq and RNA-seq data of Kdm5b KD were used from Kidder et al., 2014. GSE53093.

RNA-seq data of *Jarid2 KO* were used from Das et al., 2015. GSE58414.

RNA-seq data of undifferentiated mESCs (0hr) and differentiated state (96hr) were obtained from Canver et al., 2019. GSE140911.

Software and Algorithms

1) BEDTools (2.26.0): (Quinlan and Hall, 2010)

http://bedtools.readthedocs.io/en/latest/

2) Bowtie 2 (2.3.2): (Langmead and Salzberg, 2012)

http://bowtie-bio.sourceforge.net/bowtie2/index.shtml

3) MACS2 (2.1.1.20160309): (Zhang et al., 2008)

https://github.com/taoliu/MACS

4) DeepTools (2.5.3): (Ramírez et al., 2016)

https://deeptools.readthedocs.io/en/develop/

5) STAR (2.5.3a): (Dobin et al., 2013)

https://github.com/alexdobin/STAR

6) HTSeq package (0.12.4): (Anders et al., 2015)

https://github.com/htseq/htseq

7) wigToBigWig, Kent tools (v4): (Kent et al., 2010)

http://hgdownload.cse.ucsc.edu/admin/exe/

8) DESeq2 R package (1.14.1; Bioconductor 3.4): (Love et al., 2014)

https://bioconductor.org/packages/release/bioc/html/DESeq2.html

9) DiffBind R package (2.2.12; Bioconductor 3.4): (Ross-Innes et al., 2012)

https://bioconductor.org/packages/release/bioc/html/DiffBind.html

10) R statistical program (3.3.2):

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

URL https://www.R-project.org/.

11) ggplot2 R package (2.2.1):

ggplot2: Elegant Graphics for Data Analysis.

Springer-Verlag New York, 2009. ISBN: 978-0-387-98140-6

http://ggplot2.org

12) pheatmap R package (1.0.8):

pheatmap: Pretty Heatmaps. R package version 1.0.8.

https://CRAN.R-project.org/package=pheatmap

13) NIH DAVID website (6.8): (Dennis et al., 2003)

https://david.ncifcrf.gov/

14) ggpubr R package (0.4.0)

'ggplot2': Based Publication Ready Plots

https://rpkgs.datanovia.com/ggpubr/

15) Haystack: (Pinello et al., 2018)

https://github.com/lucapinello/Haystack

REFERENCES

Adamo, A., Sesé, B., Boue, S., Castaño, J., Paramonov, I., Barrero, M.J., and Izpisua Belmonte, J.C. (2011). LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. Nat. Cell Biol. *13*, 652–659.

Agger, K., Christensen, J., Cloos, P.A., and Helin, K. (2008). The emerging functions of histone demethylases. Current Opinion in Genetics & Development *18*, 159–168.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics *25*, 25–29.

Cao, K., Collings, C.K., Marshall, S.A., Morgan, M.A., Rendleman, E.J., Wang, L., Sze, C.C., Sun, T., Bartom, E.T., and Shilatifard, A. (2017). SET1A/COMPASS and shadow enhancers in the regulation of homeotic gene expression. Genes Dev. *31*, 787–801.

Cao, K., Collings, C.K., Morgan, M.A., Marshall, S.A., Rendleman, E.J., Ozark, P.A., Smith, E.R., and Shilatifard, A. (2018). An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. Sci Adv *4*, eaap8747.

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

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., et al. (2016a). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. Nat Commun *7*, 1–8.

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., et al. (2016b). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. Nat Commun *7*, 1–8.

Das, P.P., Hendrix, D.A., Apostolou, E., Buchner, A.H., Canver, M.C., Beyaz, S., Ljuboja, D., Kuintzle, R., Kim, W., Karnik, R., et al. (2015). PRC2 Is Required to Maintain Expression of the Maternal Gtl2-Rian-Mirg Locus by Preventing De Novo DNA Methylation in Mouse Embryonic Stem Cells. CellReports 1–57.

Das, P.P., Shao, Z., Beyaz, S., Apostolou, E., Pinello, L., De Los Angeles, A., O'Brien, K., Atsma, J.M., Fujiwara, Y., Nguyen, M., et al. (2014). Distinct and Combinatorial Functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in Mouse Embryonic Stem Cell Identity. Molecular Cell *53*, 32–48.

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. *4*, P3.

Di Croce, L., and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. Nature Publishing Group *20*, 1147–1155.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Elkon, R., and Agami, R. (2017). Characterization of noncoding regulatory DNA in the human genome. Nat. Biotechnol. *35*, 732–746.

Fang, F., Xu, Y., Chew, K.-K., Chen, X., Ng, H.-H., and Matsudaira, P. (2014). Coactivators p300 and CBP maintain the identity of mouse embryonic stem cells by mediating long-range chromatin structure. Stem Cells *32*, 1805–1816.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell *143*, 212–224.

Fursova, N.A., Blackledge, N.P., Nakayama, M., Ito, S., Koseki, Y., Farcas, A.M., King, H.W., Koseki, H., and Klose, R.J. (2019). Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. Molecular Cell 1–26.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. Molecular Cell *45*, 344–356.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat. Cell Biol. *15*, 373–384.

Healy, E., Mucha, M., Glancy, E., Fitzpatrick, D.J., Conway, E., Neikes, H.K., Monger, C., Van Mierlo, G., Baltissen, M.P., Koseki, Y., et al. (2019). PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. Molecular Cell 1–23.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nature Methods *12*, 115–121.

Højfeldt, J.W., Hedehus, L., Laugesen, A., Tatar, T., Wiehle, L., and Helin, K. (2019). Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. Molecular Cell 1–18.

Inoue, H., and Yamanaka, S. (2011). The Use of Induced Pluripotent Stem Cells in Drug Development. 1–7.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W.T.C., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3

methylation in Polycomb repression. Nat. Struct. Mol. Biol. *21*, 569–571.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207.

Kim, J., Cantor, A.B., Orkin, S.H., and Wang, J. (2009). Use of in vivo biotinylation to study protein–protein and protein–DNA interactions in mouse embryonic stem cells. Nat Protoc *4*, 506–517.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. Cell *132*, 1049–1061.

Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010). A Myc Network Accounts for Similarities between Embryonic Stem and Cancer Cell Transcription Programs. Cell *143*, 313–324.

Kooistra, S.M., and Helin, K. (2012). Molecular mechanisms and potential functions of histone demethylases. Nature Publishing Group *13*, 297–311.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*, 357–359.

Laugesen, A., and Helin, K. (2014). Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. Stem Cell *14*, 735–751.

Lee, M.G., Wynder, C., Bochar, D.A., Hakimi, M.-A., Cooch, N., and Shiekhattar, R. (2006). Functional Interplay between Histone Demethylase and Deacetylase Enzymes. Mol. Cell. Biol. *26*, 6395–6402.

Liang, G., He, J., and Zhang, Y. (2012). Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. Nat. Cell Biol. *14*, 457–466.

Liu, Y., Zhang, X., Blumenthal, R.M., and Cheng, X. (2013). A common mode of recognition for methylated CpG. Trends Biochem. Sci. *38*, 177–183.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Mitra, P., Xie, R., Harper, J.W., Stein, J.L., Stein, G.S., and van Wijnen, A.J. (2007). HiNF-P is a bifunctional regulator of cell cycle controlled histone H4 gene transcription. J. Cell. Biochem. *101*, 181–191.

Mosammaparast, N., and Shi, Y. (2010). Reversal of Histone Methylation: Biochemical and Molecular Mechanisms of Histone Demethylases. Annu. Rev. Biochem. *79*, 155–179.

Mouse ENCODE Consortium, Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R., et al. (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol. *13*, 418.

Mudge, J.M., and Harrow, J. (2015). Creating reference gene annotation for the mouse C57BL6/J genome assembly. Mamm. Genome *26*, 366–378.

Orkin, S.H., and Hochedlinger, K. (2011). Chromatin Connections to Pluripotency and Cellular Reprogramming. Cell *145*, 835–850.

Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306–310.

Pastor, W.A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T.J., Lukianchikov, A., Liu, X., Polo, J.M., Jacobsen, S.E., et al. (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. Nat. Cell Biol. *20*, 553–564.

Pedersen, M.T., and Helin, K. (2010). Histone demethylases in development and disease. Trends in Cell Biology *20*, 662–671.

Pedersen, M.T., Kooistra, S.M., Radzisheuskaya, A., Laugesen, A., Johansen, J.V., Hayward, D.G., Nilsson, J., Agger, K., and Helin, K. (2016). Continual removal of H3K9 promoter methylation by Jmjd2 demethylases is vital for ESC self-renewal and early development. The EMBO Journal *35*, 1550–1564.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell *139*, 1290–1302.

Pinello, L., Farouni, R., and Yuan, G.-C. (2018). Haystack: systematic analysis of the variation of epigenetic states and cell-type specific regulatory elements. Bioinformatics *34*, 1930–1933.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., et al. (2011). Combinatorial Patterning of Chromatin Regulators Uncovered by Genome-wide Location Analysis in Human Cells. Cell *147*, 1628–1639.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Robinton, D.A., and Daley, G.Q. (2012). The promise of induced pluripotent stem cells in research and therapy. Nature *481*, 295–305.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer.

Nature *481*, 389–393.

Rothstein, M., and Simoes-Costa, M. (2020). Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. Genome Res. *30*, 35–48.

Seruggia, D., Oti, M., Tripathi, P., Canver, M.C., LeBlanc, L., Di Giammartino, D.C., Bullen, M.J., Nefzger, C.M., Sun, Y.B.Y., Farouni, R., et al. (2019). TAF5L and TAF6L Maintain Self-Renewal of Embryonic Stem Cells via the MYC Regulatory Network. Molecular Cell 1–24.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.-C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303–1314.

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

Thomson, J.A. (1998). Embryonic Stem Cell Lines Derived from Human Blastocysts. Science *282*, 1145–1147.

Venters, B.J., Wachi, S., Mavrich, T.N., Andersen, B.E., Jena, P., Sinnamon, A.J., Jain, P., Rolleri, N.S., Jiang, C., Hemeryck-Walsh, C., et al. (2011). A Comprehensive Genomic Binding Map of Gene and Chromatin Regulatory Proteins in Saccharomyces. Molecular Cell *41*, 480–492.

Wang, S.-P., Tang, Z., Chen, C.-W., Shimada, M., Koche, R.P., Wang, L.-H., Nakadai, T., Chramiec, A., Krivtsov, A.V., Armstrong, S.A., et al. (2017). A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. Molecular Cell *67*, 308–321.e6.

