iScience

Article

Integrative analysis reveals histone demethylase LSD1 promotes RNA polymerase II pausing

Integrative analysis to define LSD1 interaction partners

Split genomic domains

Cluster factors by domain

Test enrichment of RNAPII pausing factors

>400 public ChIP-seq datasets

Transcription factors
Chromatin remodellers
Histone modifications

Genomic domains

Histone mark emission annotations

Correlation with LSD1

● RNAPII pausing
○ Others

Ranking of all factors

Decrease in RNAPII pausing at LSD1 sites upon LSD1 knockdown

Perform ChIP-seq

Calculate changes in RNAPII pausing

ChIP-seq signal

Input
LSD1
RNAPII
RNAPII

Input
Control
Control
Knockdown

Normalised Pol2 signal

Control

LSD1 Knockdown

— Input
— Pol2

TSS   TES   TSS   TES

Differential regulation of LSD1 and MYC regulated sites

● H3K27me3
● H3K27ac
● H3K4me3

LSD1-specific
pausing index
0 — Max

CoREST
CDK9
HDAC1/2
LSD1
POL2
Cell-type genes

MYC-specific
pausing index
0 — Max

MYC
POL2
Housekeeping genes

Co-localized
pausing index
0 — Max

CoREST
HDAC1/2
LSD1
CDK9
SPT5
MYC
NELF-A
POL2
TBP
Housekeeping genes

Hani Jieun Kim, Pishun Li, Taiyun Kim, Andrew J. Oldfield, Xiaofeng Zheng, Pengyi Yang

zheng.x@hunau.edu.cn (X.Z.)
pengyi.yang@sydney.edu.au (P.Y.)

Highlights

LSD1 promotes RNA polymerase II pausing in mouse embryonic stem cells

LSD1 knockdown causes global reduction of RNAPII pausing

Co-localized sites of LSD1 and MYC are enriched for RNAPII pausing and releasing factors

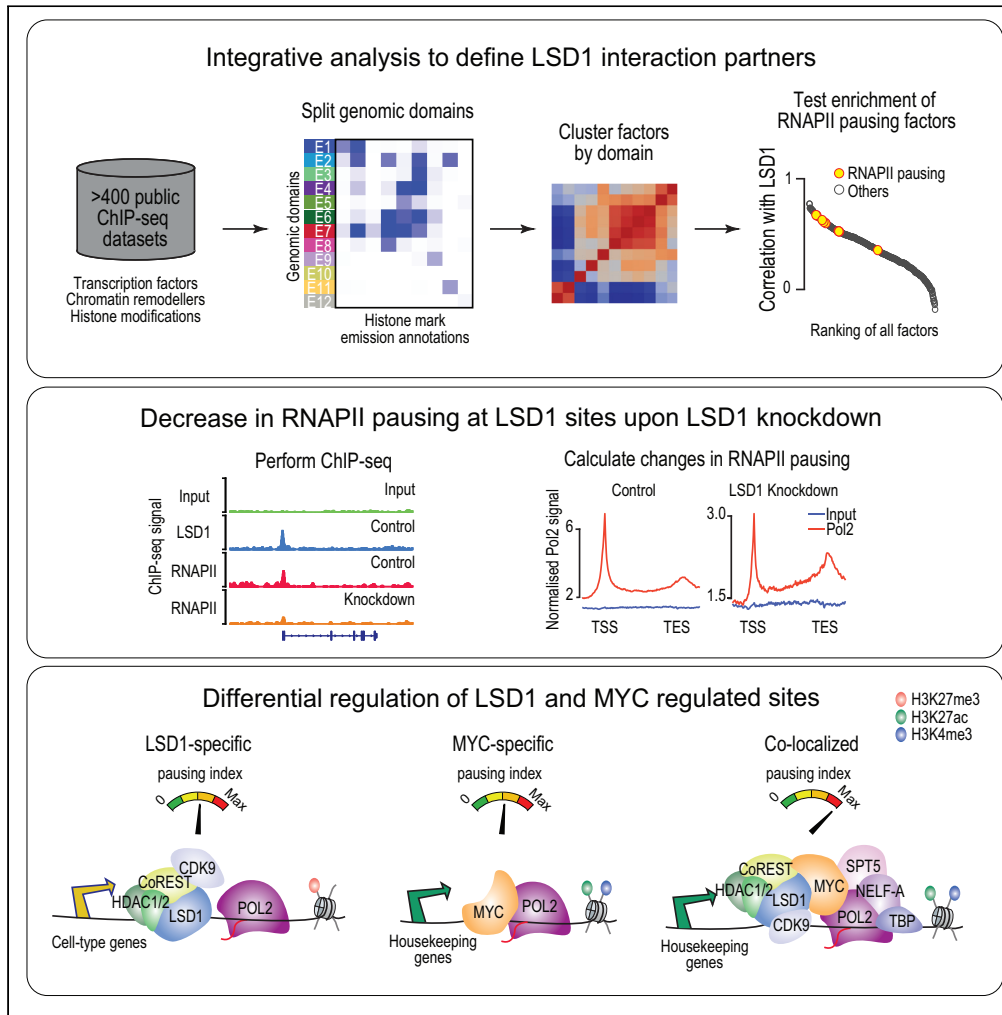MYC recruitment to co-regulated genes is dependent on LSD1 but not vice versa

iScience

Article

Integrative analysis reveals histone demethylase LSD1 promotes RNA polymerase II pausing

Hani Jieun Kim,[1,3,6] Pishun Li,[2,6] Taiyun Kim,[1,3,4] Andrew J. Oldfield,[5] Xiaofeng Zheng,[2,*] and Pengyi Yang[1,3,4,7,*]

SUMMARY

Lysine-specific demethylase 1 (LSD1) is well-known for its role in decommissioning enhancers during mouse embryonic stem cell (ESC) differentiation. Its role in gene promoters remains poorly understood despite its widespread presence at these sites. Here, we report that LSD1 promotes RNA polymerase II (RNAPII) pausing, a rate-limiting step in transcription regulation, in ESCs. We found the knockdown of LSD1 preferentially affects genes with higher RNAPII pausing. Next, we demonstrate that the co-localization sites of LSD1 and MYC, a factor known to regulate pause-release, are enriched for other RNAPII pausing factors. We show that LSD1 and MYC directly interact and MYC recruitment to genes co-regulated with LSD1 is dependent on LSD1 but not vice versa. The co-regulated gene set is significantly enriched for housekeeping processes and depleted of transcription factors compared to those bound by LSD1 alone. Collectively, our integrative analysis reveals a pleiotropic role of LSD1 in promoting RNAPII pausing.

INTRODUCTION

Lysine-specific demethylase 1 (LSD1) (Shi et al., 2004), also known as KDM1A/AOF2, plays a pleiotropic role in a broad spectrum of biological processes including stem cell pluripotency, differentiation, and cancer (Amente et al., 2013). It binds to both promoters and enhancers to specifically catalyze the demethylation at H3K4 and H3K9 (Shi et al., 2004). In embryonic stem cells (ESCs), the inhibition or ablation of LSD1 has been shown to lead to severe proliferative defects and cell death (Adamo et al., 2011; Wang et al., 2009; Whyte et al., 2012). Many studies have provided further insight into the role of LSD1 as enhancers in diverse cell types including ESCs (AlAbdi et al., 2020; Petell et al., 2016; Tatsumi et al., 2020; Vinckier et al., 2020; Whyte et al., 2012). For example, LSD1 has been reported to decommission enhancers of ESC-specific genes during mouse ESC differentiation (Whyte et al., 2012). While these studies demonstrate the functional role of LSD1 at enhancers in ESCs, the functional implication of LSD1 binding at gene promoters remains largely unknown.

At the promoters of actively transcribed genes, pausing the RNA polymerase II (RNAPII) during early elongation is an obligate part of the transcription cycle that is experienced by RNAPII at almost all genes (Core and Adelman, 2019) across the cell and tissue types (Day et al., 2016). Blocking the release of the paused RNAPII through the use of potent small molecule inhibitors has been shown to globally trap RNAPII at promoters, thereby abrogating nearly all RNA synthesis in mammalian cells (Henriques et al., 2013; Jonkers and Lis, 2015; Rahl et al., 2010). Release of the paused RNAPII in promoter-proximal regions has been shown to be the rate-limiting step in productive elongation, making the efficiency of pause release a central determinant of gene expression. Although early studies have shown that paused RNAPII is preferentially found at many developmental genes in the *Drosophila melanogaster* embryo (Zeitlinger et al., 2007), subsequent research has demonstrated that pausing is widespread in higher eukaryotes and is not only limited to developmental genes but also genes associated with essential biological processes such as cell proliferation and stress response (Vinckier et al., 2020). For example, in ESCs, RNAPII pausing has been implicated in key processes such as self-renewal (Min et al., 2011) and differentiation (Core and Adelman, 2019; Gaertner and Zeitlinger, 2014; Jonkers and Lis, 2015; Levine, 2011). Thus, the identification of factors that modulate RNAPII pausing and release at genes of different functions is of great interest toward understanding gene regulation.

[1]Computational Systems Biology Group, Children's Medical Research Institute, Westmead, NSW 2145, Australia

[2]College of Veterinary Medicine, Hunan Agricultural University, Changsha 410128, China

[3]School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006, Australia

[4]Charles Perkins Centre, University of Sydney, Sydney, NSW 2006, Australia

[5]Institut de Génétique Humaine, Université de Montpellier, CNRS-UMR9002, 34000, Montpellier, France

[6]These authors contributed equally

[7]Lead contact

*Correspondence: zheng.x@hunau.edu.cn (X.Z.), pengyi.yang@sydney.edu.au (P.Y.)

https://doi.org/10.1016/j.isci.2022.105049

Many events and factors are involved in the establishment and release of paused RNAPII (Chen et al., 2018; Gamot et al., 2014; Gardini et al., 2014; Meng and Bartholomew, 2018). After recruitment to the promoter region of a gene, RNAPII begins transcribing a short, nascent RNA approximately 25–50 nucleotides long. However, RNAPII halts within the initially transcribed region and remains promoter-proximally paused until it receives further signals. The paused state is stabilized by SPT5 and NELF-A (Yamaguchi et al., 2013), with NELF-A preventing the reactivation of the RNAPII catalytic site (Henriques et al., 2018; Vos et al., 2018a, 2018b). The release of RNAPII into productive RNA synthesis is triggered by the activity of the kinase, positive transcription elongation factor b (P-TEFb), where CDK9 acts as the catalytic subunit of P-TEFb. Recruitment of P-TEFb was proposed to be facilitated by c-MYC (MYC), BRD4, various subunits of the Mediator, and super elongation complexes (Li et al., 2018; Liu et al., 2015; Rahl et al., 2010). Critically, the phosphorylation of SPT5 by P-TEFb has been shown to be directly linked to pause release (Cheng et al., 2012; Guo et al., 2000; Marshall and Price, 1995; Yamada et al., 2006), causing the dissociation of NELF-A from RNAPII to enable reactivation and continued the elongation of the nascent RNA (Cheng and Price, 2007; Wu et al., 2003; Yamada et al., 2006).

Here, we report that LSD1 promotes RNAPII pausing in mouse ESCs. We found that LSD1 presence is pervasive at gene promoters and in particular co-localizes with a number of factors associated with paused RNAPII. Correlation of LSD1 binding sites against a compendium of ChIP-seq datasets demonstrated a strong enrichment of factors involved in RNAP II pausing with LSD1 and the knockdown of LSD1 preferentially affects genes with higher RNAPII pausing than those with lower pausing. Our functional experiments demonstrate that the LSD1 knockdown causes the global reduction of RNAPII pausing of genes and in particular those that are bound by LSD1. Interestingly, our analysis demonstrated shared and unique binding profiles between LSD1 and MYC, a well-known player of RNAPII release in mammalian cells (Rahl et al., 2010), having been demonstrated to recruit the elongation factor SPT5 (Baluapuri et al., 2019) as well as facilitate the recruitment of P-TEFB (Rahl et al., 2010). To further understand the potential interplay between LSD1 and MYC in the regulation of RNAPII pausing, we investigated the co-localization sites of these two factors and found that they are associated with significantly higher RNAPII pausing and more enriched with pausing and releasing factors. Finally, we reveal that LSD1 targets different gene categories depending on MYC occupancy. Collectively, our results unveil LSD1 in regulating RNAPII pausing and thereby gene expression and highlight a role of LSD1 at promoters that may be distinct from its role in ESC differentiation.

