Integrated transcriptome and epigenome analyses identify alternative splicing as a novel candidate linking histone modifications to embryonic stem cell fate decision

Yungang Xu[1,2], Guangxu Jin[2,3], Liang Liu[2], Dongmin Guo[2], Xiaobo Zhou[1,2,3,*]

[1]Center for Systems Medicine, School of Medical Bioinformatics, University of Texas Health Science Center at Houston, TX, 77030, USA

[2]Center for Bioinformatics and Systems Biology, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA

[3]Wake Forest Baptist Comprehensive Cancer Center, Winston-Salem, NC 27157, USA

[*]Correspondence: xiaobo.zhou@uth.tmc.edu

Running title: AS links histone modification to ESC fate decision

Keywords: alternative splicing, histone modification, embryonic stem cell, cell fate decision, cell cycle machinery

Abstract

Understanding embryonic stem cell (ESC) fate decision between self-renewal and proper differentiation is important for developmental biology and regenerative medicine. Attentions have focused on underlying mechanisms involving histone modifications (HMs), alternative pre-mRNA splicing (AS), and cell-cycle progression. However, their intricate interrelations and joint contributions to ESC fate decision remain unclear. We performed integrative analyses on transcriptomic and epigenomic data derived from human ESC (hESC) and five types of differentiated cells. We identified thousands of AS exons and revealed their development- and lineage-dependent characterizations. Following the observation that dynamic HM changes predominantly occur in AS exons upon hESCs differentiation, we identified 3 of 16 investigated HMs (H3K36me3, H3K27ac, and H4K8ac) that are strongly associated with 52.8% of hESC differentiation-related AS events. Further analyses showed that the HM-associated AS genes predominantly function in G2/M phases and ATM/ATR-mediated DNA damage response pathway for cell differentiation, whereas HM-unassociated AS genes enrich in G1 phase and pathways for self-renewal. These results imply a potential epigenetic mechanism by which some HMs contribute to ESC fate decision through the AS regulation of specific pathways and cell-cycle genes. We exemplified the potential mechanism by a cell cycle-related transcription factor, PBX1, which regulates the pluripotency regulatory network by binding to NANOG. We found that the isoform switch between *PBX1a* and *PBX1b* is strongly associated with H3K36me3 alteration and implicated in hESC fate determination. Supported by extended dataset from Roadmap/ENCODE projects, we identified the alternative splicing of PBX1 as a novel candidate linking H3K36me3 to embryonic stem cell fate decision.

Introduction

Embryonic stem cells (ESCs), the pluripotent stem cells derived from the inner cell mass of a blastocyst, provide a vital tool for studying the regulation of early embryonic development and cell-fate decision, and hold the promise for regenerative medicine (Thomson et al. 1998). The past few years have witnessed remarkable progress in understanding the ESC fate decision, i.e. either pluripotency maintenance (self-renewal) or proper differentiation (Graveley 2011). The underlying mechanisms have been largely expanded from the core pluripotent transcription factors (TFs) (Boyer et al. 2005), signaling pathways (Watabe and Miyazono 2009; Lian et al. 2013; Ogaki et al. 2013; Sakaki-Yumoto et al. 2013; Clevers et al. 2014; Itoh et al. 2014), specific microRNAs (Tay et al. 2008; Shenoy and Blelloch 2014), and long non-coding RNAs (Fatica and Bozzoni 2014) to alternative pre-mRNA splicing (AS) (Salomonis et al. 2010; Gabut et al. 2011), histone modifications (HMs) (Hawkins et al. 2010; Gifford et al. 2013; Xie et al. 2013b; Boland et al. 2014), and cell-cycle machinery (Vallier 2015). These emerging mechanisms suggest their intricate interrelations and potential joint contributions to ESC pluripotency and differentiation, which, however, remain unknown.