Welstead, G.G., Creyghton, M.P., Bilodeau, S., Cheng, A.W., Markoulaki, S., Young, R.A., and Jaenisch, R. (2012). X-linked H3K27me3 demethylase Utx is required for embryonic development in a sex-specific manner. Proc. Natl. Acad. Sci. U.S.a. *109*, 13004–13009.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221–225.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

Xie, L., Pelz, C., Wang, W., Bashar, A., Varlamova, O., Shadle, S., and Impey, S. (2011). KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. The EMBO Journal *30*, 1473–1484.

Young, R.A. (2011). Control of the Embryonic Stem Cell State. Cell *144*, 940–954.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E.,

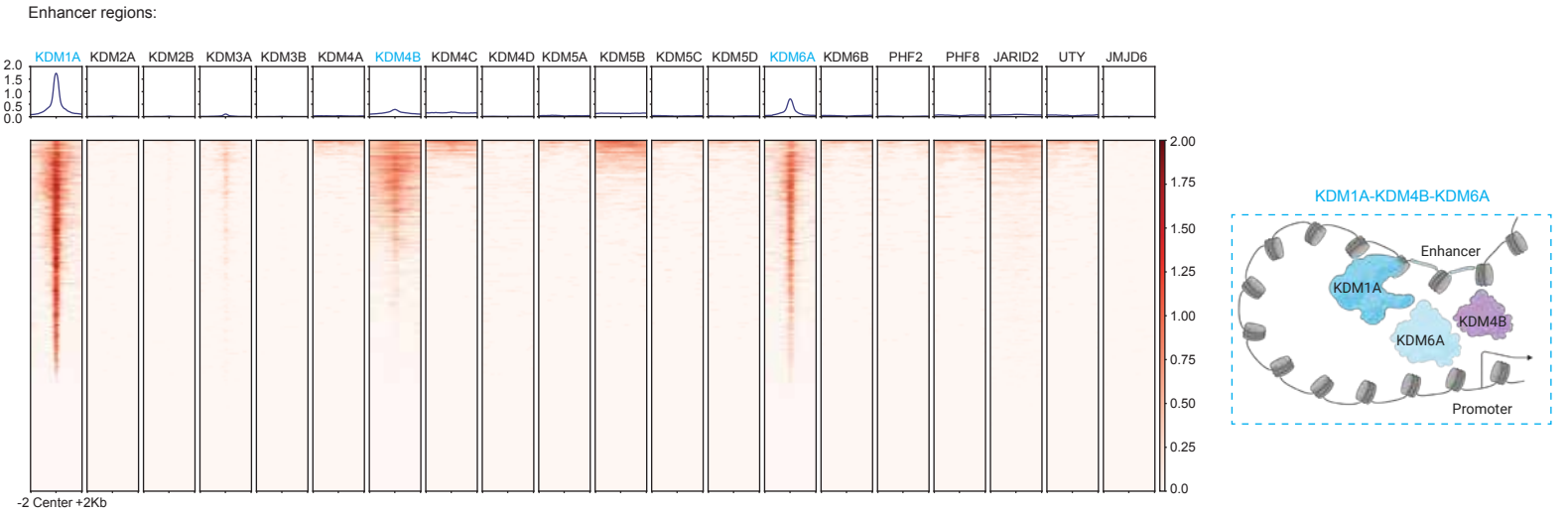
Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2010). Charting histone modifications and the functional organization of mammalian genomes. Nature Publishing Group *12*, 7–18.

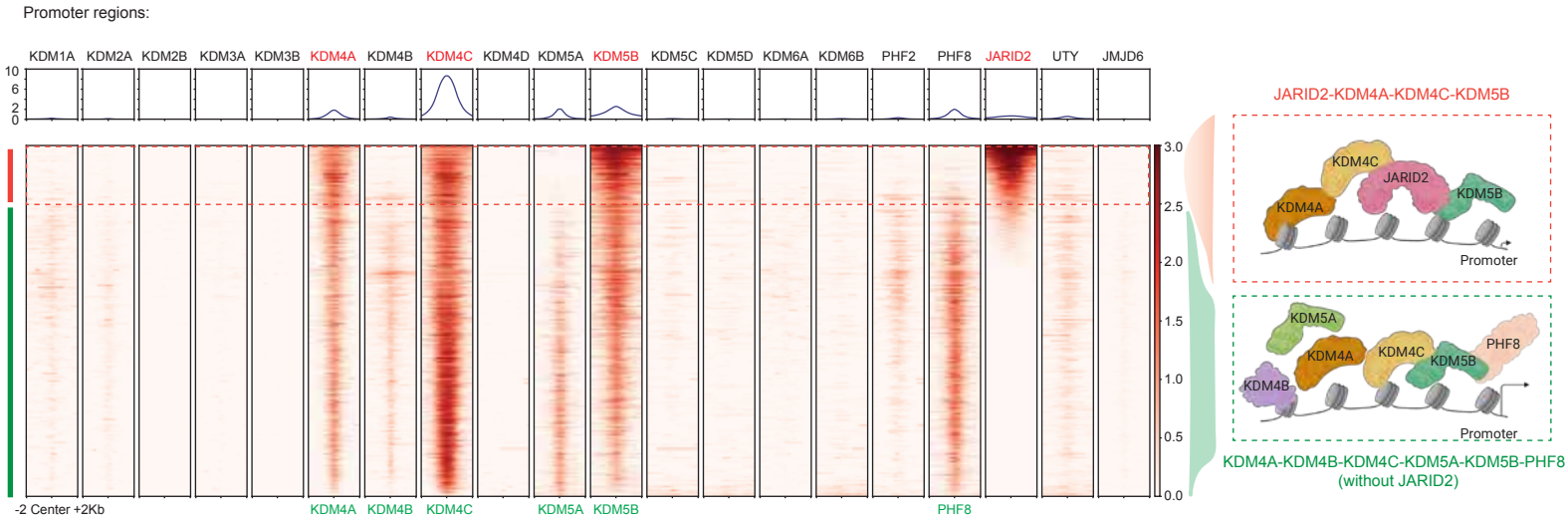
(2014). KDM5B focuses H3K4 methylation near promoters and enhancers during embryonic stem cell self-renewal and differentiation. 1–19.