RESULTS

LSD1 co-localizes with key components of the RNAPII pausing machinery at gene promoters in embryonic stem cells

We previously developed an analytic tool, PAD (http://pad2.maths.usyd.edu.au/), for identifying TFs, chromatin remodelers, and histone modifications that co-localize on chromatin based on their genomic binding profiles in ESCs, as measured using ChIP-seq (Yang et al., 2017). Here, we extended PAD to enable targeted co-localization analysis within 12 functional genomic regions identified by ChromHMM (Ernst and Kellis, 2017) (Figure S1; see STAR Methods). Previous studies have largely focused on the role of LSD1 as enhancers (Vinckier et al., 2020; Whyte et al., 2012), and despite the pervasiveness of LSD1 binding at promoters (Whyte et al., 2012), its role at promoters remains largely uninvestigated. To investigate the role of LSD1 at promoters, we correlated its binding profiles with other DNA- and chromatin-binding proteins at promoter regions using PAD. Our analysis shows a strong co-localization of LSD1 at promoters with core RNAPII pausing factors (RNAPII, CDK9, NELF-A, and SPT5) and additional factors such as MYC (Baluapuri et al., 2019; Price, 2010; Rahl et al., 2010) and BRD4 (Devaiah et al., 2012; Gatchalian et al., 2018; Itzen et al., 2014; Lu et al., 2015; Patel et al., 2013) (Figure 1A). In particular, LSD1 co-localized most strongly with CDK9, the enzymatic subunit of the pause release factor P-TEFb (Booth et al., 2018; Larochelle et al., 2012; Lu et al., 2016) across the three genomic regions (Figure 1B). When we further investigated the co-localization at the binding sites of LSD1, we found that while the ChIP-seq signal of LSD1-CoREST complex components, CoREST, HDAC1, and HDAC2, were observed at comparable levels at promoters, bivalent domains, and enhancers (Figure 1C), the co-localization of LSD1 with RNAPII pausing and release factors was the strongest at promoters (Figure 1C).

We next assessed the binding profile of LSD1 (n = 14106) across the genome and found that approximately 71.5% of all LSD1 binding sites in ESCs are found at the promoters (Figure 1D). This was akin to the strong promoter-proximal binding preference of other RNAPII pausing and release factors (Figure 1D). The
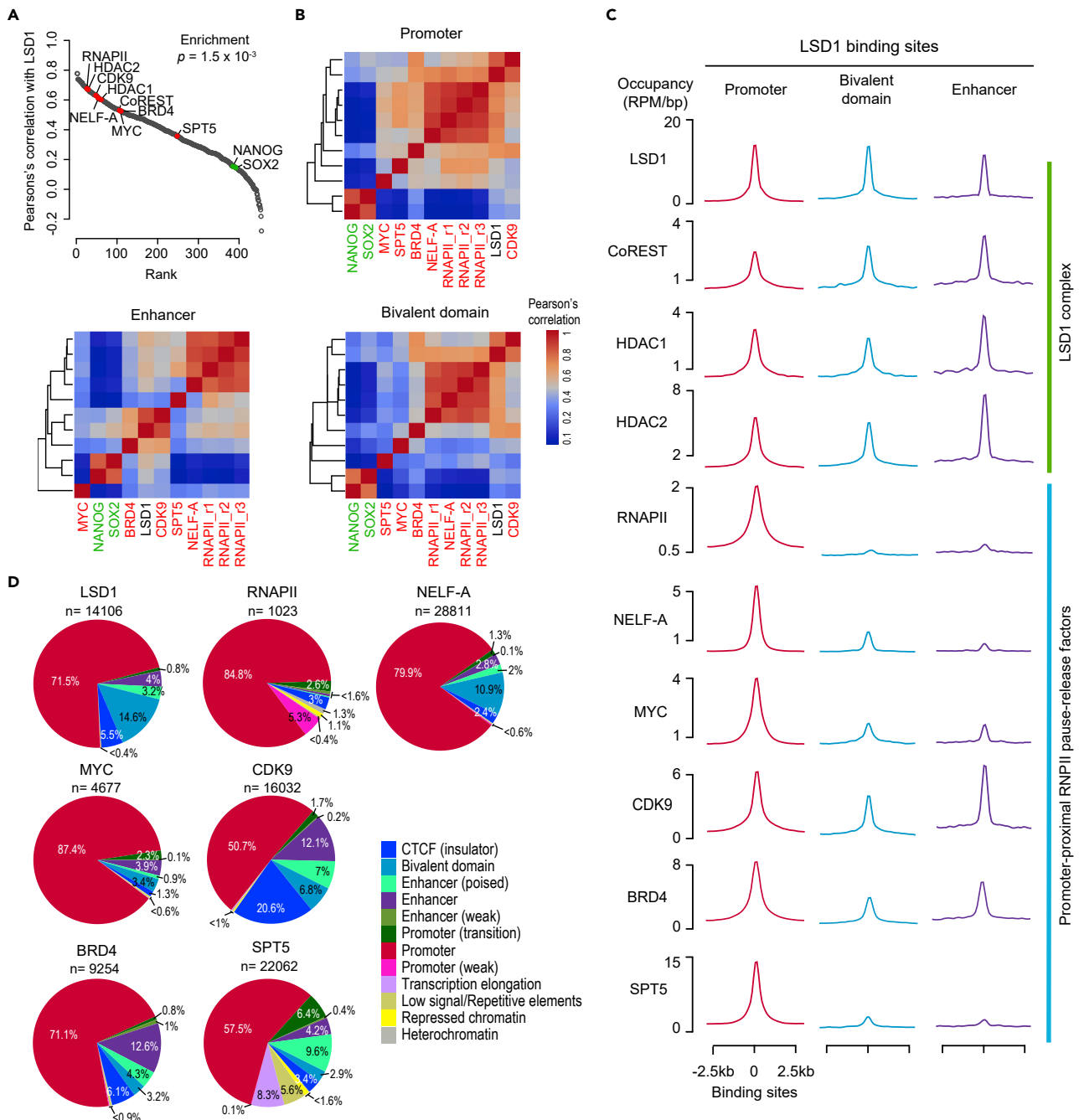
Figure 1. Co-localization of LSD1 with RNAPII pausing machinery across regulatory regions

(A) DNA binding proteins are ranked by the correlation of binding sites with LSD1 at promoter regions. A panel of genes known to be associated with RNAPII pausing are highlighted in red and two ESC pluripotency-associated TFs are highlighted in green for contrast. Gene set enrichment test was applied to the RNAPII-associated proteins (those in red) with respect to their correlation with LSD1.

(B) Co-localization heatmaps of LSD1 and select of RNAPII-associated factors across three regulatory regions (promoters, bivalent domains, and enhancers).

(C) Binding sites of LSD1, components of the LSD1 complex (CoREST, HDAC1, and HDAC2), and factors associated with the RNAPII pausing machinery (RNAPII, NELF-A, MYC, CDK9, BRD4, and SPT5) at the three regulatory regions.

(D) Pie charts showing the proportion of binding sites of factors across the 12 regulatory regions.
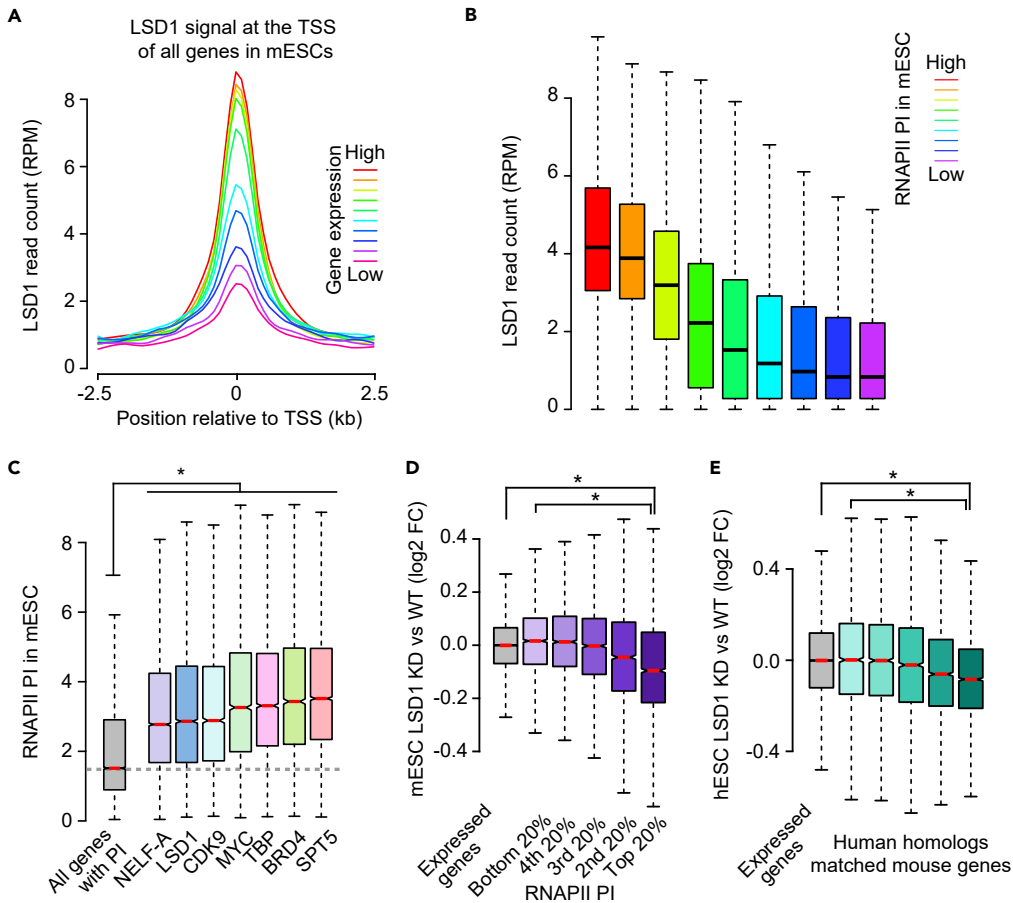
See also Figure S1.

Figure 2. Knockdown of LSD1 affects genes with higher RNAPII pausing than those with lower pausing (see also Figure S2)

(A) Gene sets were partitioned according to the level of expression in ESCs and the level of LSD1 signal (RPM) (+/− 1kb around the transcription start site [TSS]) was quantified for each gene set.

(B) Boxplot of LSD1 signal (RPM) at gene promoters grouped according to RNAPII pausing index (PI), calculated as the ratio of RNAPII binding across the gene body against RNAPII binding at the TSS (see STAR Methods).

(C) Boxplot of RNAPII PI of genes bound by NELF-A, LSD1, CDK9, MYC, TBP, BRD4, and SPT5.

(D and E) Boxplot of log2 fold-change in gene expression for gene sets grouped in terms of RNAPII PI after knockdown of LSD1 in (d) ESCs and (e) human ESCs (hESCs). * denotes statistical significance (p < 0.05) using the Wilcox rank sum test. For each boxplot, center line, median; box limits, upper and lower quartiles; whiskers, 1.5 times the interquartile range; points, outliers.

prevalence of LSD1 binding at promoters is striking particularly given that previous studies have largely focused on the role of LSD1 at enhancers (Whyte et al., 2012), even though less than 8% of LSD1 binding sites fall into the enhancer regions (as defined by ChromHMM). Altogether, these data show strong recruitment of LSD1 to gene promoters and its co-localization with factors associated with RNAPII pausing.