Alternative splicing (AS) is a most important pre-mRNA processing to increase the diversity of transcriptome and proteome in tissue- and development-dependent manners (Pan et al. 2008). The estimates based on RNA-seq revealed that up to 94%, 60% and 25% of genes in human, *Drosophila melanogaster* and *Caenorhabdities elegans*, respectively, undergo AS (Pan et al. 2008; Wang et al. 2008a; Gerstein et al. 2010; Graveley et al. 2011; Ramani et al. 2011). AS also provides a powerful mechanism to control developmental decision in ESCs (Yeo et al. 2007; Salomonis et al. 2009; Lu et al. 2013). Specific isoforms are necessary to maintain both the identity and activity of stem cells, and switching to different isoforms ensures proper differentiation (Kalsotra and Cooper 2011). Especially, the AS of TFs plays major roles in ESC fate determination, such as *FGF4* (Mayshar et al. 2008) and *FOXP1* (Gabut et al. 2011) for hESC, *Tcf3* (Salomonis

et al. 2010) and *Sall4* (Rao et al. 2010) for mouse ESCs (mESCs). Understanding the precise regulations on AS would contribute to elucidation of ESC fate decision and has attracted extensive efforts. For many years, studies aiming to shed light on this process focused on the RNA level, characterizing the manner by which splicing factors and auxiliary proteins interact with splicing signals, thereby enabling, facilitating and regulating RNA splicing. These *cis*-acting RNA elements and *trans*-acting splicing factors (SFs) have been assembled into splicing code (Barash et al. 2010), revealing a number of AS regulators critical for ESC differentiation, such as MBNL (Han et al. 2013) and SON (Lu et al. 2013). However, these genetic controls are far from sufficient to explain the faithful regulation of AS (Schwartz et al. 2009). Especially in some cases that tissue-specific AS patterns exist despite the identity in sequences and ubiquitously expression of involved SFs (Kornblihtt et al. 2004; Wang and Cooper 2007), indicating additional regulatory layers leading to specific AS patterns. As Expected, we increasingly understand that splicing is not an isolated process; rather it occurs co-transcriptionally and is presumably also regulated by transcription-related processes. Emerging provocative studies have unveiled that AS is subject to extensive controls not only from genetic but also epigenetic mechanisms due to its co-transcriptional occurrence (Luco et al. 2011). The epigenetic mechanisms, such as HMs, benefits ESCs by providing an epigenetic memory for splicing decisions, so that the splicing pattern could be passed on during self-renewal and be modified during differentiation without the requirement of establishing new AS rules (Luco et al. 2011).

HMs have long been thought to play crucial roles in ESCs maintenance and differentiation by determining what parts of the genome are expressed. Specific genomic regulatory regions, such as enhancers and promoters, undergo dynamic changes in HMs during ESC differentiation to transcriptionally permit or repress the expression of genes required for cell fate decision (Xie et al. 2013a). For example, the co-occurrence of the active (H3K4me3) and repressive (H3K27me3) HMs at the promoters of developmentally regulated genes defines the bivalent domains, resulting

in the poised states of these genes (Sharov and Ko 2007). These poised states will be dissolved upon differentiation to allow these genes to be active or more stably repressed depending on the lineage being specified, which enables the ESCs to change their identities (Singh et al. 2015). In addition to above roles in determining transcripts abundance, HMs are emerging as major regulators to define the transcripts structure by determining how the genome is spliced when being transcribed, adding another layer of regulatory complexity beyond the genetic splicing code (Kouzarides 2007). A number of HMs, such as H3K4me3 (Sims et al. 2007), H3K9me3 (Piacentini et al. 2009), H3K36me3 (Luco et al. 2010; Pradeepa et al. 2012), and acetylation of H3 (Gunderson and Johnson 2009), have been proven to regulate AS by either directly recruiting chromatin-associated factors and SFs or indirectly modulating transcriptional elongation rate (Luco et al. 2011). However, few studies focused on epigenetic regulations on AS in the context of cell fate decision.