Figure 1

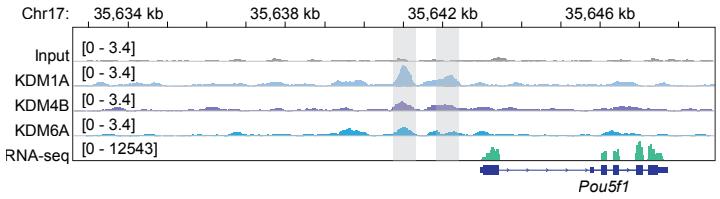
A



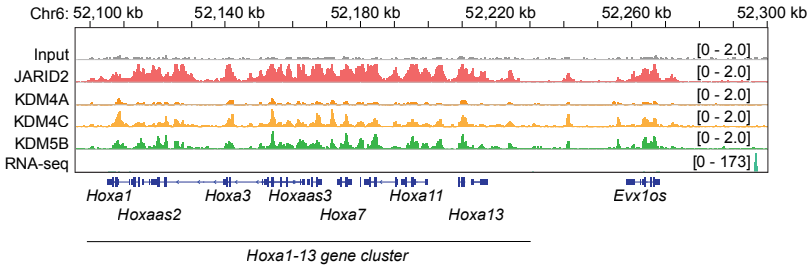
B



C



D



E

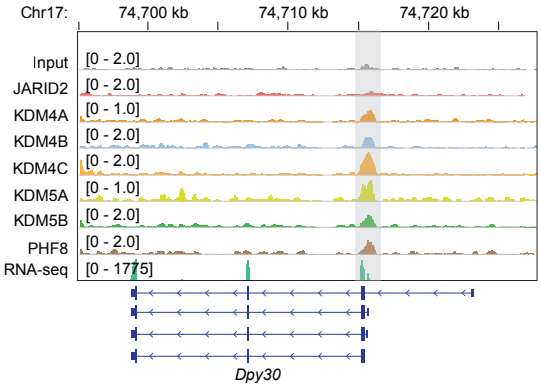


Figure 2

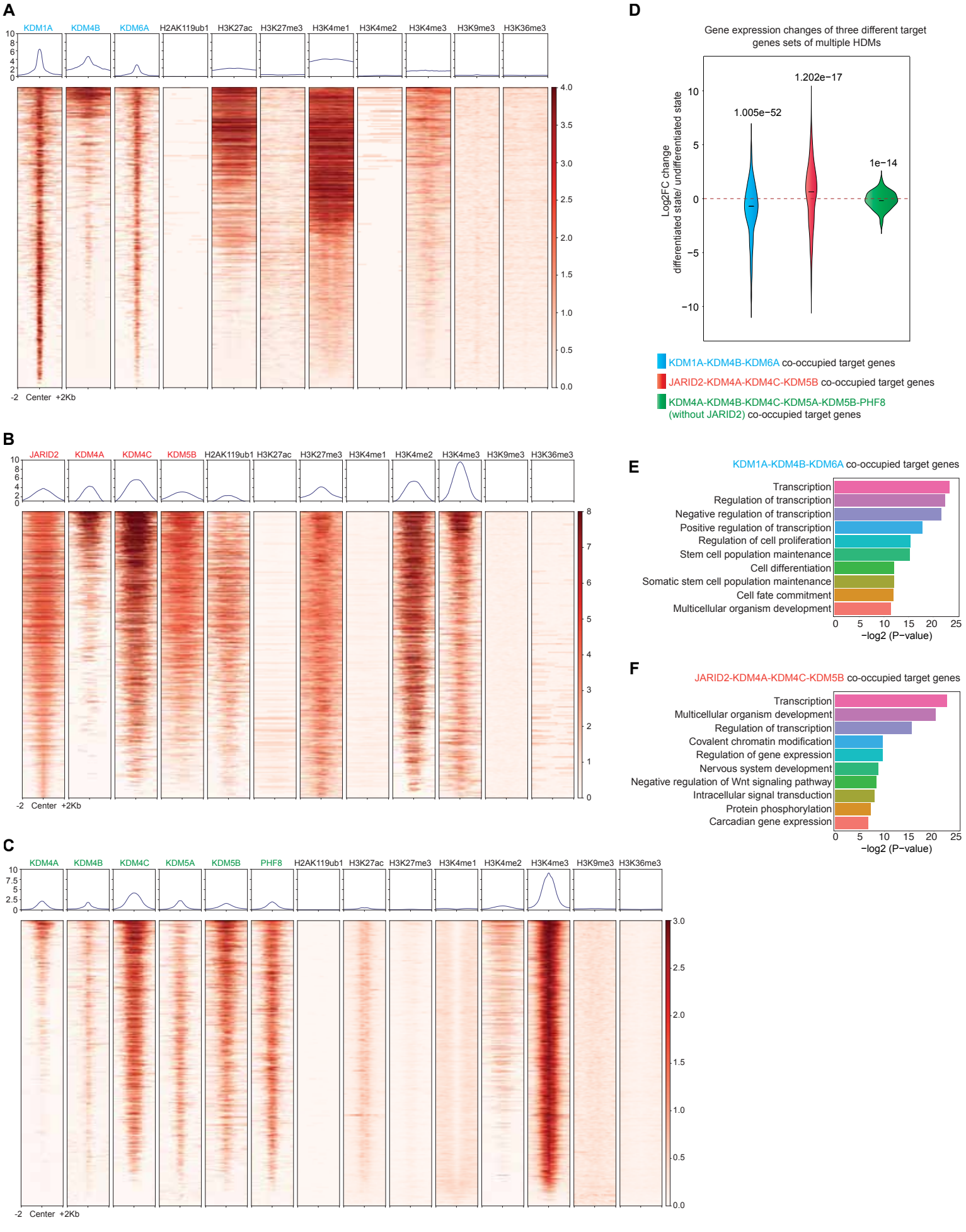

A
KDM1A
KDM4B
KDM6A
H2AK119ub1
H3K27ac
H3K27me3
H3K4me1
H3K4me2
H3K4me3
H3K9me3
H3K36me3
-2 Center +2Kb
B
JARID2
KDM4A
KDM4C
KDM5B
H2AK119ub1
H3K27ac
H3K27me3
H3K4me1
H3K4me2
H3K4me3
H3K9me3
H3K36me3
-2 Center +2Kb
C
KDM4A
KDM4B
KDM4C
KDM5A
KDM5B
PHF8
H2AK119ub1
H3K27ac
H3K27me3
H3K4me1
H3K4me2
H3K4me3
H3K9me3
H3K36me3
-2 Center +2Kb
D
Gene expression changes of three different target genes sets of multiple HDMs
Log2FC change differentiated state/ undifferentiated state
1.005e−52
1.202e−17
1e−14
KDM1A-KDM4B-KDM6A co-occupied target genes
JARID2-KDM4A-KDM4C-KDM5B co-occupied target genes
KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2) co-occupied target genes
E
KDM1A-KDM4B-KDM6A co-occupied target genes
Transcription
Regulation of transcription
Negative regulation of transcription
Positive regulation of transcription
Regulation of cell proliferation
Stem cell population maintenance
Cell differentiation
Somatic stem cell population maintenance
Cell fate commitment
Multicellular organism development
−log2 (P−value)
F
JARID2-KDM4A-KDM4C-KDM5B co-occupied target genes
Transcription
Multicellular organism development
Regulation of transcription
Covalent chromatin modification
Regulation of gene expression
Nervous system development
Negative regulation of Wnt signaling pathway
Intracellular signal transduction
Protein phosphorylation
Carcadian gene expression
−log2 (P−value)

Figure 3

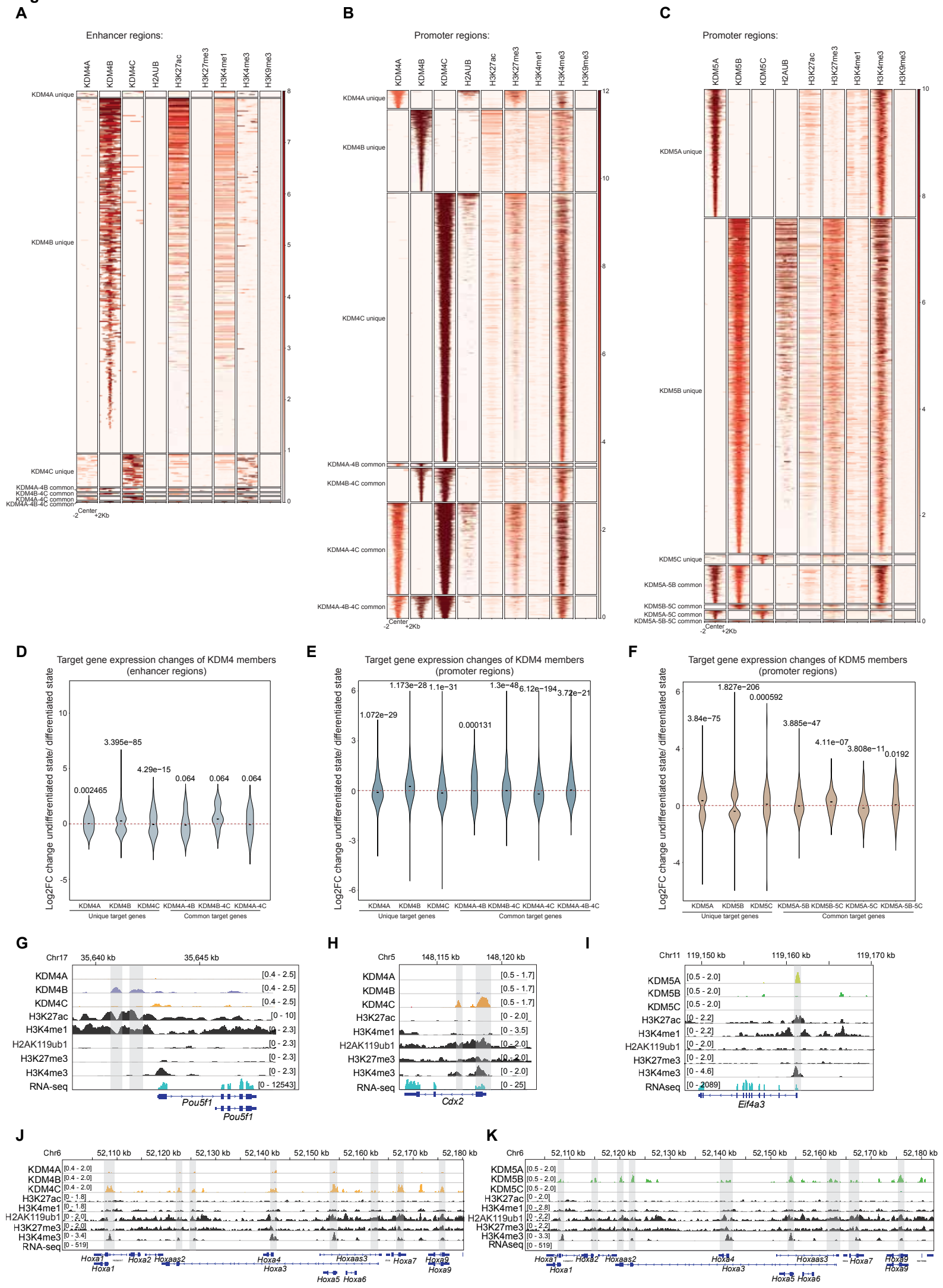

A
Enhancer regions:
KDM4A
KDM4B
KDM4C
H2AUB
H3K27ac
H3K27me3
H3K4me1
H3K4me3
H3K9me3
KDM4A unique
KDM4B unique
KDM4C unique
KDM4A-4B common
KDM4B-4C common
KDM4A-4C common
KDM4A-4B-4C common
Center
-2
+2Kb
B
Promoter regions:
KDM4A unique
KDM4B unique
KDM4C unique
KDM4A-4B common
KDM4B-4C common
KDM4A-4C common
KDM4A-4B-4C common
C
Promoter regions:
KDM5A
KDM5B
KDM5C
KDM5A unique
KDM5B unique
KDM5C unique
KDM5A-5B common
KDM5B-5C common
KDM5A-5C common
KDM5A-5B-5C common
D
Target gene expression changes of KDM4 members
(enhancer regions)
Log2FC change undifferentiated state/ differentiated state
0.002465
3.395e−85
4.29e−15
0.064
0.064
0.064
KDM4A
KDM4B
KDM4C
KDM4A-4B
KDM4B-4C
KDM4A-4C
Unique target genes
Common target genes
E
Target gene expression changes of KDM4 members
(promoter regions)
1.072e−29
1.173e−28
1.1e−31
0.000131
1.3e−48
6.12e−194
3.72e−21
F
Target gene expression changes of KDM5 members
(promoter regions)
3.84e−75
1.827e−206
0.000592
3.885e−47
4.11e−07
3.808e−11
0.0192
KDM5A
KDM5B
KDM5C
KDM5A-5B
KDM5B-5C
KDM5A-5C
KDM5A-5B-5C
G
Chr17
35,640 kb
35,645 kb
KDM4A
KDM4B
KDM4C
H3K27ac
H3K4me1
H2AK119ub1
H3K27me3
H3K4me3
RNA-seq
Pou5f1
Pou5f1
H
Chr5
148,115 kb
148,120 kb
Cdx2
I
Chr11
119,150 kb
119,160 kb
119,170 kb
KDM5A
KDM5B
KDM5C
H3K27ac
H3K4me1
H2AK119ub1
H3K27me3
H3K4me3
RNAseq
Eif4a3
J
Chr6
52,110 kb
52,120 kb
52,130 kb
52,140 kb
52,150 kb
52,160 kb
52,170 kb
52,180 kb
Hoxa1
Hoxa2
Hoxaas2
Hoxa4
Hoxaas3
Hoxa7
Hoxa9
Hoxa3
Hoxa5
Hoxa6
K