LSD1 regulates the transcription of genes associated with RNAPII pausing

Given the widespread presence of LSD1 at gene promoters, we first investigated the relationship between gene expression and LSD1 binding. We found that the higher LSD1 binding signals at gene promoters, the higher expression of the genes in ESCs (Figure 2A). Comparing gene expression of wildtype ESCs with those after LSD1 knockdown shows that genes with higher expression are more downregulated upon LSD1 knockdown, providing evidence for LSD1 binding at gene promoters for their transcription activation (Figure S2A). These results are consistent with experimental studies that have shown that LSD1 is a transcriptional activator (Amente et al., 2013; Sehrawat et al., 2018). We next sought to assess the relationship between RNAPII pausing and LSD1 binding of genes. To do this, we calculated the RNAPII pausing index (PI) of all genes based on RNAPII

ChIP-seq data (Muse et al., 2007) (see STAR Methods) and correlated them with the LSD1 ChIP-seq signal. Comparison with global run-on sequencing (GRO-seq) (Williams et al., 2015), which measures nascent RNA from transcriptionally engaged RNAPII, confirms a strong positive correlation between RNAPII ChIP-seq and GRO-seq signals (Figures S2B and 2SC). Our analysis suggests that LSD1 is preferentially bound at the promoters of paused genes (Figure 2B). Furthermore, we found that LSD1-bound genes show similar levels of RNAPII PI to those bound by other factors known to regulate RNAPII pausing or release in ESCs (Figure 2C). These results suggest that LSD1-bound genes are subject to RNAPII pausing.

We reasoned that if LSD1 has a functional role in controlling gene expression via regulation of RNAPII pausing, loss of LSD1 expression should preferentially affect the expression of LSD1 target gene sets with higher RNAPII PI much more so than those with lower RNAPII PI. To test this, we partitioned all quantified genes in ESCs into five groups and assessed their log2 fold change in expression before and after LSD1 knockdown (Foster et al., 2010). We found that the log2 fold change in gene expression after LSD1 knockdown is proportionately affected by RNAPII pausing, indicating that LSD1 knockdown preferentially affects genes with higher RNAPII PI (Figure 2D). Lastly, we confirmed our observations using data from human ESCs by showing the same pausing-dependent log2 fold change in expression with LSD1 knockdown can be observed in mouse and human ESCs (Adamo et al., 2011) (Figure 2E). Collectively, these results suggest a potential role for LSD1 in regulating gene expression through RNAPII pausing, and these findings motivated us to investigate the mechanistic role of LSD1 in modulating transcriptional programs associated with RNAPII pausing.

LSD1 knockdown causes global reduction of RNAPII pausing

Given its association with RNAPII pausing and the transcription of genes, we next investigated the direct impact of LSD1 knockdown on RNAPII pausing at gene promoters. By genome-wide ChIP-seq profiling of RNAPII with or without LSD1 knockdown (Figures 3A and 3B), we found that the depletion of LSD1 in ESC induced the reduction of RNAPII pausing at various LSD1-bound genes (Figures 3A and S3C). Meta-gene analysis revealed that the average ChIP-seq signal of RNAPII profiles across the gene bodies of all LSD1-bound genes demonstrated a clear reduction of RNAPII pausing (Figures 3B and S3D). Notably, we found that the knockdown of LSD1 led to a global reduction of RNAPII pausing (Figure 3C) and this extent of RNAPII pausing reduction was significantly greater in the LSD1-bound genes when compared to that of all the genes, even when stratified into low and high transcriptional groups (Figures 3C and S3E). Collectively, these results indicate that LSD1 promotes RNAPII pausing and the decrease of LSD1 expression causes a reduction of RNAPII pausing at gene promoters.

LSD1 and MYC co-occupancy is enriched for RNAPII pausing

MYC is known to play a major role in RNAPII release in mammalian cells (Rahl et al., 2010). It binds to promoter-proximal regions of transcribed genes to regulate the release of paused RNAPII by recruiting P-TEFb (Rahl et al., 2010). Recently, MYC has also been shown to recruit the elongation factor SPT5, a subunit of the elongation factor DSIF, to promoter-proximal RNAPII, wherein SPT5 travels with RNAPII and enhances its processivity during transcriptional elongation (Baluapuri et al., 2019). Given the important role of MYC in regulating the transcriptional activity of RNAPII and the strong co-localization between LSD1 and the pause-release factor P-TEFb, we asked whether comparing sites of LSD1 and MYC co-localization against sites where either LSD1 or MYC are found alone may provide additional insight into the role of LSD1 in RNAPII pause release.

To test this, we partitioned bindings sites of LSD1 and MYC into three categories: MYC-specific sites, co-localized sites, and LSD1-specific sites (Figure 4A). Approximately 57% of MYC binding sites were co-localized with LSD1 and 19% of LSD1 binding sites were co-localized with MYC. As a quality control, we assessed the LSD1 and MYC signals at MYC-specific and LSD1-specific sites and found that they are significantly depleted as per their definition (Figures 4A and S4A). In comparison, the co-localized sites had a comparatively similar amount of factor when compared to their respective individual sites.

Next, with the hypothesis that LSD1 co-occupancy would show us a differential binding profile of RNAPII pausing factors at MYC binding sites, we evaluated the difference in occupancy of RNAPII pausing factors (RNAPII, TBP, NELF-A, SPT5, BRD4, and CDK9) at these regions. Strikingly, we saw that compared to the LSD1/MYC-specific sites, sites co-occupied by LSD1 and MYC were significantly enriched by all RNAPII pausing factors. Interestingly, we observed that MYC-specific and co-localized sites are epigenetically more permissive with higher H3K27ac and H3K4me3 and lower H3K27me3 signals (Figures S4B and S4C)
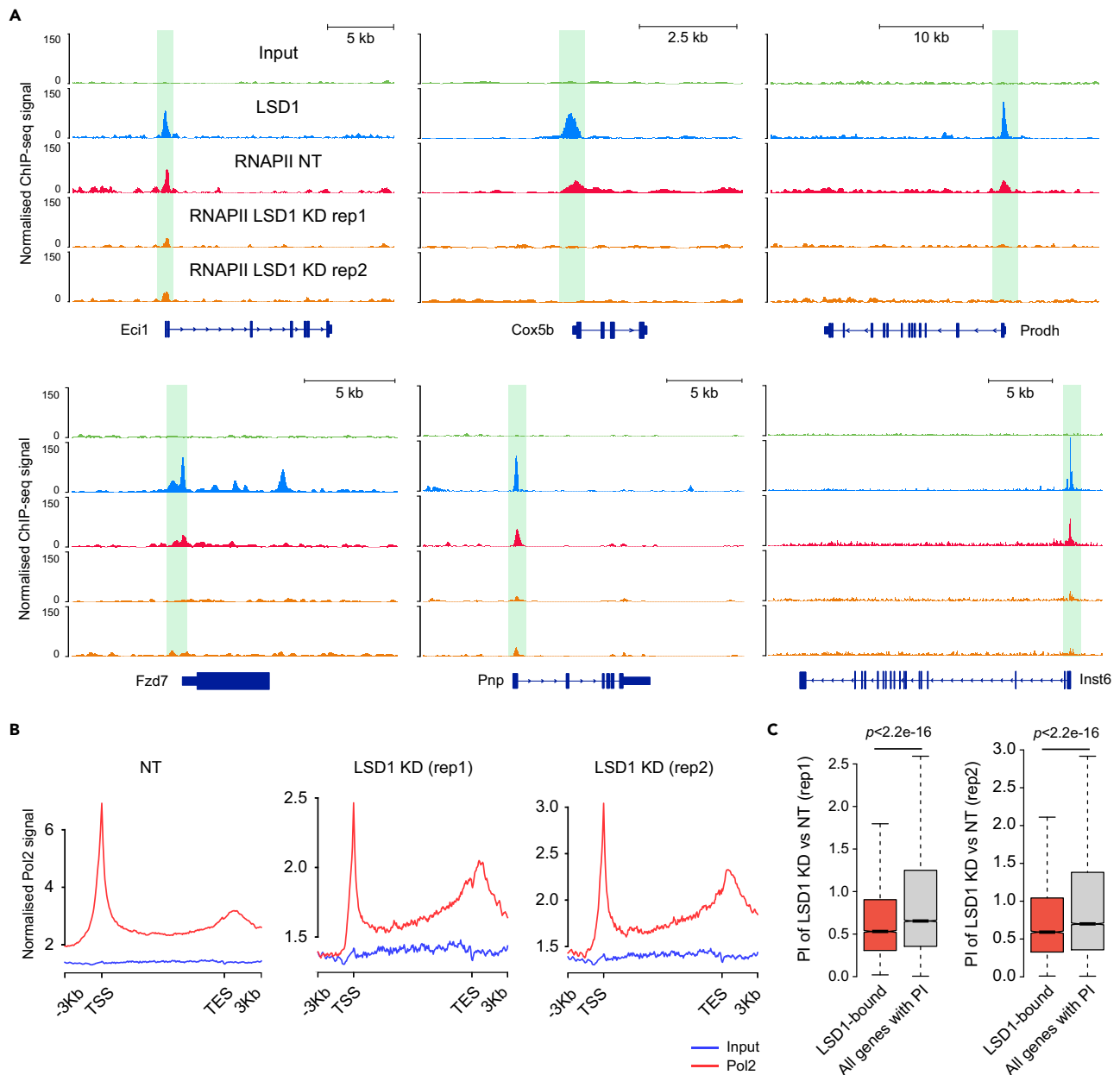
Figure 3. Impact of LSD1 knockdown on RNAPII pausing (see also Figure S3)

(A) ChIP-seq profiles of LSD1 and RNAPII with non-targeted (NT) or shLSD1 knockdown (KD) at six example genes. Promoter regions of genes are highlighted.

(B) Meta-gene body profile of input (blue) and RNAPII ChIP-seq (red) signal in NT and LSD1 KD samples.

(C) Boxplot showing the fold change (LSD1 KD versus NT) of RNAPII pausing index (PI) in LSD1-bound genes or all genes that have quantified PIs. p-values denote the statistical significance using the Wilcox rank sum test. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5 times the interquartile range; points, outliers.

than LSD1-specific sites, suggesting that their target genes may be more transcriptionally active. Collectively, these findings show that MYC and LSD1 co-localized sites are active promoters that are enriched for RNAPII pause/release factors.

To further characterize genes regulated by either LSD1 or MYC or both (Figure 4B), we identified the target genes bound by LSD1 and/or MYC and analyzed their expression profiles in ESCs. While all
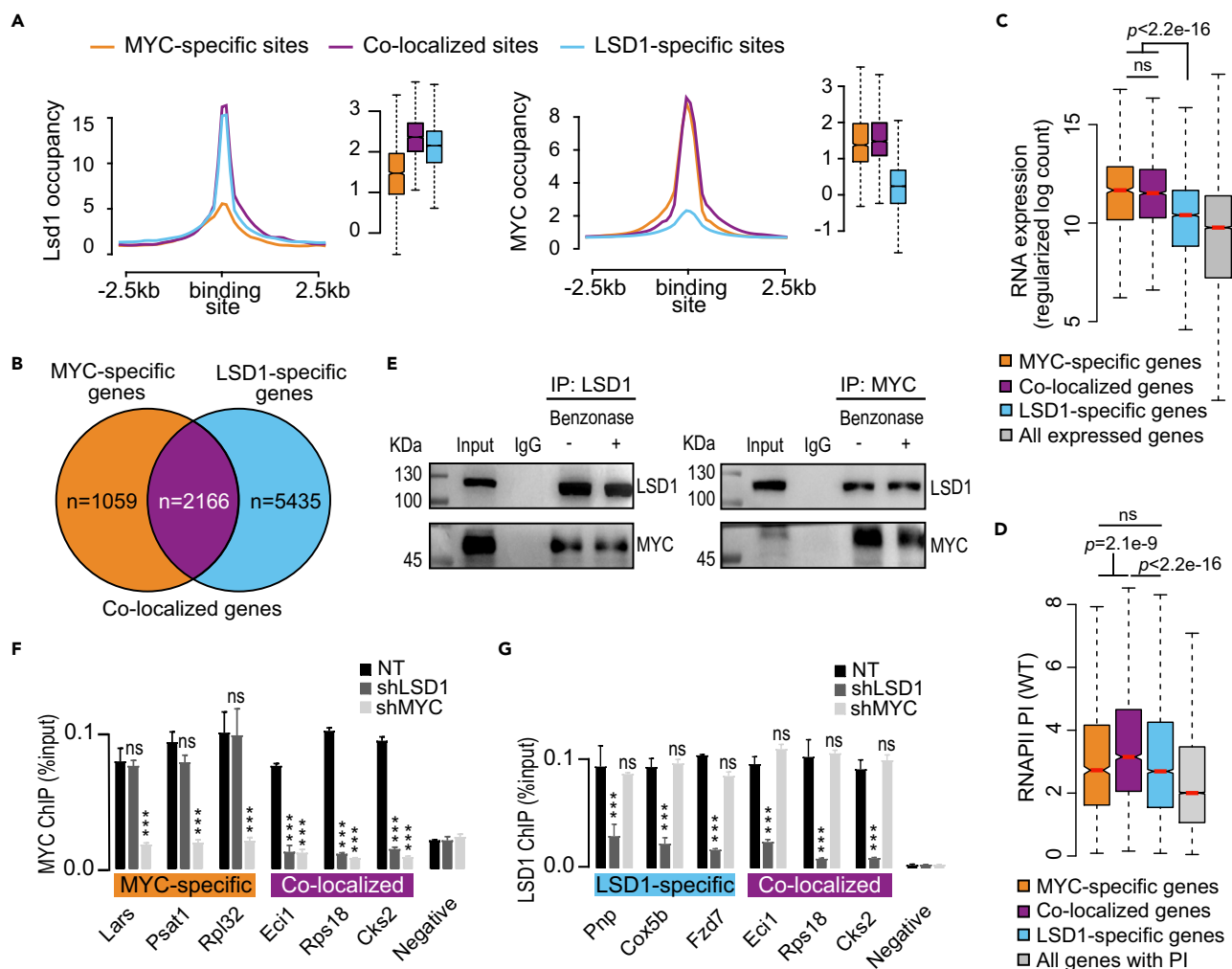
Figure 4. Partition of genome-wide binding sites into MYC-specific, co-localized, and LSD1-specific sites reveals enrichment of pausing (see also Figure S4)

(A) Density plots and boxplots of MYC and LSD1 occupancy (RPM) at the three genomic sites. The three lines denote the averaged signal across sites from each category: MYC-specific (orange), co-localized (purple), and LSD1-specific (blue) sites.

(B) Venn diagram showing the overlap of binding sites by the presence of either MYC and/or LSD1. Boxplots of regularised and log-transformed RNA expression and (d) RNAPII pausing in WT ESCs. Wilcox rank-sum test was used in (C and D). Center line, median; box limits, upper and lower quartiles; whiskers, 1.5 times the interquartile range; points, outliers.

(E) Co-immunoprecipitation experiment reveals a direct interaction between LSD1 and MYC with or without the presence of benzonase in mouse embryonic stem cells.

(F and G) Bar plot of ChIP-qPCR results showing MYC and LSD1 occupancy at selected genes in ESCs transduced with GFP (Non-targeted, NT), MYC (shMYC), or LSD1 (shLSD1) shRNA lentivirus. For ChIP-qPCRs analysis, ESCs were collected 96h after transduction. % Input values were plotted as mean ± standard error. Student's t-test; *p-value<0.05; **p-value<0.01; ***p-value<0.001, ns = no significant.

three gene sets have higher than average transcription in ESCs (Figure 4C), we found that genes targeted by MYC or both MYC and LSD1 were significantly more expressed than LSD1-specific target genes (Figure 4C), consistent with what we observed at the epigenetic level (Figures S4B and S4C). We next compared the RNAPII PI of the three sets of target genes in ESCs (Figure 4D). We found genes bound specifically by MYC, a factor well-documented for its role in RNAPII pause release (Baluapuri et al., 2019; Lu et al., 2015; De Pretis et al., 2017; Price, 2010; Rahl et al., 2010), had a similar level of RNAPII pausing compared to those bound specifically by LSD1 (Figure 4D), and consistent with our observation that the co-occupied sites are enriched for RNAPII pausing factors, we found that genes co-occupied by LSD1 and MYC exhibited significantly higher RNAPII pausing than those with LSD1 or MYC alone.

MYC binding at co-localized target genes is dependent on the presence of LSD1

To further strengthen the mechanistic insight of LSD1 in RNAPII pausing, we further investigated the interaction between MYC and LSD1. Our association studies suggested a functional link between LSD1 and MYC, as revealed by the significant enrichment of RNAPII-associated factors at target genes co-bound by LSD1 and MYC compared to the singly occupied target genes. First, we performed co-immunoprecipitation analysis to demonstrate the co-localization between LSD1 and MYC, which revealed that, indeed, they are directly interacting in ESCs and do so independently of the presence of DNA and RNA (Figure 4E). Next, we investigated whether the binding of MYC at its target sites is dependent on the presence of LSD1. Previously, MYC has been shown to be responsible for the recruitment of SPT5 (Baluapuri et al., 2019); therefore, the dependency of MYC recruitment on LSD1 has significant implications on RNAPII pausing and elongation. Leveraging on the target genes bound by LSD1 or MYC or both, we performed MYC and LSD1 ChIP qPCR at these sites under either LSD1 or MYC knockdown. Our findings revealed that the binding of MYC is dependent on the presence of LSD1 specifically at target genes that are co-regulated by these two factors, whereas MYC binding was not affected by the depletion of LSD1 in sites specifically targeted by MYC (Figure 4F). In contrast, depletion of MYC had no effect on co-regulated target genes, suggesting that the binding of LSD1 is independent of the presence of MYC (Figure 4G). Collectively, these findings support the co-operative role between LSD1 and MYC in regulating transcription.

Occupancy of LSD1 and MYC is associated with different transcription factors

As our analysis suggested differential regulation of genes targeted by both LSD1 and MYC versus those targeted by LSD1 or MYC alone, we performed *de novo* motif discovery from the DNA sequences of the three sets of binding sites (Figure S5A) and transcription factor (Figure S5B) enrichment from the three gene sets to further characterize their difference. We found that the top two enriched motifs differed between the three sets of binding sites. Specifically, we found that the sites co-occupied by LSD1 and MYC, but not those by MYC alone, were enriched with motifs of KLF family member SP1 (Kaczynski et al., 2003) (Figure S5A). This is in agreement with previous research that found that KLF family members are DNA-binding transcription factors specifically associated with RNAPII promoters with slow TBP (TATA-binding protein) turnover, which is associated with high transcriptional activity when compared to promoters with fast TBP turnover (Hasegawa and Struhl, 2019). At LSD1-specific binding sites, we found that the most enriched transcription factor in this gene set is CTCF, the binding of which at promoter-proximal regions has recently been shown to regulate distal enhancer-dependent gene transcription (Kubo et al., 2021). We note that the TF enrichment is consistent with our *de novo* motif analysis as we generated a motif resembling those of Lin28A in the LSD1-specific binding sites, which have been shown to be occupied near CTCF target sites (Tan and Yeo, 2016) (Figure S5). MYC and REST (a subunit of coREST complex) are highlighted as positive controls (Figure S5B). Collectively, these enrichment studies suggest different factors are involved in the regulation of the differentially occupied target genes.

Occupancy of LSD1 and MYC is associated with different gene functions

Lastly, we demonstrate that different gene sets are enriched in each category (Figure 5A). Whereas MYC-specific genes are enriched for pathways such as RNA splicing and miRNA processing, in agreement with previous knowledge (Kim et al., 2020), co-localized genes are largely enriched for metabolic processes, and LSD1-specific genes are enriched for pathways pertaining to cellular organization and localization (Figure 5A). Together, these results suggest that genes co-bound by LSD1 and MYC and those bound by them individually may undergo different transcriptional regulations and perform distinct biological functions.

Given the different functional enrichment of genes found by the above analysis, we next conducted a test for non-independent overlap of the three gene sets with transcription factors, which are frequently associated with enhancers and cell type-specificity (Eeckhoute et al., 2009), and housekeeping genes. Intriguingly, we found a reciprocal pattern of enrichment for transcription factors vs. housekeeping genes in the three gene sets. LSD1-specific target genes were enriched for target genes encoding transcription factors whilst MYC-specific and co-localized genes were not (Figures 5B and 5C). In contrast, co-localized and MYC-specific target genes were enriched for target genes encoding housekeeping genes whilst LSD1-specific genes were depleted, consistent with the gene set enrichment results showing that these sites were enriched for genes related to, for example, metabolic processes that take place regardless of the cell type (Figure 5A). Collectively, our findings reveal a putative role in pausing LSD1 at promoters through its association with MYC and P-TEFb that may be distinct from its enhancer-related role in ESC
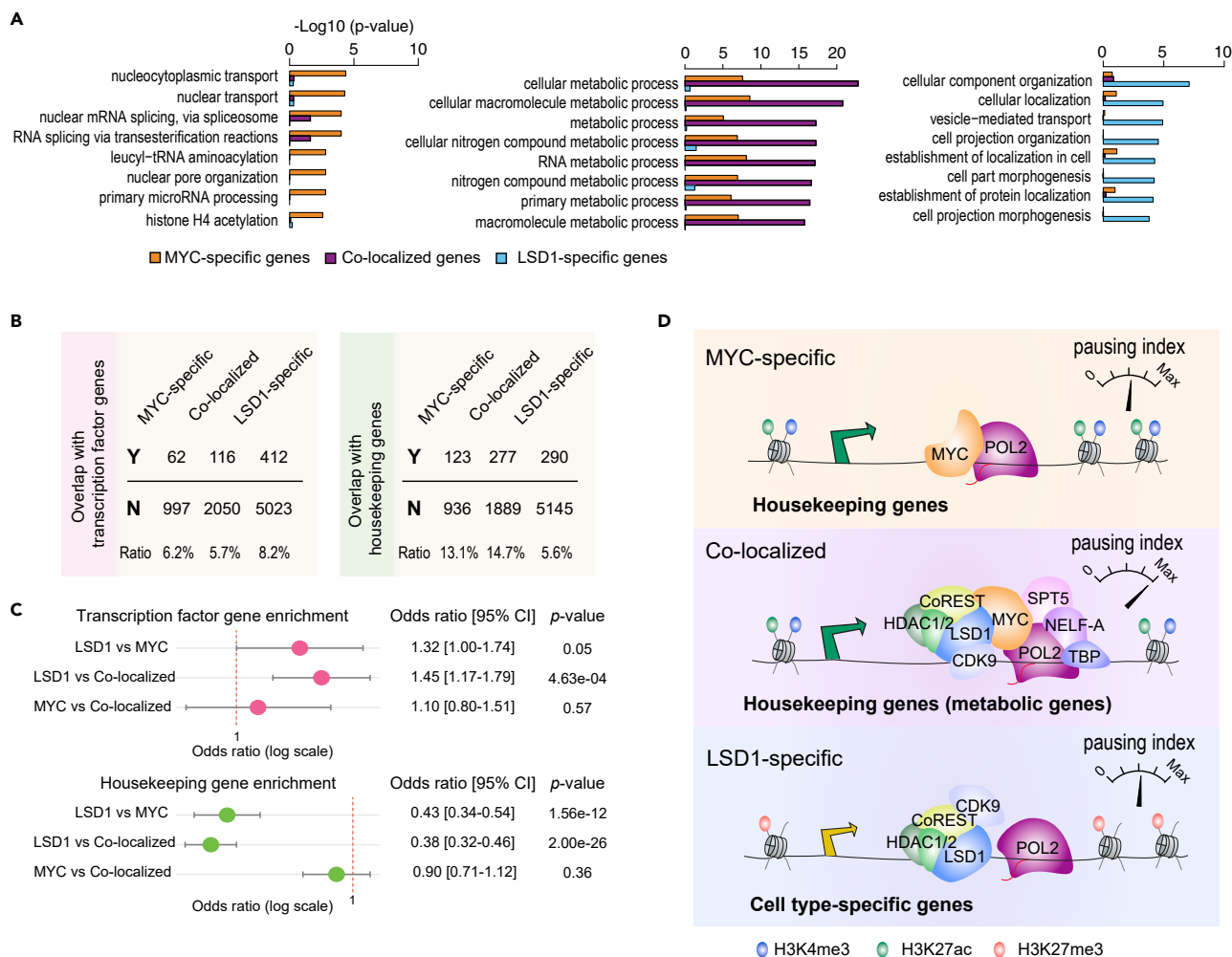
Figure 5. Differentially enrichment of transcription factors and housekeeping genes at LSD1 sites (see also Figure S5)

(A) Over-representation analysis of gene set enrichment from the three categories. X-axis denotes the degree of enrichment in terms of the negative log10 p-value.