Figure 4

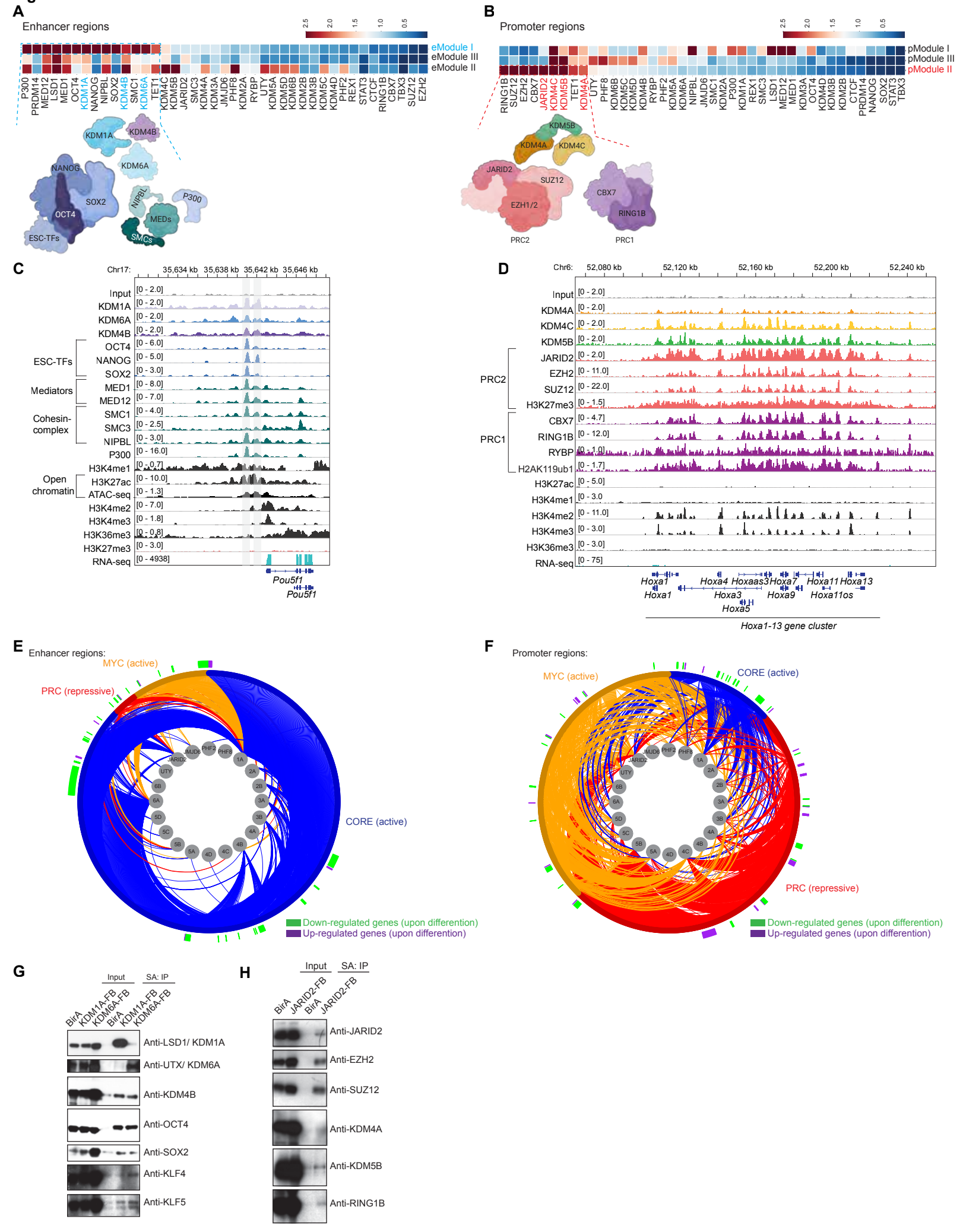

A
Enhancer regions
eModule I
eModule III
eModule II
P300 PRDM14 MED12 LSD1 MED1 OCT4 KDM1A NANOG NIPBL SOX2 KDM4B SMC1 KDM6A TET1 KDM4C KDM5B JARID2 SMC3 KDM4A KDM3A JMJD6 PHF8 KDM2A RYBP UTY KDM5A KDM5D KDM6B KDM2B KDM3B KDM5C KDM4D PHF2 REX1 STAT3 CTCF RING1B CBX7 TBX3 SUZ12 EZH2
KDM1A
KDM4B
KDM6A
NANOG
SOX2
OCT4
NIPBL
P300
MEDs
SMCs
ESC-TFs
B
Promoter regions
pModule I
pModule III
pModule II
RING1B SUZ12 EZH2 CBX7 JARID2 KDM4C KDM5B TET1 KDM4A UTY PHF8 KDM6B KDM5C KDM5D KDM4B RYBP PHF2 KDM5A KDM6A NIPBL JMJD6 SMC1 KDM2A P300 KDM1A REX1 SMC3 LSD1 MED12 MED1 KDM3A OCT4 KDM4D KDM3B KDM2B CTCF PRDM14 NANOG SOX2 STAT3 TBX3
KDM5B
KDM4A
KDM4C
JARID2
SUZ12
EZH1/2
CBX7
RING1B
PRC2
PRC1
C
Chr17:
35,634 kb 35,638 kb 35,642 kb 35,646 kb
Input
KDM1A
KDM6A
KDM4B
ESC-TFs
OCT4
NANOG
SOX2
Mediators
MED1
MED12
Cohesin-complex
SMC1
SMC3
NIPBL
P300
Open chromatin
H3K4me1
H3K27ac
ATAC-seq
H3K4me2
H3K4me3
H3K36me3
H3K27me3
RNA-seq
Pou5f1
Pou5f1
D
Chr6:
52,080 kb 52,120 kb 52,160 kb 52,200 kb 52,240 kb
Input
KDM4A
KDM4C
KDM5B
PRC2
JARID2
EZH2
SUZ12
H3K27me3
PRC1
CBX7
RING1B
RYBP
H2AK119ub1
H3K27ac
H3K4me1
H3K4me2
H3K4me3
H3K36me3
RNA-seq
Hoxa1 Hoxa4 Hoxaas3 Hoxa7 Hoxa11 Hoxa13
Hoxa1 Hoxa3 Hoxa9 Hoxa11os
Hoxa5
Hoxa1-13 gene cluster
E
Enhancer regions:
MYC (active)
PRC (repressive)
CORE (active)
Down-regulated genes (upon differentiation)
Up-regulated genes (upon differentiation)
F
Promoter regions:
MYC (active)
CORE (active)
PRC (repressive)
Down-regulated genes (upon differentiation)
Up-regulated genes (upon differentiation)
G
Input
SA: IP
BirA
KDM1A-FB
KDM6A-FB
BirA
KDM1A-FB
KDM6A-FB
Anti-LSD1/ KDM1A
Anti-UTX/ KDM6A
Anti-KDM4B
Anti-OCT4
Anti-SOX2
Anti-KLF4
Anti-KLF5
H
Input
SA: IP
BirA
JARID2-FB
BirA
JARID2-FB
Anti-JARID2
Anti-EZH2
Anti-SUZ12
Anti-KDM4A
Anti-KDM5B
Anti-RING1B

Figure 5

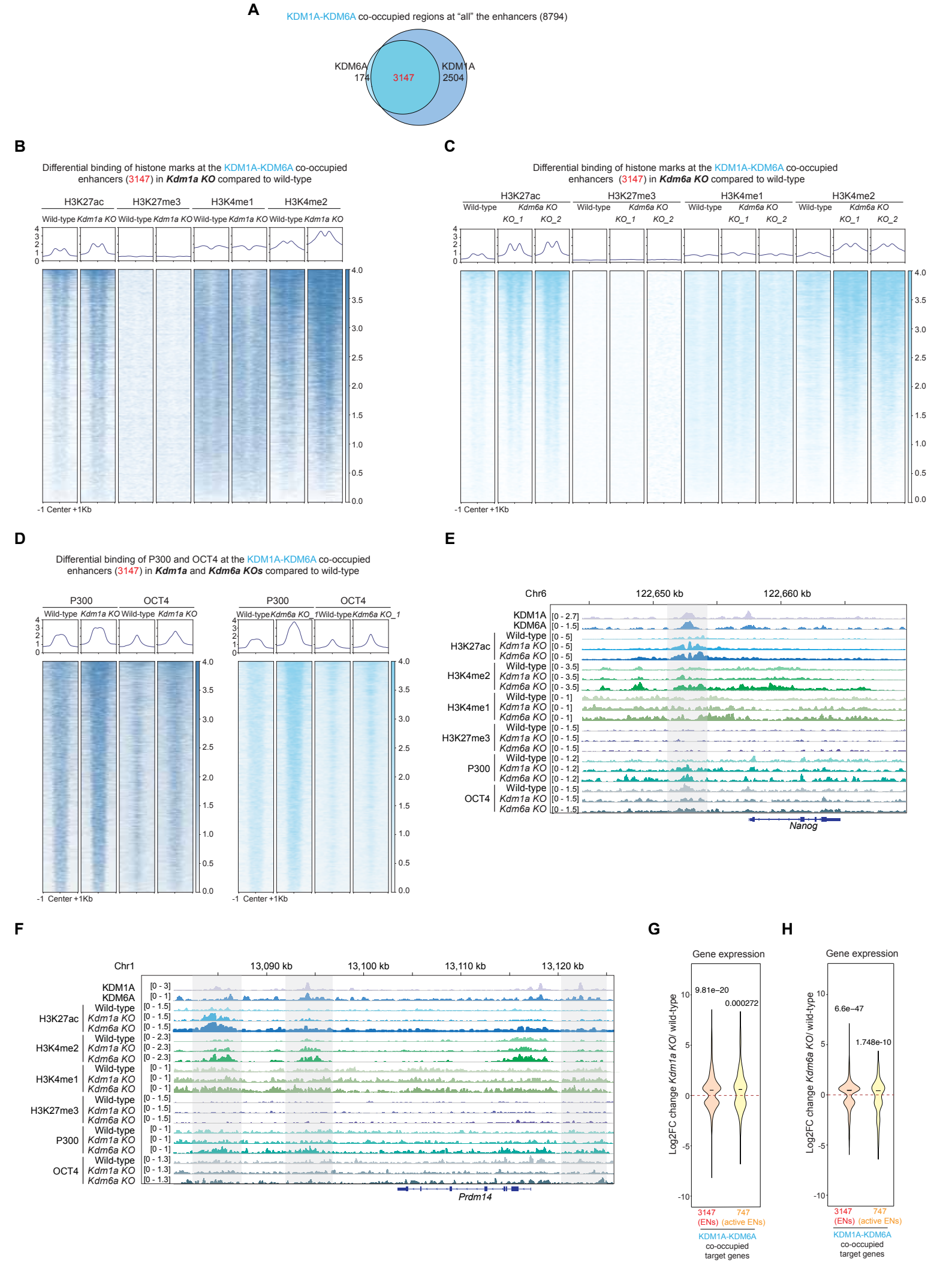

A
KDM1A-KDM6A co-occupied regions at "all" the enhancers (8794)
KDM6A
174
3147
KDM1A
2504
B
Differential binding of histone marks at the KDM1A-KDM6A co-occupied enhancers (3147) in Kdm1a KO compared to wild-type
H3K27ac
H3K27me3
H3K4me1
H3K4me2
Wild-type Kdm1a KO
-1 Center +1Kb
C
Differential binding of histone marks at the KDM1A-KDM6A co-occupied enhancers (3147) in Kdm6a KO compared to wild-type
Wild-type Kdm6a KO
KO_1 KO_2
D
Differential binding of P300 and OCT4 at the KDM1A-KDM6A co-occupied enhancers (3147) in Kdm1a and Kdm6a KOs compared to wild-type
P300
OCT4
E
Chr6
122,650 kb
122,660 kb
KDM1A
KDM6A
H3K27ac
H3K4me2
H3K4me1
H3K27me3
P300
OCT4
Wild-type
Kdm1a KO
Kdm6a KO
Nanog
F
Chr1
13,090 kb
13,100 kb
13,110 kb
13,120 kb
Prdm14
G
Gene expression
Log2FC change Kdm1a KO/ wild-type
9.81e−20
0.000272
3147 (ENs)
747 (active ENs)
KDM1A-KDM6A co-occupied target genes
H
Gene expression
Log2FC change Kdm6a KO/ wild-type
6.6e−47
1.748e-10
3147 (ENs)
747 (active ENs)
KDM1A-KDM6A co-occupied target genes

Figure 6

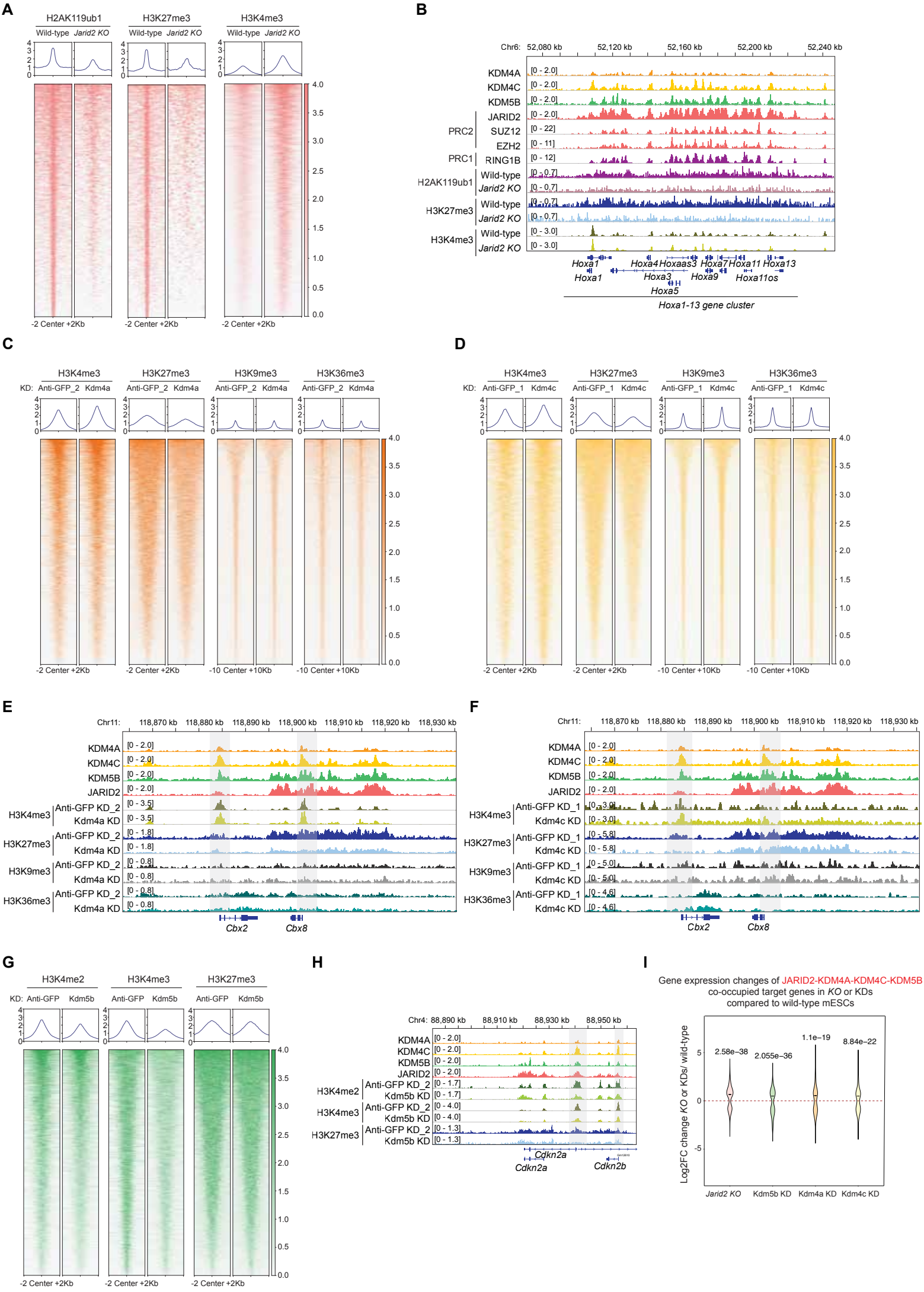
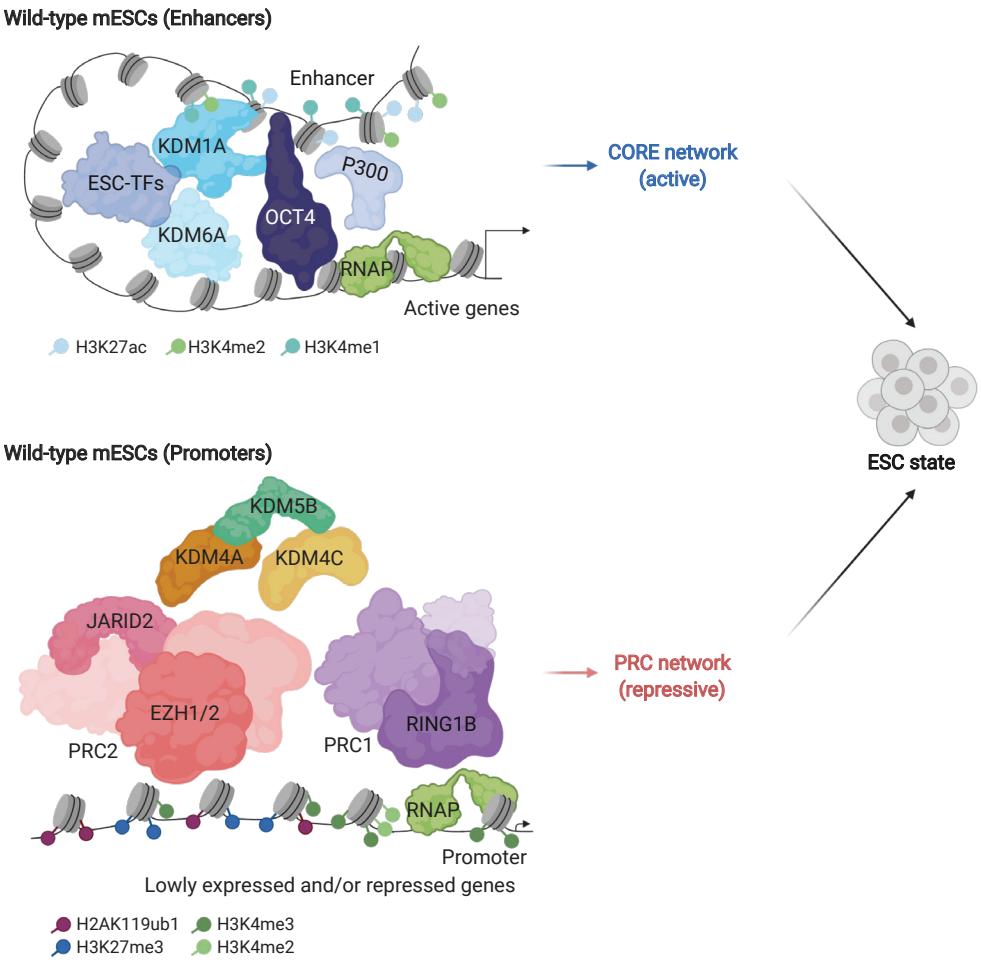

A
H2AK119ub1
H3K27me3
H3K4me3
Wild-type Jarid2 KO
-2 Center +2Kb
B
Chr6: 52,080 kb 52,120 kb 52,160 kb 52,200 kb 52,240 kb
KDM4A
KDM4C
KDM5B
JARID2
SUZ12
EZH2
RING1B
PRC2
PRC1
H2AK119ub1
H3K27me3
H3K4me3
Wild-type
Jarid2 KO
Hoxa1-13 gene cluster
C
H3K4me3
H3K27me3
H3K9me3
H3K36me3
KD: Anti-GFP_2 Kdm4a
-2 Center +2Kb
-10 Center +10Kb
D
KD: Anti-GFP_1 Kdm4c
E
Chr11: 118,870 kb 118,880 kb 118,890 kb 118,900 kb 118,910 kb 118,920 kb 118,930 kb
Anti-GFP KD_2
Kdm4a KD
Cbx2
Cbx8
F
Anti-GFP KD_1
Kdm4c KD
G
H3K4me2
Anti-GFP Kdm5b
H
Chr4: 88,890 kb 88,910 kb 88,930 kb 88,950 kb
Anti-GFP KD_2
Kdm5b KD
Cdkn2a
Cdkn2b
I
Gene expression changes of JARID2-KDM4A-KDM4C-KDM5B co-occupied target genes in KO or KDs compared to wild-type mESCs
Log2FC change KO or KDs/ wild-type
2.58e−38
2.055e−36
1.1e−19
8.84e−22
Jarid2 KO
Kdm5b KD
Kdm4a KD
Kdm4c KD

Figure 7


Wild-type mESCs (Enhancers)
Enhancer
KDM1A
ESC-TFs
P300
OCT4
KDM6A
RNAP
Active genes
H3K27ac
H3K4me2
H3K4me1
CORE network
(active)
ESC state
Wild-type mESCs (Promoters)
KDM5B
KDM4A
KDM4C
JARID2
EZH1/2
PRC2
PRC1
RING1B
RNAP
Promoter
Lowly expressed and/or repressed genes
H2AK119ub1
H3K4me3
H3K27me3
H3K4me2
PRC network
(repressive)

SUPPLEMENTAL INFORMATION

Figure S1. The HDMome signature reveals the combinatorial co-occupancy of multiple HDMs

(A) Genome-wide distribution of 20 HDMs.

(B-D) Venn diagrams show overlapped binding regions of multiple of HDMs at genome-wide, enhancers and promoters.

(C) Genome browser representation of ChIP-seq normalised reads for multiple HDMs at the *Nanog* gene locus. The RNA-seq track demonstrates expression of the *Nanog* gene in wild-type mESCs.

Related to Figure 1.

Figure S2. Multiple HDMs combinatorially regulate gene expression

(A) Gene ontology (GO) term analyses (biological processes) of KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2) co-occupied target genes.

Related to Figure 2.

Figure S3. Combinatorial gene regulatory functions of KDM4 and KDM5 members

(A-B) Venn diagrams show overlapped binding sites of KDM4A, KDM4B and KDM4C at the enhancers (A) and promoters (B).

(C-D) Genomic tracks represent ChIP-seq normalised reads for KDM4 members and histone marks at the two different gene loci (*Klf2* and *Med23*). The RNA-seq tracks demonstrates expression of these genes in wild-type mESCs.

(E-F) Venn diagrams display overlapped binding sites of KDM5A, KDM5B and KDM5C at the enhancers (E) and promoters (F).

(G) Genomic tracks represent ChIP-seq normalised reads for KDM5 members and histone marks at the *Zfp507* gene locus. The RNA-seq track demonstrates expression of this gene in wild-type mESCs.