(B) Overlap and ratios of MYC, LSD1, and their co-localization sites with transcription factors (left) and housekeeping genes (right). (a) Test of non-independent overlap of (C) using Fisher's exact test.

(D) Schematic of the differential regulation of LSD1 target genes. Sites occupied by LSD1 only and those co-occupied with RNAPII pausing machinery show differential transcriptional control, epigenetic landscape, and enrichment of transcription factors and housekeeping genes in ESCs.

differentiation (Whyte et al., 2012) and demonstrate differential transcriptional control, epigenetic land-scape, and enrichment of transcription factors and housekeeping genes between LSD1 sites distinguished by RNAPII pausing (Figure 5D).

DISCUSSION

Here we show how the association of LSD1 at the promoters of actively transcribed genes corresponds to genes engaged in RNAPII pausing in ESCs and demonstrate a functional role of LSD1 in RNAPII pausing. LSD1 has been shown to decommission enhancers during ESC differentiation (Vinckier et al., 2020; Whyte et al., 2012), but its role in the promoters of genes remains elusive. Although the regulation of LSD1 at gene promoters can be attributed to both its binding at distal regulatory elements such as enhancers and promoter-proximal regions, we found that the majority of the LSD1 binding sites are located at promoter-proximal regions (71.5%, Figure 2D). Several lines of evidence suggest the involvement of LSD1 in RNAPII pausing. We observed that the LSD1 signal was strongly and positively correlated with gene expression, supporting a transactivating role of LSD1 in gene transcription. This is consistent with

a recent study that demonstrates that LSD1 is recruited to promoters by MYCB to promote active transcription (Ahmed and Streit, 2018). Previously, LSD2 the only mammalian homolog of LSD1 has also been shown to have an important function in active gene transcription elongation (Fang et al., 2010). Moreover, our genome-wide chromatin-binding profiles demonstrated that LSD1 binding sites at RNAPII paused sites are strongly co-localized with RNAPII pause-release factors (Figure 1C). A recent study has demonstrated a functional role of SIRT6, a histone deacetylase, in the release of RNAPII pausing by preventing the eviction of NELF-E from the chromatin (Etchegaray et al., 2019), highlighting the direct involvement of chromatin remodellers in RNAPII pausing and release.

Previous studies have shown that LSD1 physically interacts with RNAPII (Latos et al., 2015), MYC (Baluapuri et al., 2019; Weimann et al., 2013), and SNAIL (Carmichael et al., 2020), which in Drosophila embryos has been shown to inhibit the release of paused RNAPII (Bothma et al., 2011). Among 22 methyltransferases, LSD1 was the only one found to be in direct physical contact with MYC in the methyltransferase interactome (Baluapuri et al., 2019; Weimann et al., 2013). We confirmed LSD1 as a physical interactor of MYC in mouse ESCs through co-immunoprecipitation analysis (Figure 4E). Our findings further demonstrate a causal link between LSD1 and MYC function by investigating the LSD1 dependency of MYC genomic binding. Our results revealed that the recruitment of MYC is dependent on the presence of LSD1 specifically at target genes that are co-regulated by these two factors but not vice versa. Collectively, these findings strongly support the co-operative role between LSD1 and MYC in regulating transcription and highlight at least two distinct streams of MYC-regulated transcription that are either LSD1 dependent or independent.

A recent study demonstrated a generic role of MYC in governing the transition of RNAPII from a paused state into a transcriptionally engaged mode by directly recruiting SPT5 to RNAPII (Baluapuri et al., 2019). Consistent with the function of MYC as a universal amplifier of gene expression (Nie et al., 2012), the ablation of MYC led to global depletion of SPT5 in chromatin (Baluapuri et al., 2019). Strikingly, we found that among the MYC binding sites, the presence of LSD1 facilitated the enrichment of RNAPII pausing factors at the promoters of genes and a higher pausing index (Figure 4). In comparison, genes bound by MYC demonstrated lower RNAPII pausing index but still retained a transcriptional level comparable to genes bound by both factors. We note that whilst we defined sites or genes to be either bound or unbound in this study, the LSD1 signal is continuous (Figure S4A) and so are the other RNAPII pausing-associated factors (Figure S4B), and hence the effect of LSD1 is in general global as was seen in Figure 3C and was also demonstrated by the changes of the gene expression in Figure 2D.

Consistent with these findings, it has previously been shown that genetic ablation and chemical inhibition of LSD1 in pluripotent cell lines (mouse ESCs and F9 cells) increased the level of H3K56ac (Yin et al., 2014), a histone modification associated with the gene body (Jonkers and Lis, 2015) and recently identified as a key histone mark, along with H3K9ac, that modulates transcriptional pausing and elongation (Etchegaray et al., 2019). In another study, LSD1 ablation and catalytic inhibition led to an upregulation of H4K16ac (Yin et al., 2014), a histone mark that has been implicated in the regulation of RNAPII promoter-proximal pausing by recruiting BRD4 and P-TEFb (Kapoor-Vazirani et al., 2011; Zippo et al., 2009). Thus, our integrative analysis suggests a role for LSD1 in the regulation of RNAPII pause release through its direct association with critical pause release factors including MYC. We note that these secondary changes brought about by LSD1 knockdown in the level of histone modifications that have been shown to impact gene transcription may partially contribute toward the global decrease in RNPII occupancy observed in our study with LSD1 knockdown (Figure 3C). In addition, our genome-wide binding profiles reveal a strong co-localization of LSD1 and CDK9 (Figure 1C), the enzymatic subunit of P-TEFb, whose activity has been proposed as the rate-limiting step in paused RNAPII release (Gressel et al., 2017; Larochelle et al., 2012; Parua et al., 2018). The co-localization of LSD1 and CDK9 was observed at both promoter and enhancer elements, where CDK9 has been shown to be involved in enhancer-associated RNAPII pausing (Bacon and D'Orso, 2019; Ghavi-helm et al., 2014; Henriques et al., 2018). Future studies on LSD1 and CDK9 may shed light on their relationship in regulating RNAPII pausing.

In ESCs, the catalytic activity of LSD1 has been shown to render enhancers inactive and is associated with the silencing of target genes (Whyte et al., 2012). Studies have shown LSD1 occupies the regulatory domains of pluripotency genes, including transcription factors, and decommissions them only when cells undergo lineage-specific differentiation (Vinckier et al., 2020; Whyte et al., 2012). Although we found that LSD1-specific genes are enriched for transcription factors, supporting its role in regulating cell

type-specific differentiation, genes co-localized by LSD1 and MYC are significantly depleted of transcription factors and rather enriched for housekeeping genes, and in particular those that are involved in the regulation of cell type-invariant metabolic processes. Consistent with this, loss of LSD1 function has been shown to be associated with significant proliferative defects in neural (Sun et al., 2010) and embryonic stem cells (Yin et al., 2014) and has been implicated in cellular growth pathways and in the metastatic and oncogenic potential of several types of cancer (Amente et al., 2013; Wang et al., 2007, 2009).

Our integrative analysis presents evidence for a previously unanticipated role of the histone demethylase LSD1 in cooperation with other RNAPII pausing factors in modulating cell type-specific and cell type-invariant genes and forms the basis for further investigation. Future work to assess the various phosphorylated forms of the carboxy-terminal tail of RNAP2 upon LSD1 knockdown and to evaluate the role of LSD1 activity toward its binding with MYC and its influence on RNAP2 pausing will be required to further elucidate the mechanism of LSD1 action in RNAP2 pausing.

Limitations of the study

Whilst our study highlights the role of LSD1 in RNAPII pausing, several limitations of the study remain to be addressed that will further shed light on its mechanism of action. For example, experiments probing the role of the LSD1 histone demethylase activity and its relationship with RNAPII pausing, the direct interaction of LSD1 with RNAPII pausing related factors, and the changes in the levels of key indicators of RNAPII pausing and releasing (such as RNAPII CTD Ser2 and Ser5 phosphorylation) should further reveal important mechanistic insights. We believe these are important future research avenues that will elucidate the exact mechanism of action of LSD1 in RNAPII pausing.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Lentiviral shRNA constructs, lentivirus preparation and transduction
  - RT-qPCR
  - Co-immunoprecipitation
  - Western blot analysis
  - ChIP-qPCR and ChIP-seq
  - PAD implementation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis and visualization of ChIP-seq data
  - Binding sites and target genes
  - Calculation of RNAPII pausing index
  - Gene expression analysis
  - Motif and transcription factor enrichment analysis
  - Functional enrichment analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105049.

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS

The authors declare no competing interests.

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. https://doi.org/10.1038/ncb2246.

Ahmed, M., and Streit, A. (2018). Lsd1 interacts with cMyb to demethylate repressive histone marks and maintain inner ear progenitor identity. Development *145*, dev160325. https://doi.org/10.1242/dev.160325.

AlAbdi, L., Saha, D., He, M., Dar, M.S., Utturkar, S.M., Sudyanti, P.A., McCune, S., Spears, B.H., Breedlove, J.A., Lanman, N.A., and Gowher, H. (2020). Oct4-Mediated inhibition of Lsd1 activity promotes the active and primed state of pluripotency enhancers. Cell Rep. *30*, 1478–1490.e6. https://doi.org/10.1016/j.celrep.2019.11.040.

Amente, S., Lania, L., and Majello, B. (2013). The histone LSD1 demethylase in stemness and cancer transcription programs. Biochim. Biophys. Acta *1829*, 981–986. https://doi.org/10.1016/j.bbagrm.2013.05.002.

Bacon, C.W., and D'Orso, I. (2019). CDK9: a signaling hub for transcriptional control. Transcription *10*, 57–75. https://doi.org/10.1080/21541264.2018.1523668.

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. *34*, W369–W373. https://doi.org/10.1093/nar/gkl198.

Baluapuri, A., Hofstetter, J., Dudvarski Stankovic, N., Endres, T., Bhandare, P., Vos, S.M., Adhikari, B., Schwarz, J.D., Narain, A., Vogt, M., et al. (2019). MYC recruits SPT5 to RNA polymerase II to promote processive transcription elongation. Mol. Cell *74*, 674–687.e11. https://doi.org/10.1016/j.molcel.2019.02.031.

Gene Ontology Consortium, Christie, K.R., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Burgess, S., Buza, T., Gresham, C., et al. (2015). Gene ontology consortium: going

forward. Nucleic Acids Res. *43*, D1049–D1056. https://doi.org/10.1093/nar/gku1179.

Booth, G.T., Parua, P.K., Sansó, M., Fisher, R.P., and Lis, J.T. (2018). Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. Nat. Commun. *9*, 543. https://doi.org/10.1038/s41467-018-03006-4.

Bothma, J.P., Magliocco, J., and Levine, M. (2011). The snail repressor inhibits release, not elongation, of paused Pol II in the drosophila embryo. Curr. Biol. *21*, 1571–1577. https://doi.org/10.1016/j.cub.2011.08.019.

Carmichael, C.L., Wang, J., Nguyen, T., Kolawole, O., Benyoucef, A., De Mazière, C., Milne, A., Samuel, S., Gillinder, K.R., Hediyeh-zadeh, S., et al. (2020). The EMT modulator SNAI1 contributes to AML pathogenesis via its interaction with LSD1. Blood *136*, 957–973. https://doi.org/10.1182/blood.2019002548.

Chen, F.X., Smith, E.R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *19*, 464–478. https://doi.org/10.1038/s41580-018-0010-5.

Cheng, B., and Price, D.H. (2007). Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J. Biol. Chem. *282*, 21901–21912. https://doi.org/10.1074/jbc.M702936200.

Chen, X., Xu, H., Yuan, P., Fang, F., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Cheng, B., Li, T., Rahl, P.B., Adamson, T.E., Loudas, N.B., Guo, J., Varzavand, K., Cooper, J.J., Hu, X., Gnatt, A., et al. (2012). Functional association of gdown1 with RNA polymerase II poised on human genes. Mol. Cell *45*, 38–50. https://doi.org/10.1016/j.molcel.2011.10.022.

Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. *33*, 960–982. https://doi.org/10.1101/gad.325142.119.

Day, D.S., Zhang, B., Stevens, S.M., Ferrari, F., Larschan, E.N., Park, P.J., and Pu, W.T. (2016). Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types. Genome Biol. *17*, 120. https://doi.org/10.1186/s13059-016-0984-2.

De Pretis, S., Kress, T.R., Morelli, M.J., Sabò, A., Locarno, C., Verrecchia, A., Doni, M., Campaner, S., Amati, B., and Pelizzola, M. (2017). Integrative analysis of RNA polymerase II and transcriptional dynamics upon MYC activation. Genome Res. *27*, 1658–1664. https://doi.org/10.1101/gr.226035.117.

Devaiah, B.N., Lewis, B.A., Cherman, N., Hewitt, M.C., Albrecht, B.K., Robey, P.G., Ozato, K., Sims, R.J., and Singer, D.S. (2012). BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA Polymerase II carboxy-terminal domain. Proc. Natl. Acad. Sci. USA *109*, 6927–6932. https://doi.org/10.1073/pnas.1120422109.

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. https://doi.org/10.1093/bioinformatics/bts635.

Eeckhoute, J., Métivier, R., and Salbert, G. (2009). Defining specificity of transcription factor regulatory activities. J. Cell Sci. *122*, 4027–4034. https://doi.org/10.1242/jcs.054916.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. *12*, 2478–2492. https://doi.org/10.1038/nprot.2017.124.

Etchegaray, J.P., Zhong, L., Li, C., Henriques, T., Ablondi, E., Nakadai, T., Van Rechem, C., Ferrer, C., Ross, K.N., Choi, J.E., et al. (2019). The histone deacetylase SIRT6 restrains transcription elongation via promoter-proximal pausing. Mol. Cell *75*, 683–699.e7. https://doi.org/10.1016/j.molcel.2019.06.034.

Fang, R., Barbera, A.J., Xu, Y., Rutenberg, M., Leonor, T., Bi, Q., Lan, F., Mei, P., Yuan, G.C., Lian, C., et al. (2010). Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 Methylation. Mol. Cell *39*,

222–233. https://doi.org/10.1016/j.molcel.2010.07.008.

Foster, C.T., Dovey, O.M., Lezina, L., Luo, J.L., Gant, T.W., Barlev, N., Bradley, A., and Cowley, S.M. (2010). Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. Mol. Cell Biol. 30, 4851–4863. https://doi.org/10.1128/mcb.00521-10.

Gaertner, B., and Zeitlinger, J. (2014). RNA polymerase II pausing during development. Development 141, 1179–1183. https://doi.org/10.1242/dev.088492.

Gamot, A., Martin, P., Malirat, N., Koval, S., Stadelmayer, B., Parrinello, H., Cuvier, O., Raffel, R., Sobhian, B., and Severac, D. (2014). Polymerase II Pause/Release and Processivity at Coding Genes. https://doi.org/10.1038/ncomms6531.

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Article integrator regulates transcriptional initiation and pause release following activation. Mol. Cell 56, 128–139. https://doi.org/10.1016/j.molcel.2014.08.004.

Gatchalian, J., Malik, S., Ho, J., Lee, D.-S., Kelso, T.W.R., Shokhirev, M.N., Dixon, J.R., and Hargreaves, D.C. (2018). A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. Nat. Commun. 9, 5139. https://doi.org/10.1038/s41467-018-07528-9.

Ghavi-helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512, 96–100. https://doi.org/10.1038/nature13417.

Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. Elife 6, e29736. https://doi.org/10.7554/eLife.29736.

Guo, S., Yamaguchi, Y., Schilbach, S., Wada, T., Lee, J., Goddard, A., French, D., Handa, H., and Rosenthal, A. (2000). A regulator of transcriptional elongation controls vertebrate neuronal development. Nature 408, 366–369. https://doi.org/10.1038/35042590.

Hasegawa, Y., and Struhl, K. (2019). Promoter-specific dynamics of TATA-binding protein association with the human genome. Genome Res. 29, 1939–1950. https://doi.org/10.1101/gr.254466.119.

Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2013). Stable pausing by rna polymerase II provides an opportunity to target and integrate regulatory signals. Mol. Cell 52, 517–528. https://doi.org/10.1016/j.molcel.2013.10.001.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev. 32, 26–41. https://doi.org/10.1101/gad.309351.117.

Itzen, F., Greifenberg, A.K., Bösken, C.A., and Geyer, M. (2014). Brd4 activates P-TEFb for RNA polymerase II CTD phosphorylation. Nucleic Acids Res. 42, 7577–7590. https://doi.org/10.1093/nar/gku449.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 16, 167–177. https://doi.org/10.1038/nrm3953.

Jothi, R., Cuddapah, S., Barski, A., Cui, K., and Zhao, K. (2008). Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res. 36, 5221–5231. https://doi.org/10.1093/nar/gkn488.

Kaczynski, J., Cook, T., and Urrutia, R. (2003). Sp1- and Krüppel-like transcription factors. Genome Biol. 4, 206. https://doi.org/10.1186/gb-2003-4-2-206.

Kapoor-Vazirani, P., Kagey, J.D., and Vertino, P.M. (2011). SUV420H2-Mediated H4K20 trimethylation enforces RNA polymerase II promoter-proximal pausing by blocking hMOF-dependent H4K16 acetylation. Mol. Cell Biol. 31, 1594–1609. https://doi.org/10.1128/mcb.00524-10.

Keenan, A.B., Torre, D., Lachmann, A., Leong, A.K., Wojciechowicz, M.L., Utti, V., Jagodnik, K.M., Kropiwnicki, E., Wang, Z., and Ma'ayan, A. (2019). ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res. 47, W212–W224. https://doi.org/10.1093/nar/gkz446.

Kim, H.J., Osteil, P., Humphrey, S.J., Cinghu, S., Oldfield, A.J., Patrick, E., Wilkie, E.E., Peng, G., Suo, S., Jothi, R., et al. (2020). Transcriptional network dynamics during the progression of pluripotency revealed by integrative statistical learning. Nucleic Acids Res. 48, 1828–1842. https://doi.org/10.1093/nar/gkz1179.

Kubo, N., Ishii, H., Xiong, X., Bianco, S., Meitinger, F., Hu, R., Hocker, J.D., Conte, M., Gorkin, D., Yu, M., et al. (2021). Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. Nat. Struct. Mol. Biol. 28, 152–161. https://doi.org/10.1038/s41594-020-00539-5.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Larochelle, S., Amat, R., Glover-Cutter, K., Sansó, M., Zhang, C., Allen, J.J., Shokat, K.M., Bentley, D.L., and Fisher, R.P. (2012). Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. Nat. Struct. Mol. Biol. 19, 1108–1115. https://doi.org/10.1038/nsmb.2399.

Latos, P.A., Goncalves, A., Oxley, D., Mohammed, H., Turro, E., and Hemberger, M. (2015). Fgf and Esrrb integrate epigenetic and transcriptional networks that regulate self-renewal of trophoblast stem cells. Nat. Commun. 6, 7776. https://doi.org/10.1038/ncomms8776.

Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. Cell 145, 502–511. https://doi.org/10.1016/j.cell.2011.04.021.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, P., Wang, L., Bennett, B.D., Wang, J., Li, J., Qin, Y., Takaku, M., Wade, P.A., Wong, J., and Hu, G. (2017). Rif1 promotes a repressive chromatin state to safeguard against endogenous retrovirus activation. Nucleic Acids Res. 45, 12723–12738. https://doi.org/10.1093/nar/gkx884.

Li, Y., Liu, M., Chen, L.F., and Chen, R. (2018). P-TEFb: finding its ways to release promoter-proximally paused RNA polymerase II. Transcription 9, 88–94. https://doi.org/10.1080/21541264.2017.1281864.

Liu, X., Kraus, W.L., and Bai, X. (2015). Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. Trends Biochem. Sci. 40, 516–525. https://doi.org/10.1016/j.tibs.2015.07.003.

Lu, H., Xue, Y., Xue, Y., Yu, G.K., Arias, C., Lin, J., Fong, S., Faure, M., Weisburd, B., Ji, X., et al. (2015). Compensatory induction of MYC expression by sustained CDK9 inhibition via a BRD4-dependent mechanism. Elife 4. e06535–26. https://doi.org/10.7554/eLife.06535.

Lu, X., Zhu, X., Li, Y., Liu, M., Yu, B., Wang, Y., Rao, M., Yang, H., Zhou, K., Wang, Y., et al. (2016). Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. Nucleic Acids Res. 44, 6853–6867. https://doi.org/10.1093/nar/gkw571.

Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J. Biol. Chem. 270, 12335–12338. https://doi.org/10.1074/jbc.270.21.12335.

Meng, H., and Bartholomew, B. (2018). Emerging roles of transcriptional enhancers in chromatin looping and promoter-proximal pausing of RNA polymerase II. J. Biol. Chem. 293, 13786–13794. https://doi.org/10.1074/jbc.R117.813485.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev. 25, 742–754. https://doi.org/10.1101/gad.2005511.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. 39, 1507–1511. https://doi.org/10.1038/ng.2007.21.

Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., et al. (2012). c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. Cell 151, 68–79. https://doi.org/10.1016/j.cell.2012.08.033.

Oldfield, A.J., Yang, P., Conway, A.E., Cinghu, S., Freudenberg, J.M., Yellaboina, S., and Jothi, R. (2014). Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. Mol. Cell 55, 708–722. https://doi.org/10.1016/j.molcel.2014.07.005.

Parua, P.K., Booth, G.T., Sansó, M., Benjamin, B., Tanny, J.C., Lis, J.T., and Fisher, R.P. (2018). A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. Nature *558*, 460–464. https://doi.org/10.1038/s41586-018-0214-z.

Patel, M.C., Debrosse, M., Smith, M., Dey, A., Huynh, W., Sarai, N., Heightman, T.D., Tamura, T., and Ozato, K. (2013). BRD4 coordinates recruitment of pause release factor P-TEFb and the pausing complex NELF/DSIF to regulate transcription elongation of interferon-stimulated genes. Mol. Cell Biol. *33*, 2497–2507. https://doi.org/10.1128/mcb.01180-12.