Related to Figure 3.

Figure S4. HDM modules cooperate with key ESC regulatory networks for gene regulation

(A) Enriched TF motifs at the KDM1A-KDM4B-KDM6A co-occupied enhancers.

(B) Enriched TF motifs at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters.

(C) Western blots of superpose-6 gel-filtration column fractions showing the eluted fraction numbers and molecular weight of the markers (arrows).

Related to Figure 4.

Figure S5. KDM1A and KDM6A combinatorially modulate P300/H3K27ac, H3K4me1, H3K4me2 deposition and OCT4 recruitment at the enhancers for target gene expression

(A) Heat map represents co-occupancy of KDM1A, KDM1A-FB, KDM6A, KDM6A-FB, ESC-TFs (OCT4, NANOG, SOX2), MLL4, P300 and histone marks (H3K7ac, H3K4me1, H3K4me2 and H3K27me3). Our KDM6A and KDM6A-FB ChIP-seq data, and published KDM6A (pub) ChIP-seq data were used for the comparison.

(B-C) Western blot analysis confirmed the *Kdm1a* and *Kdm6a KO (Knockout)* mESC lines. Highlighted *KO* mESC lines (blue in colour) were used for further experiments. Two individual *Kdm6a KOs* were used to assess the reproducibility of the data.

(D) Venn diagram shows KDM1A and KDM1A-FB (5,651) co-occupied sites; KDM6A and KDM6A-FB (3,321) co-occupied sites at “all” the enhancers (8794).

(E) Differential binding of H3K7ac, H3K27me3, H3K4me1 and H3K4me2 at the KDM1A and KDM1A-FB (5,651) co-occupied enhancers in *Kdm1a KO* compared to wild-type.

(F) Differential binding of H3K7ac, H3K27me3, H3K4me1 and H3K4me2 at the KDM6A and KDM6A-FB (3,321) co-occupied enhancers in *Kdm6a KOs* compared to wild-type.

(G) Venn diagram shows KDM1A and KDM1A-FB (2,021) co-occupied sites and KDM6A and KDM6A-FB (833) co-occupied sites at the “active” enhancers (4,346). “Active” enhancers contained H3K27ac and H3K4me1 enhancer marks, but without of H3K4me3 and ± 2kb of TSS as promoter regions (i.e without ± 2kb TSS, NO H3K4me3, +H3K27ac, +H3K4me1). The “active” enhancers (4,346) represent as a sub-set of “all” enhancers (8,794).

(H) Differential binding of H3K7ac, H3K27me3, H3K4me1 and H3K4me2 at the KDM1A and KDM1A-FB (2,021) co-occupied “active” enhancers in *Kdm1a KO* compared to wild-type.

(I) Differential binding of H3K7ac, H3K27me3, H3K4me1 and H3K4me2 at the KDM6A and KDM6A-FB (833) co-occupied “active” enhancers in *Kdm6a KOs* compared to wild-type.

(J) Gene expression changes of KDM1A target genes (5,651 and 2,021) in *Kdm1a KO* compared to wild-type.

(K) Gene expression changes of KDM6A target genes (3,321 and 833) in *Kdm6a KO* compared to wild-type.

(L) Venn diagram show high-confidence KDM1A-KDM6A co-occupied sites (747) at the “active” mESC-specific enhancers (4,346). KDM1A and KDM1A-FB co-occupied sites (2,021), and KDM6A and KDM6A-FB (833) co-occupied sites at the “active” enhancers were used to estimate the KDM1A-KDM6A overlapping binding sites (747) (see Figure S5G).

(M-N) Heat maps exhibit differential binding of H3K7ac, H3K27me3, H3K4me1 and H3K4me2 at the KDM1A-KDM6A co-occupied “active” enhancers (747) in *Kdm1a KO* (M) and *Kdm6a KOs* (N). Two individual *Kdm6a KOs* were used to examine the reproducibility of the data.

(O) Heat maps display differential binding of P300 and OCT4 at the KDM1A-KDM6A co-occupied “active” enhancers (747) in *Kdm1a* and *Kdm6a KOs*.

Related to Figure 5.

Figure S6. JARID2, KDM4A, KDM4C and KDM5B co-operatively control H2AK119ub1 (of PRC1) and bivalent marks (of PRC2) at the promoters for target gene repression

(A-C) Western blot analysis confirmed the *Jarid2 KO,* Kdm4a KD and Kdm5b KD mESC lines. The highlighted Kdm4a KD and Kdm5b KD mESC lines (blue) were used for further experiments.

(D) Genomic tracks demonstrate differential binding of histone marks (H2AK119ub1, H3K27me3 and H3K4me3) at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters of *Cdkn2a, 2b* gene loci in the *Jarid2 KO* compared to wild-type. The highlighted regions show differential binding of histone marks (*Jarid2 KO* v. wild-type) at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters.

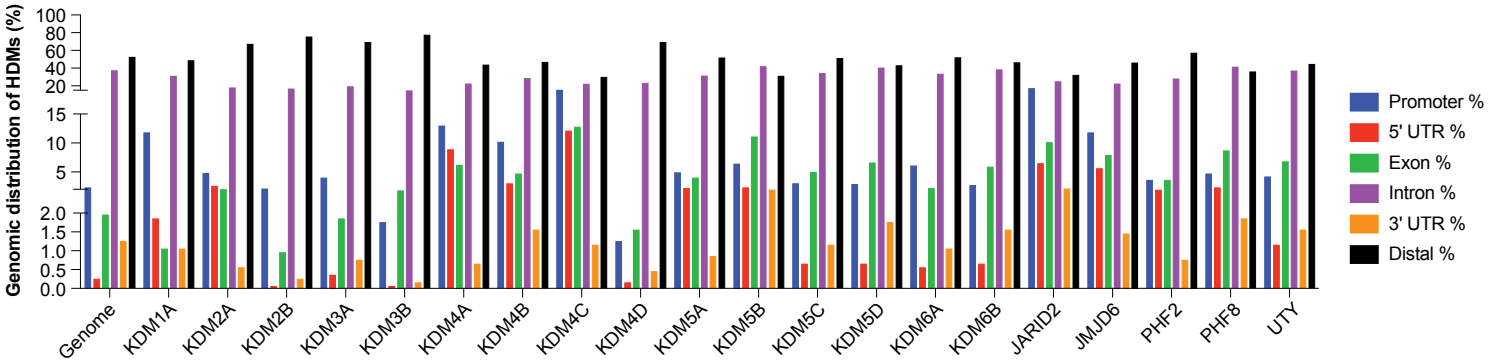
(E-F) Genomic tracks display differential binding of H3K4me3, H3K27me3, H3K9me3 and H3K36me3 at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters of *Cdkn2a, 2b* gene loci in the Kdm4a (E) and Kdm4c (F) KDs compared to the anti-GFP KD (control). The highlighted regions show differential binding of histone marks (KD v. control) at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters.

(G) Genomic tracks show differential binding of H3K4me2, H3K4me3 and H3K27me3 at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters of the *Hoxa1-13* gene cluster in the Kdm5b KD compared to the anti-GFP KD (control).

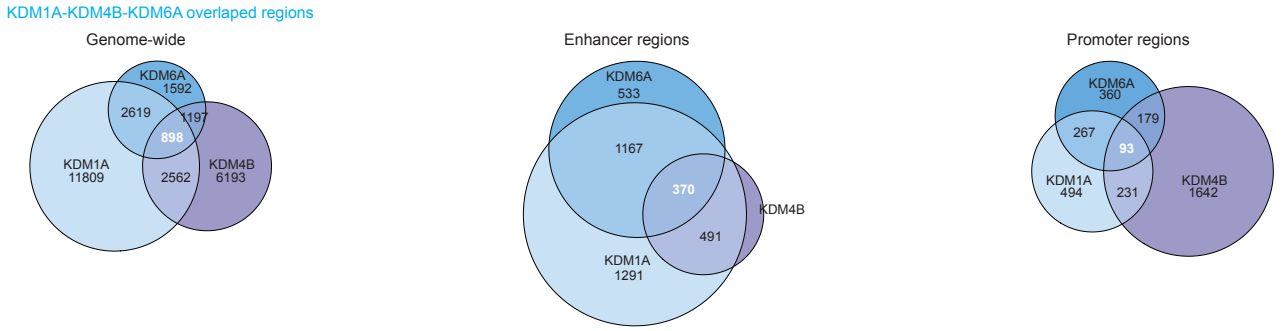
Related to Figure 6.