Petell, C.J., Alabdi, L., He, M., San Miguel, P., Rose, R., and Gowher, H. (2016). An epigenetic switch regulates de novo DNA methylation at a subset of pluripotency gene enhancers during embryonic stem cell differentiation. Nucleic Acids Res. *44*, 7605–7617. https://doi.org/10.1093/nar/gkw426.

Price, D.H. (2010). Regulation of RNA polymerase II elongation by c-Myc. Cell *141*, 399–400. https://doi.org/10.1016/j.cell.2010.04.016.

Quinlan, A.R. (2014). BEDTools: the Swiss-Army tool for genome feature analysis. Curr. Protoc. Bioinf. *47*, 11.12.1–11.1234. https://doi.org/10.1002/0471250953.bi1112s47.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). C-Myc regulates transcriptional pause release. Cell *141*, 432–445. https://doi.org/10.1016/j.cell.2010.03.030.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). DeepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191. https://doi.org/10.1093/nar/gku365.

Sehrawat, A., Gao, L., Wang, Y., Bankhead, A., McWeeney, S.K., King, C.J., Schwartzman, J., Urrutia, J., Bisson, W.H., Coleman, D.J., et al. (2018). LSD1 activates a lethal prostate cancer gene network independently of its demethylase function. Proc. Natl. Acad. Sci. USA *115*, E4179–E4188. https://doi.org/10.1073/pnas.1719168115.

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. Mol. Cell *74*, 1148–1163.e7. https://doi.org/10.1016/j.molcel.2019.03.025.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953. https://doi.org/10.1016/j.cell.2004.12.012.

Sun, G., Alzayady, K., Stewart, R., Ye, P., Yang, S., Li, W., and Shi, Y. (2010). Histone demethylase LSD1 regulates neural stem cell proliferation. Mol. Cell Biol. *30*, 1997–2005. https://doi.org/10.1128/mcb.01116-09.

Tan, F.E., and Yeo, G.W. (2016). Blurred boundaries: the RNA binding protein Lin28A is also an epigenetic regulator. Mol. Cell *61*, 1–2. https://doi.org/10.1016/j.molcel.2015.12.018.

Tatsumi, G., Kawahara, M., Yamamoto, R., Hishizawa, M., Kito, K., Suzuki, T., Takaori-Kondo, A., and Andoh, A. (2020). LSD1-mediated repression of GFI1 super-enhancer plays an essential role in erythroleukemia. Leukemia *34*, 746–758. https://doi.org/10.1038/s41375-019-0614-6.

Vinckier, N.K., Patel, N.A., Geusz, R.J., Wang, A., Wang, J., Matta, I., Harrington, A.R., Wortham, M., Wetton, N., Wang, J., et al. (2020). LSD1-mediated enhancer silencing attenuates retinoic acid signalling during pancreatic endocrine cell development. Nat. Commun. *11*, 2082. https://doi.org/10.1038/s41467-020-16017-x.

Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018b). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. Nature *560*, 607–612. https://doi.org/10.1038/s41586-018-0440-4.

Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018a). Structure of paused transcription complex Pol II–DSIF–NELF. Nature *560*, 601–606. https://doi.org/10.1038/s41586-018-0442-2.

Wang, G.G., Allis, C.D., and Chi, P. (2007). Chromatin remodeling and cancer, part I: covalent histone modifications. Trends Mol. Med. *13*, 363–372. https://doi.org/10.1016/j.molmed.2007.07.003.

Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L., et al. (2009). LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. Cell *138*, 660–672. https://doi.org/10.1016/j.cell.2009.05.050.

Weimann, M., Grossmann, A., Woodsmith, J., Özkan, Z., Birth, P., Meierhofer, D., Benlasfer, N., Valovka, T., Timmermann, B., Wanker, E.E., et al. (2013). A Y2H-seq approach defines the human protein methyltransferase interactome. Nat. Methods *10*, 339–342. https://doi.org/10.1038/nmeth.2397.

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. https://doi.org/10.1038/nature10805.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., et al. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo,

D.C., Hu, G., and Adelman, K. (2015). Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol. Cell *58*, 311–322. https://doi.org/10.1016/j.molcel.2015.02.003.

Wu, C.H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., and Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes Dev. *17*, 1402–1414. https://doi.org/10.1101/gad.1091403.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol. Cell *21*, 227–237. https://doi.org/10.1016/j.molcel.2005.11.024.

Yamaguchi, Y., Shibata, H., and Handa, H. (2013). Biochimica et Biophysica Acta Transcription elongation factors DSIF and NELF : promoter-proximal pausing and beyond. Biochim. Biophys. Acta *1829*, 98–104. https://doi.org/10.1016/j.bbagrm.2012.11.007.

Yang, P., Humphrey, S.J., Cinghu, S., Pathania, R., Oldfield, A.J., Kumar, D., Perera, D., Yang, J.Y.H., James, D.E., Mann, M., and Jothi, R. (2019). Multi-omic profiling reveals dynamics of the phased progression of pluripotency. Cell Syst. *8*, 427–445.e10. https://doi.org/10.1016/j.cels.2019.03.012.

Yang, P., Oldfield, A., Kim, T., Yang, A., Yang, J.Y.H., and Ho, J.W.K. (2017). Integrative analysis identifies co-dependent gene expression regulation of BRG1 and CHD7 at distal regulatory sites in embryonic stem cells. Bioinformatics *33*, 1916–1920. https://doi.org/10.1093/bioinformatics/btx092.

Yin, F., Lan, R., Zhang, X., Zhu, L., Chen, F., Xu, Z., Liu, Y., Ye, T., Sun, H., Lu, F., and Zhang, H. (2014). LSD1 regulates pluripotency of embryonic stem/carcinoma cells through histone deacetylase 1-mediated deacetylation of histone H4 at lysine 16. Mol. Cell Biol. *34*, 158–179. https://doi.org/10.1128/mcb.00631-13.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. *39*, 1512–1516. https://doi.org/10.1038/ng.2007.26.

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. Cell *138*, 1122–1136. https://doi.org/10.1016/j.cell.2009.07.031.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-RNAP2	Santa Cruz	Cat#sc-47701; RRID:AB_677353
Anti-cMYC	Abcam	Cat#ab32072; RRID:AB_731658
Anti-LSD1	Gift from Jiemin Wong	N/A
Anti-beta-tubulin	Abmart	Cat#M30109; RRID:AB_2916070
Chemicals, peptides, and recombinant proteins		
Puromycin	Selleck	S7417
Recombinant Mouse LIF Protein	Millipore	ESG1107
Protease inhibitor cocktail	Roche	11836153001
Critical commercial assays		
GeneJET RNA Purification Kit	ThermoFisher Scientific	K0731
PerfectStart® Uni RT&qPCR Kit	Transgen Biotech	AUQ-01
PerfectStart® Green qPCR SuperMix	Transgen Biotech	AQ601-01
AMPure XP	Beckman Coulter	A63881
Deposited data		
RNAPII ChIP-seq datasets of WT and LSD1 knockdown ESCs	This paper	GEO: GSE173089
Mouse ESCs MYC ChIP-seq	Chen et al., 2008	GEO: GSE11431
Mouse ESCs NANOG ChIP-seq	Whyte et al., 2013	GEO: GSE44286
Mouse ESCs Sox2 ChIP-seq	Whyte et al., 2013	GEO: GSE44286
Mouse ESCs SPT5 ChIP-seq	Rahl et al. (2010)	GEO: GSE20530
Mouse ESCs NELF-A ChIP-seq	Rahl et al. (2010)	GEO: GSE20530
Mouse ESCs BRD4 ChIP-seq	Gatchalian et al. (2018)	GEO: GSE111264
Mouse ESCs LSD1 ChIP-seq	Whyte et al. (2012)	GEO: GSE27841
Mouse ESCs CDK9 ChIP-seq	Whyte et al., 2013	GEO: GSE44286
Mouse ESCs TBP ChIP-seq	https://www.ncbi.nlm.nih.gov/geo/	GEO: GSE22303
Mouse ESCs CoREST ChIP-seq	Whyte et al. (2012)	GEO: GSE27841
Mouse ESCs HDAC1-2 ChIP-seq	Whyte et al. (2012)	GEO: GSE27841
Mouse ESCs H3K27ac ChIP-seq	Yang et al. (2019)	GEO: GSE117896
Mouse ESCs H3K4me3 ChIP-seq	Yang et al. (2019)	GEO: GSE117896
Mouse ESCs H3K27me3 ChIP-seq	Yang et al. (2019)	GEO: GSE117896
Mouse ESCs GRO-seq	Williams et al. (2015)	GEO: GSE43390
Mouse ESCs LSD1 KD microarray	Foster et al. (2010)	GEO: GSE21131
Mouse ESCs cMYC KD RNA-seq	Seruggia et al. (2019)	GEO: GSE113329
Human ESCs LSD1 KD microarray	Adamo et al. (2011)	GEO: GSE24844
Mouse ESCs 0h RNA-seq	Yang et al. (2019)	GEO: GSE117896
Experimental models: Cell lines		
E14TG2a mouse embryonic stem cells	ATCC	CRL-1821
Oligonucleotides		
shRNA-#1 targeting sequence for LSD1: GAGTTGAAAGAGCTTCTTAAT	This paper	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Continued		
shRNA-#2 targeting sequence for LSD1: CCACAAGTCAAACCTTTATTT	This paper	N/A
shRNA-#1 targeting sequence for c-MYC: ACTTCACCAACAGGAACTATG	This paper	N/A
shRNA-#1 targeting sequence for c-MYC: TGGAGATGATGACCGAGTTAC	This paper	N/A
Primers for RT-qPCR. See Table S1	This paper	N/A
Primers for ChIP-qPCR. See Table S2	This paper	N/A
Software and algorithms		
PAD2	This paper	http://pad2.maths.usyd.edu.au
R version 4.1.0	R Development Core Team, 2016	https://www.R-project.org/
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	Li et al. (2009)	http://samtools.sourceforge.net
deepTools 2.0	Ramírez et al. (2014)	https://deeptools.readthedocs.io/en/develop/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for reagent and resource may be directed to and will be fulfilled by the lead contact, Dr. Pengyi Yang (pengyi.yang@sydney.edu.au).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The authors declare that all data supporting the findings of this study are available and can be accessed from the following accession codes. The accession codes for previously reported datasets are reported in Key Resources Table. The accession codes for RNAPII ChIP-seq datasets of WT and LSD1 knockdown ESCs generated from this study is GEO: GSE173089. All published computational tools and programs used in this study are clearly referenced in the STAR Methods section. Custom code used for data analysis in this study are available from Github repository (https://github.com/HaniJieunKim/POL2LSD1Manuscript). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

E14TG2a mouse ESCs (ATCC) were routinely cultured on gelatin-coated plates in DMEM (GIBCO) supplemented with 15% ESC-qualified fetal bovine serum (GIBCO), 10 mM 2-mercaptoethanol, 1mM nonessential amino acids, and 1,000 U/ml of LIF (Millipore). For cell maintenance, ESCs were cultured in the serum-free ESGRO medium (Millipore). 293T cells were maintained in DMEM with 10% fetal bovine serum (Biological Industries) at 37°C with 5% CO_2.

METHOD DETAILS

Lentiviral shRNA constructs, lentivirus preparation and transduction

shRNA constructs for mouse Lsd1 were designed to target 21-base-pair (bp) transcript-specific regions. Complementary single-stranded oligos were annealed and cloned into AgeI and EcoRI sites of the pLKO.1-TRC lentivirus vector as suggested by the RNAi consortium. The sequences targeted by the shRNAs are as follows: Lsd1-1#, GAGTTGAAAGAGCTTCTTAAT; Lsd1-2#, CCACAAGTCAAACCTTTATTT; c-Myc-1#, ACTTCACCAACAGGAACTATG; c-Myc-2#, TGGAGATGATGACCGAGTTAC. Lsd1-2# and c-Myc-2# were used in the RT-qPCR and ChIP-qPCR analyses. All plasmids were verified by sequencing.