Figure S1

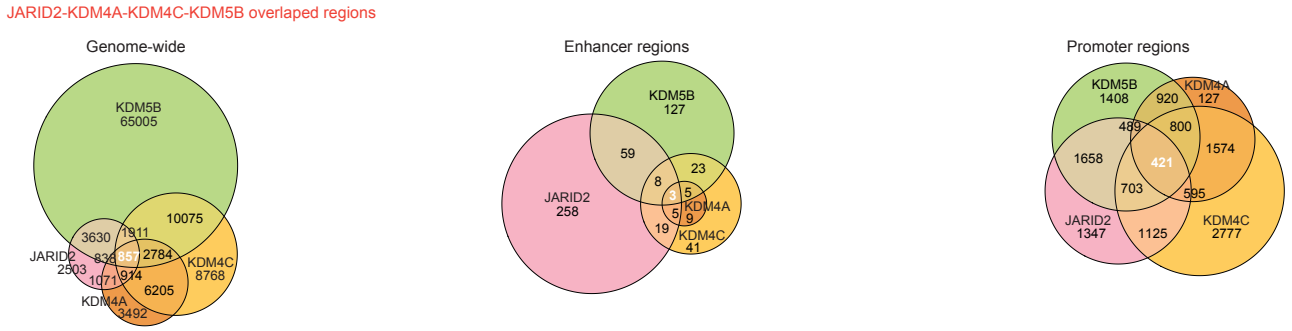
A



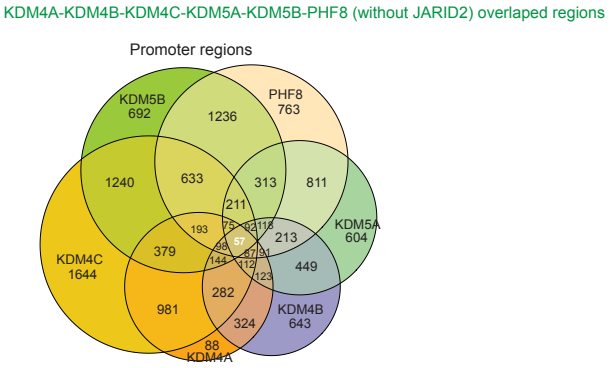
B



C



D



E

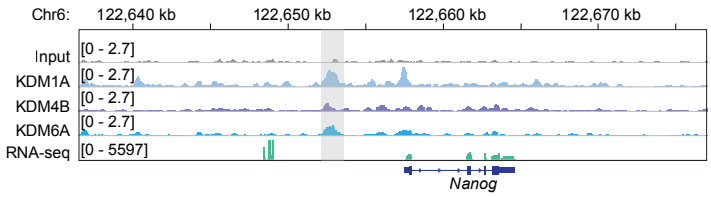


Figure S2

A

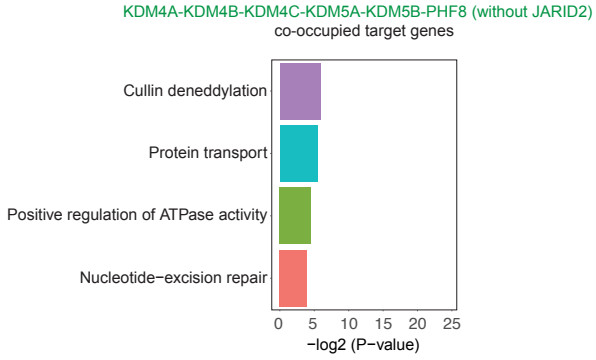
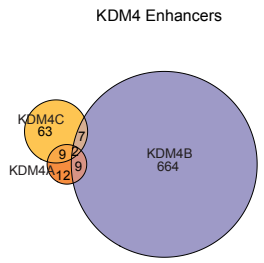
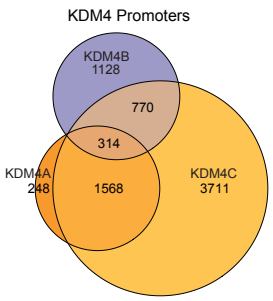

KDM4A-KDM4B-KDM4C-KDM5A-KDM5B-PHF8 (without JARID2)
co-occupied target genes
Cullin deneddylation
Protein transport
Positive regulation of ATPase activity
Nucleotide−excision repair
0
5
10
15
20
25
−log2 (P−value)

Figure S3

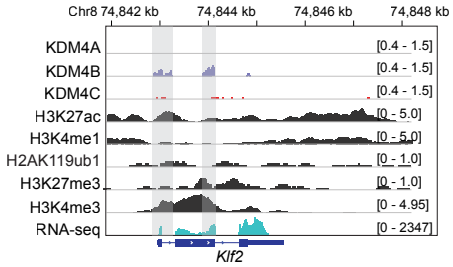
A



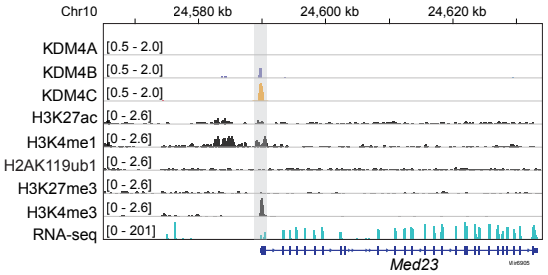
B



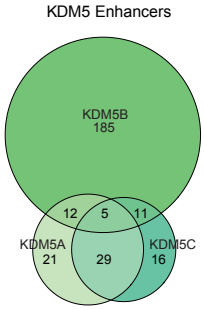
C



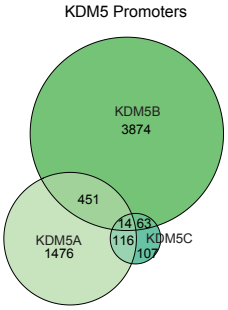
D



E



F



G

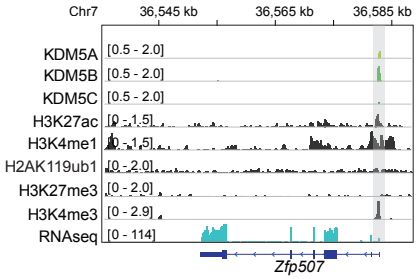


Figure S4

A

TF motifs at the KDM1A-KDM4B-KDM6A co-occupied enhancers

MOTIFs	Names	P-values
	KLF5	7.71e-15
	KLF4	9.99e-08
	ESRRB	1.50e-08
	OCT4::SOX2	2.21e-07
	SOX2	1.57e-04

B

TF motifs at the JARID2-KDM4A-KDM4C-KDM5B co-occupied promoters

MOTIFs	Names	P-values
	ZBTB33	1.29e-04
	HINFP	1.29e-04
	TFAP2A	6.88e-05
	TFAP2C	3.20e-03
	ZBTB7A	1.03e-03

C

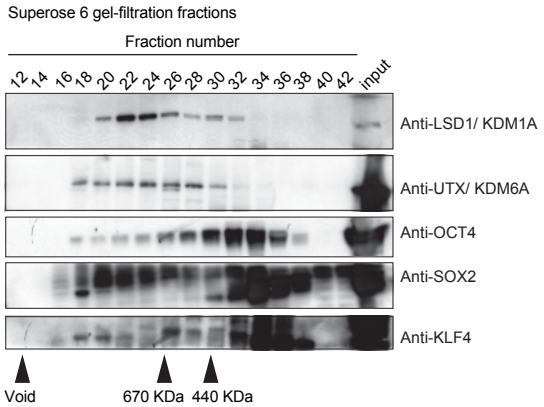
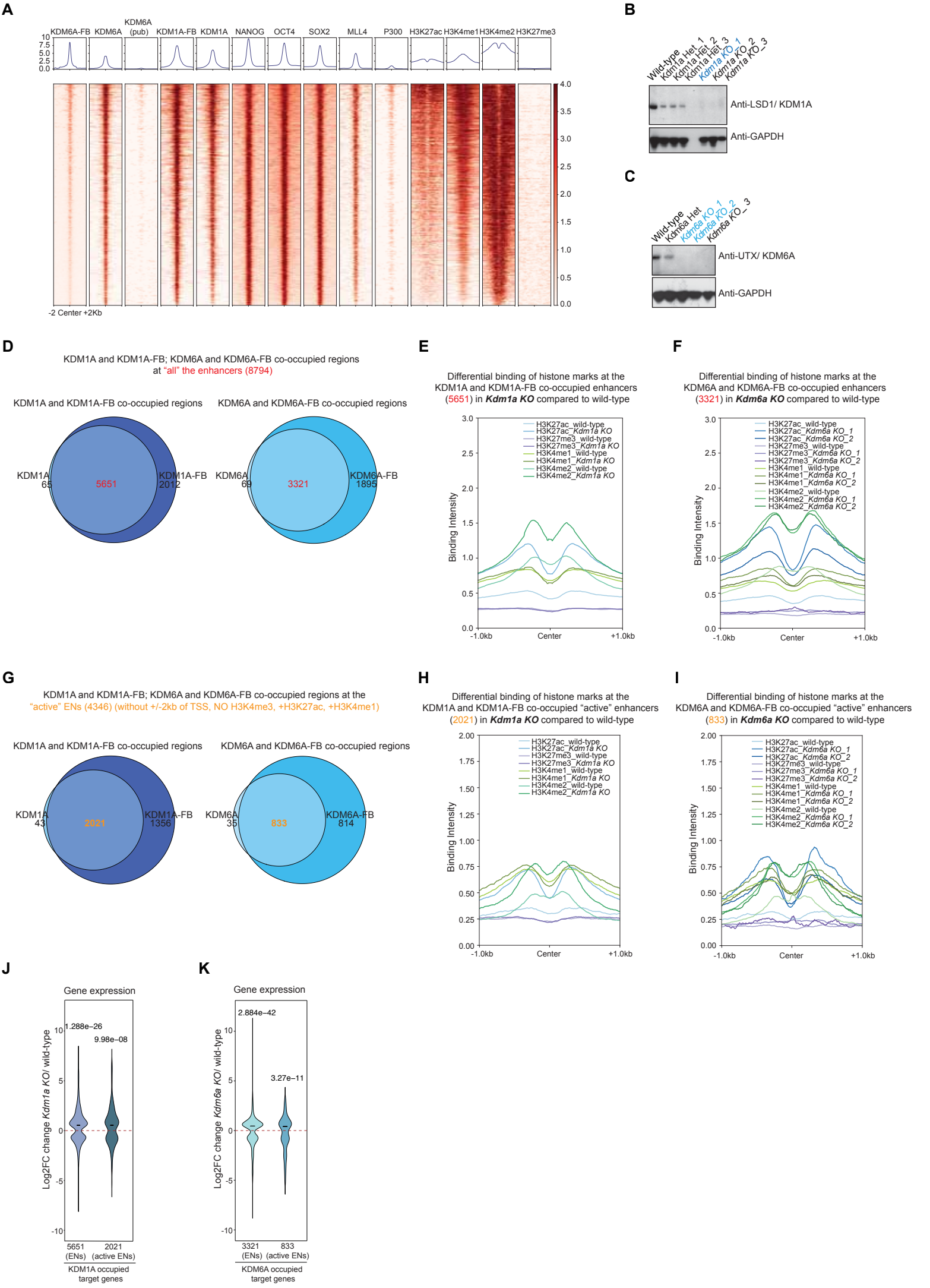
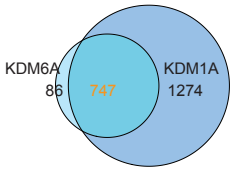


Figure S5


A
KDM6A-FB
KDM6A
KDM6A (pub)
KDM1A-FB
KDM1A
NANOG
OCT4
SOX2
MLL4
P300
H3K27ac
H3K4me1
H3K4me2
H3K27me3
-2 Center +2Kb
B
Wild-type
Kdm1a Het_1
Kdm1a Het_2
Kdm1a Het_3
Kdm1a KO_1
Kdm1a KO_2
Kdm1a KO_3
Anti-LSD1/ KDM1A
Anti-GAPDH
C
Wild-type
Kdm6a Het
Kdm6a KO_1
Kdm6a KO_2
Kdm6a KO_3
Anti-UTX/ KDM6A
Anti-GAPDH
D
KDM1A and KDM1A-FB; KDM6A and KDM6A-FB co-occupied regions
at "all" the enhancers (8794)
KDM1A and KDM1A-FB co-occupied regions
KDM6A and KDM6A-FB co-occupied regions
KDM1A 65
5651
KDM1A-FB 2012
KDM6A 69
3321
KDM6A-FB 1895
E
Differential binding of histone marks at the
KDM1A and KDM1A-FB co-occupied enhancers
(5651) in Kdm1a KO compared to wild-type
Binding Intensity
-1.0kb
Center
+1.0kb
F
Differential binding of histone marks at the
KDM6A and KDM6A-FB co-occupied enhancers
(3321) in Kdm6a KO compared to wild-type
Binding Intensity
-1.0kb
Center
+1.0kb
G
KDM1A and KDM1A-FB; KDM6A and KDM6A-FB co-occupied regions at the
"active" ENs (4346) (without +/-2kb of TSS, NO H3K4me3, +H3K27ac, +H3K4me1)
KDM1A and KDM1A-FB co-occupied regions
KDM6A and KDM6A-FB co-occupied regions
KDM1A 43
2021
KDM1A-FB 1356
KDM6A 35
833
KDM6A-FB 814
H
Differential binding of histone marks at the
KDM1A and KDM1A-FB co-occupied "active" enhancers
(2021) in Kdm1a KO compared to wild-type
Binding Intensity
-1.0kb
Center
+1.0kb
I
Differential binding of histone marks at the
KDM6A and KDM6A-FB co-occupied "active" enhancers
(833) in Kdm6a KO compared to wild-type
Binding Intensity
-1.0kb
Center
+1.0kb
J
Gene expression
Log2FC change Kdm1a KO/ wild-type
1.288e−26
9.98e−08
5651 (ENs)
2021 (active ENs)
KDM1A occupied target genes
K
Gene expression
Log2FC change Kdm6a KO/ wild-type
2.884e−42
3.27e−11
3321 (ENs)
833 (active ENs)
KDM6A occupied target genes