Lentivirus production and transduction were carried out by transfecting 293T cells. Briefly, shRNA constructs were cotransfected with the envelope plasmid pMD2.G (Addgene) and the packing vector psPAX2(pMD2.G(Addgene_12259) and psPAX2 (Addgene_12260), kindly gifted from Didier Trono, using lipofiter reagents (Hanbio). Viruses were harvested 48 h after transfection and tittered by colony formation assay. Then, the viruses were used to transduce E14 ESCs. To select for infected cells, puromycin was added to the media at a concentration of 1 μg/mL after 24 h of infection cells were treated with puromycin for two days and harvested for RNA extraction using RT-qPCR at day 4.

RT-qPCR

Total RNA was extracted using the GeneJET RNA Purification Kit (ThermoFisher Scientific, K0731). cDNA was synthesized from 1 μg of total RNA using the PerfectStart® Uni RT&qPCR Kit for qPCR (Transgen Biotech, AUQ-01) according to the manufacturer's instructions. RT-qPCR was carried out with PerfectStart® Green qPCR SuperMix (Transgen Biotech, AQ601-01) using Roche LightCycler 480 (Roche). The relative expression level was determined by using the $2-\Delta Cq$ method and normalized to GAPDH expression. The primers used in the paper are listed in Table S1.

Co-immunoprecipition

E14 ESCs were collected and lysed in lysis buffer (50 mM Tris–HCl [pH 7.5], 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, protease inhibitor cocktail and 1 mM DTT). The lysates were treated with 20 U/mL benzonase (Solarbio) and then were cleared by centrifugation at 12 000 rpm for 20 min at 4°C. The supernatant was diluted with binding buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 8% glycerol, 1 mM EDTA, 1× protease inhibitor cocktail and 1 mM DTT) until the concentration of Triton X-100 decreased to 0.2%. Antibodies (c-Myc, Abcam#ab32072; LSD1, a gift kindly provided by Jiemin Wong) were added to the lysate at a final concentration of 1 μg/mg and incubated overnight at 4°C, followed by antibody-protein complex capture with Protein A/G Magnetic Beads (Pierce, cat#88802). After washing (50 mM Tris–HCl [pH 7.5], 150 mM NaCl, 0.1% Triton X-100, 1 mM EDTA, protease inhibitor cocktail and 1 mM DTT), complexes were eluted, boiled in 1× SDS loading buffer (50mM Tris-HCl [pH6.8], 2% SDS, 100mM DTT, 10% Glycerol and 0.1% bromophenol blue dye) and analyzed by western blot.

Western blot analysis

Cells were lysed in ice-cold cell lysis buffer (50 mM Tris [pH 7.5], 150 mM NaCl, 1% Triton X-100, 1 mM EDTA) supplemented with protease inhibitor cocktail (Roche, cat#11836153001) and DTT. The proteins were separated on 10% SDS-PAGE gels and electrotransferred onto Nitrocellulose Transfer Membrane (PALL, P-N66485) that were placed in 5% skimmed milk for blocking. Probing was performed with specific primary antibodies (4°C, overnight) and horseradish peroxidase-conjugated secondary antibodies (Room temperature for 1 h). The primary antibodies used in this study targeted LSD1 (gifts of Prof.Jiemin Wong, dilution 1:1000; HUABIO # ER1802-12, dilution 1:1000), c-Myc (Abcam#ab32072, dilution 1:1000; ZENBIO #380784, dilution 1:1000), β-Tubulin (Abmart #M30109, dilution 1:2000), Pol II (Santa Cruz #sc-47701, dilution 1:1000), Phospho-Pol II (Ser2) (ZENBIO #381832, dilution 1:1000), Phospho-Pol II (Ser5) (ZENBIO #381833 dilution 1:1000), GAPDH(HUABIO #ET1601-4, dilution 1:1000).

ChIP-qPCR and ChIP-seq

We followed the established ChIP experiment protocol (Li et al., 2017) where lentivirus transduced ESCs were fixed using 1% formaldehyde for 10 min, and 0.125 M glycine was added to stop the fixation. Cells were then harvested, and DNA was fragmented to 300–500 base pairs (bp) by sonication with Scientz08-III sonicator (SCIENTZ). 10% of the total supernatant was retained for sequencing the input control. Immunoprecipitation was performed with antibodies (Pol II, Santa Cruz#sc-47701; c-Myc, Abcam#ab32072; LSD1 antibody, a gift from Jiemin Wong) conjugated to Dynabeads protein A/G beads (Pierce, cat#88802). The DNA from ChIP was eluted, reverse-cross-linked, extracted by AMPure XP beads (Beckman). qPCR from ChIP experiments was carried out with PerfectStart® Green qPCR SuperMix (Transgen Biotech, AQ601-01) using Roche LightCycler 480 (Roche). For primer sequences used in ChIP-qPCR validation experiments, see Table S2. For ChIP-seq, 1 ng of ChIP DNA or input DNA was used to generate sequencing libraries using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). Libraries were sequenced on the NovaSeq 6000 (Illumina) using the 75-nt sequencing protocol.

PAD implementation

Public ChIP-seq data analysis

All public ChIP-seq data analyzed in this study were generated from mouse ESC lines (Yang et al., 2017). Reads from each ChIP-seq dataset were aligned to the mouse genome (mm9 assembly). For each DNA-binding protein, mapped reads were binned across the mouse genome with 1kb bin size and quantified as reads per base/kilobase per million reads (RPKM) and used in PAD clustering (see next section). ChIP-seq read density plots were generated by calculating the number of reads within ± 2.5 kb upstream and downstream of sites of interest in 100 bp windows and normalized to RPKM and plotted as histograms. Data for heatmaps were generated in a similar manner.

PAD framework

PAD2 was developed in Python 3.7 based on Django web framework and the interactive plots in PAD2 were rendered using Plotly.js, an extension from PAD clustering (Yang et al., 2017) implementation. PAD clustering (Yang et al., 2017) (http://pad2.maths.usyd.edu.au) was previously developed to characterize the co-localization of TFs and epigenomic marks at various genomic regions based on their ChIP-seq profiling. Here, we extended PAD by increasing the number of partitions for genomic binding sites and histone modifications to 12 functional genomic regions identified by ChromHMM (Ernst and Kellis, 2017) in ESCs (Figure S1).

Mapping of genome to functional regions

A collection of transcription factors, chromatin remodellers and histone mark ChIP-seq datasets have been processed and quantified in 1kb bin across the genome. For each of the peak files, we mapped them to the 12 genomic functional regions (Figure S1) by using the intersect method in BEDTools v2.28.0 (Quinlan, 2014) and calculate their fold change.

The fold change of a protein binding at each functional region is defined as follows:

$$\text{fold change of } p_i = \frac{p_i}{q_i}$$

where p_i denote the percentage of protein binding at genomics region i, and q_i denote the percentage of genomic region i covers in a whole genome.

Clustering of DNA binding proteins

To perform clustering of DNA binding proteins at different genomic regions, we combine a set of peak files from a user-specified selection of DNA binding proteins and functional regions to a matrix. We then perform hierarchical clustering with Pearson's correlation as a similarity metric.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis and visualization of ChIP-seq data

ChIP-seq data (pair-end reads) generated from this study was aligned using STAR (v2.5.4b) (Dobin et al., 2013), with following parameters: –outFilterMultimapNmax 1 –alignEndsType EndToEnd –outFilterMismatchNmax 3 –alignIntronMax 1. BedTools (v2.26.0) (Quinlan, 2014) and SamTools (v1.7) (Li et al., 2009) were used to manipulate and convert files. Coverage analyses on bed and bam files were performed using BedTools and DeepTools (v2.0) (Ramírez et al., 2014) were applied to generate and visualize the matrices as heatmaps and to create the average ChIP-seq profiles.

Binding sites and target genes

To define binding sites for each DNA-binding factor, aligned reads were processed using SISSRs with a common input (Oldfield et al., 2014) and those with a p < 0.001, a stringent cut-off, were called as binding sites (Jothi et al., 2008). MYC-specific and LSD1-specific sites were defined as any MYC or LSD1 peaks that lack LSD1 or MYC ChIP-seq signals, respectively, at the same locus. Co-localized sites were determined as loci where MYC and LSD1 peaks overlap (within 500 bp window). Based on the mm9 RefSeq annotation, genes with closest TSSs to binding sites (within 1.5 kb) of a DNA-binding factor are assigned as the target genes of that factor.

Calculation of RNAPII pausing index

To calculate the RNAPII pausing index, we used the method described by (Day et al., 2016). Specifically, for each gene, we calculated the pausing index as follows:

$$Pausing\ Index(PI)\ =\ \frac{Read\ count\ in\ TSS\ region/L1}{Read\ count\ in\ gene\ body/L2}$$

where the transcription start site (TSS) region of a gene is defined as the −50 bp to +300 bp around the TSS and the gene body is defined as +300 bp downstream of the TSS to +3 kb past the transcription end site. To segregate genes into different tiers of pausing, we categorized genes into five pausing groups based on their RNAPII pausing index at the promoter so that each group contained roughly the same number of genes. The same pausing groups (i.e., homologous genes) were used to evaluate the effect of LSD1 knockdown in hESCs. Differences in median RNAPII pausing indices were calculated using the Wilcoxon Rank Sum test where the number of asterisks denote the level of statistical significance: ***, $p < 0.001$; **, $p < 0.01$; and *, $p < 0.05$.

Gene expression analysis

Previously published data from LSD1 knockout and WT mouse ESC samples measured using Illumina MouseWG-6 v2.0 expression beadchip were downloaded from the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) with accession numbers GSE21131(Foster et al., 2010). Log2 fold change for each gene was then calculated by averaging across duplicates in either LSD1 KD or WT samples and subtracting log2 transformed values in LSD1 KD measurement with WT measurement. The same approach was taken to process data from MYC knockdown and WT mouse ESC samples measured using Illumina HiSeq 2500 (GSE113329 (Seruggia et al., 2019)) and microarray data from LSD1 knockdown and WT human ESC samples measured using Affymetrix Human Promoter 1.0R Array from human ESCs (GSE24844)(Adamo et al., 2011). RNA-seq data from ESCs in naïve state (GSE117896) was used for defining expressed genes (Yang et al., 2019). Specifically, expressed genes were defined as any gene with regularised log expression equal or higher than 5.

Motif and transcription factor enrichment analysis

For motif enrichment analysis, DNA sequences (500 bp) flanking the center of the binding sites for each factor were first extracted from mm9 assembly using the 'getfasta' of the bedtools (Quinlan and Hall, 2010). These sequences were subsequent searched using MEME (Bailey et al., 2006) using a minimum and maximum window size of 5 and 15, respectively, and the zero or one motif occurrence per sequence (zoops) option. The top two most enrichment motifs (based on the MEME reported E-value) were presented and annotated using known motifs. For the transcription factor enrichment analysis, we used the target gene sets and the ChIP-X Enrichment Analysis 3 (ChEA3) software to generate the gene-set associated TF ranks (Keenan et al., 2019). The transcription factor target over-representation analysis is performed using putative target genes determined by ChIP-seq experiments from ENCODE. A Fisher's exact test (FET) with a background of 20,000 is used to quantify the significance in the overlap between the input gene set to the TF target gene sets, and the top ten TFs most closely associated were visualized.

Functional enrichment analysis

Functional enrichment analysis, in terms of overrepresentation of genes from a pathway, was performed using Fisher's exact test against the gene ontology (GO) terms from the GO database (Blake et al., 2015). p-values were adjusted using the Benjamini-Hochberg method.