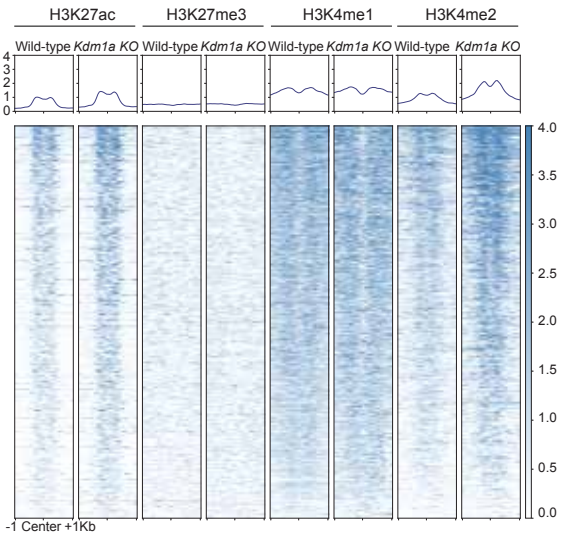
L

KDM1A-KDM6A co-occupied regions at "active" ENs (4346)
(without +/-2kb of TSS, NO H3K4me3, +H3K27ac, +H3K4me1)



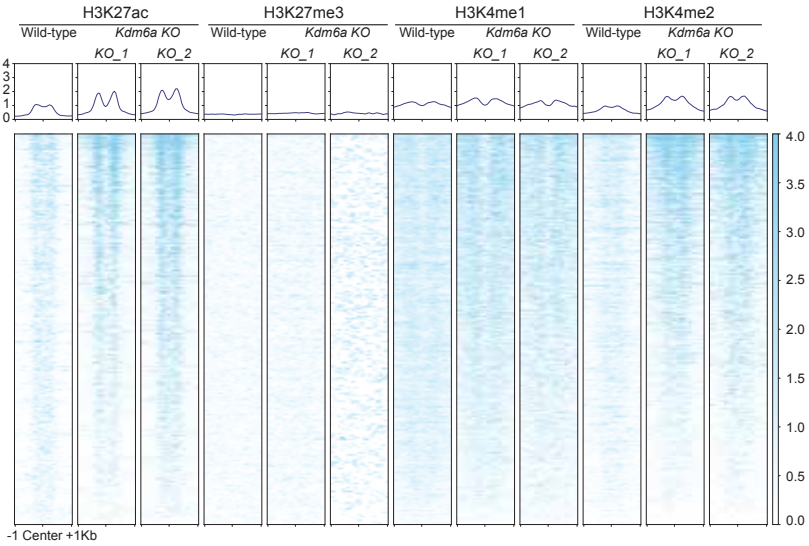
M

Differential binding of histone marks at the KDM1A-KDM6A co-occupied "active" enhancers (747) in ***Kdm1a KO*** compared to wild-type



N

Differential binding of histone marks at the KDM1A-KDM6A co-occupied "active" enhancers (747) in ***Kdm6a KO*** compared to wild-type



O

Differential binding of P300 and OCT4 at the KDM1A-KDM6A co-occupied "active" enhancers (747) in ***Kdm1a*** and ***Kdm6a KOs*** compared to wild-type

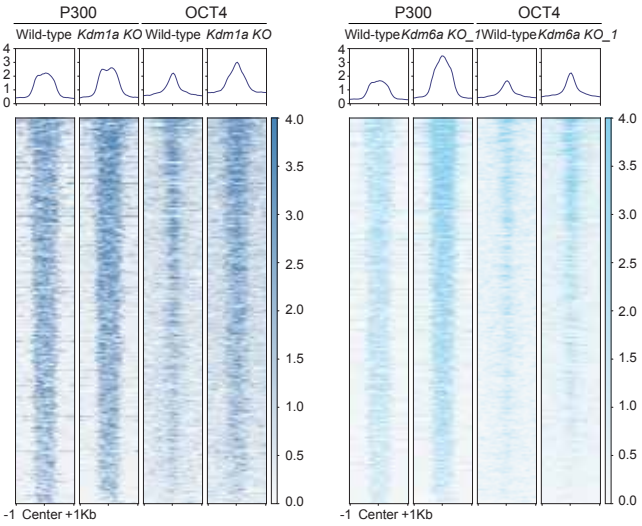
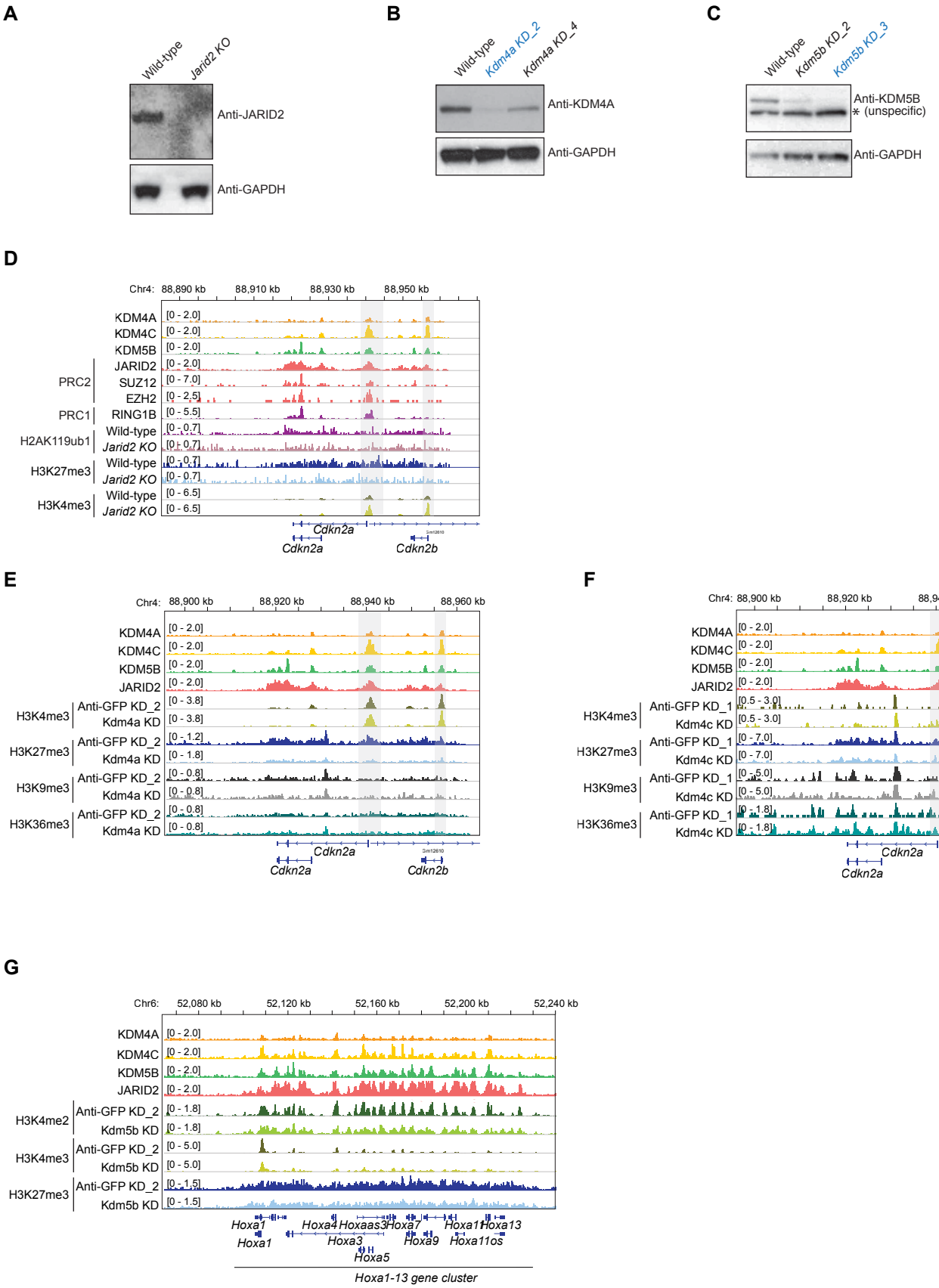


Figure S6


A
Wild-type
Jarid2 KO
Anti-JARID2
Anti-GAPDH
B
Wild-type
Kdm4a KD_2
Kdm4a KD_4
Anti-KDM4A
Anti-GAPDH
C
Wild-type
Kdm5b KD_2
Kdm5b KD_3
Anti-KDM5B
* (unspecific)
Anti-GAPDH
D
Chr4: 88,890 kb
88,910 kb
88,930 kb
88,950 kb
KDM4A
KDM4C
KDM5B
JARID2
PRC2
SUZ12
EZH2
PRC1
RING1B
H2AK119ub1
Wild-type
Jarid2 KO
H3K27me3
H3K4me3
Cdkn2a
Cdkn2b
E
Chr4: 88,900 kb
88,920 kb
88,940 kb
88,960 kb
H3K4me3
H3K27me3
H3K9me3
H3K36me3
Anti-GFP KD_2
Kdm4a KD
F
Chr4: 88,900 kb
88,920 kb
88,940 kb
88,960 kb
Anti-GFP KD_1
Kdm4c KD
G
Chr6: 52,080 kb
52,120 kb
52,160 kb
52,200 kb
52,240 kb
H3K4me2
Kdm5b KD
Hoxa1
Hoxa4
Hoxaas3
Hoxa7
Hoxa1
Hoxa13
Hoxa3
Hoxa9
Hoxa11os
Hoxa5
Hoxa1-13 gene cluster