Additionally, cell-cycle machinery dominates the mechanisms underlying ESC pluripotency and differentiation (Dalton 2015; Vallier 2015). Changes of cell fates require going through cell-cycle progression. Studies in mESCs (Coronado et al. 2013) and hESCs (Pauklin and Vallier 2013; Gonzales et al. 2015) found that the cell fate specification starts in G1 phase, when ESCs can sense differentiation signals. Cell fate commitment is only achieved in G2/M phases, when pluripotency is dissolved through cell-cycle dependent mechanisms. However, whether the HMs and AS, and their interrelations, are involved in these cell cycle-dependent mechanisms remains unclear. Therefore, it is intuitive to expect that HMs could contribute to ESC pluripotency and differentiation by regulating the AS of genes required for specific processes, such cell-cycle progression. Nevertheless, we do not even have a comprehensive view of how HMs relate to AS outcome at genome-wide during ESC differentiation. Therefore, further studies are required to elucidate the extent to which the HMs are associated with specific splicing repertoire and their joint contributions to ESC fate decision between self-renewal and proper differentiation.

To address these gaps in current knowledge, we performed genome-wide association studies between transcriptome and epigenome of the differentiations from the hESCs (H1 cell line) to five differentiated cell types, representing multiple lineages of different developmental levels (Xie et al. 2013b). First, we used RNA-seq data to identify exons that are alternatively spliced during hESC differentiation. We identified 3,513 mutually exclusive exons (MXE) and 3,678 skipped exons (SE) that are differentially spliced between the hESCs and differentiated cells. These hESC differentiation-related AS events involve ~20% of expressed genes and characterize the multiple lineage differentiations. Second, we profiled 16 HMs with ChIP-seq data available for all six cell types, including nine types of acetylation and seven types of methylation. Following the observation that the dynamic changes of most HMs are enriched in AS exons and significantly different between inclusion-gain and inclusion-loss exons, we found that 3 of the 16 investigated HMs (H3K36me3, H3K27ac, and H4K8ac) are strongly associated with 52.8% of hESC differentiation-related AS exons. We then linked the association between HMs and AS to cell-cycle progression based on the additional discovery that the AS genes predominantly function in cell-cycle progression. More intriguingly, we found that HMs and AS are associated in G2/M phases and involved in ESC fate decision through promoting pluripotency state dissolution, repressing self-renewal, or both. Specially, we suggested a candidate of the H3K36me3-associated isoform switch from PBX1a to PBX1b that is implicated in hESC differentiation by repressing the pluripotency regulatory network. Collectively, we presented a potential mechanism conveying the HM information into cell fate decision through the regulation of AS, which will drive further mechanistic and experimental studies on the involvements of HMs in cell fate decision via determining the transcript structure rather than only the transcript abundance.

Results

Alternative splicing characterizes hESC differentiation

The role of AS in regulation of ES cell fates adds another notable regulatory layer to the know mechanisms that govern stemness and differentiation (Aaronson and Meshorer 2013). To screen the AS events associated with ES cell fate decision, we investigated a panel of RNA-seq data during hESC (H1 cell line) differentiation (Xie et al. 2013b). We considered four cell types directly differentiated from H1 cells, including trophoblast-like cells (TBL), mesendoderm (ME), neural progenitor cells (NPC), and mesenchymal stem cells (MSC) (Figure S1A). We also considered IMR90, a cell line for primary human fetal lung fibroblast, as an example of terminally differentiated cells (Figure S1A). Comparisons between the differentiated cell types and hESCs represent five cell lineages of different developmental levels (Figure S1A). We identified two types of AS exons with the changes of 'per spliced in' (ΔPSIs) are more than 0.1 (inclusion-loss) or less than -0.1 (inclusion-gain), and with the FDRs are less than 0.05 based on the measurement used by rMATS (Shen et al. 2014) (Figure S1B, see Methods). They are referred to as hESC differentiation-related AS exons, including 3,513 mutually exclusive exons (MXE) and 3,678 skipped exons (SE) (Figure S1C, Supplemental Table S1).

These hESC differentiation-related AS exons possess the typical properties as previously described (Magen and Ast 2005; Zheng et al. 2005): (1) most of their hosting genes are not differentially expressed between hESCs and differentiated cells (Figure S1D); (2) they tend to be shorter with much longer flanking introns compared to the average length of all exons and introns (RefSeq annotation), respectively (Figures S2A, B); (3) the arrangement of shorter AS exons surrounded by longer introns is consistent across cell lineages and AS types (Figures S2C, D); and (4) the lengths of AS exons are more often divisible by three to preserve the reading frame (Figure S2E).

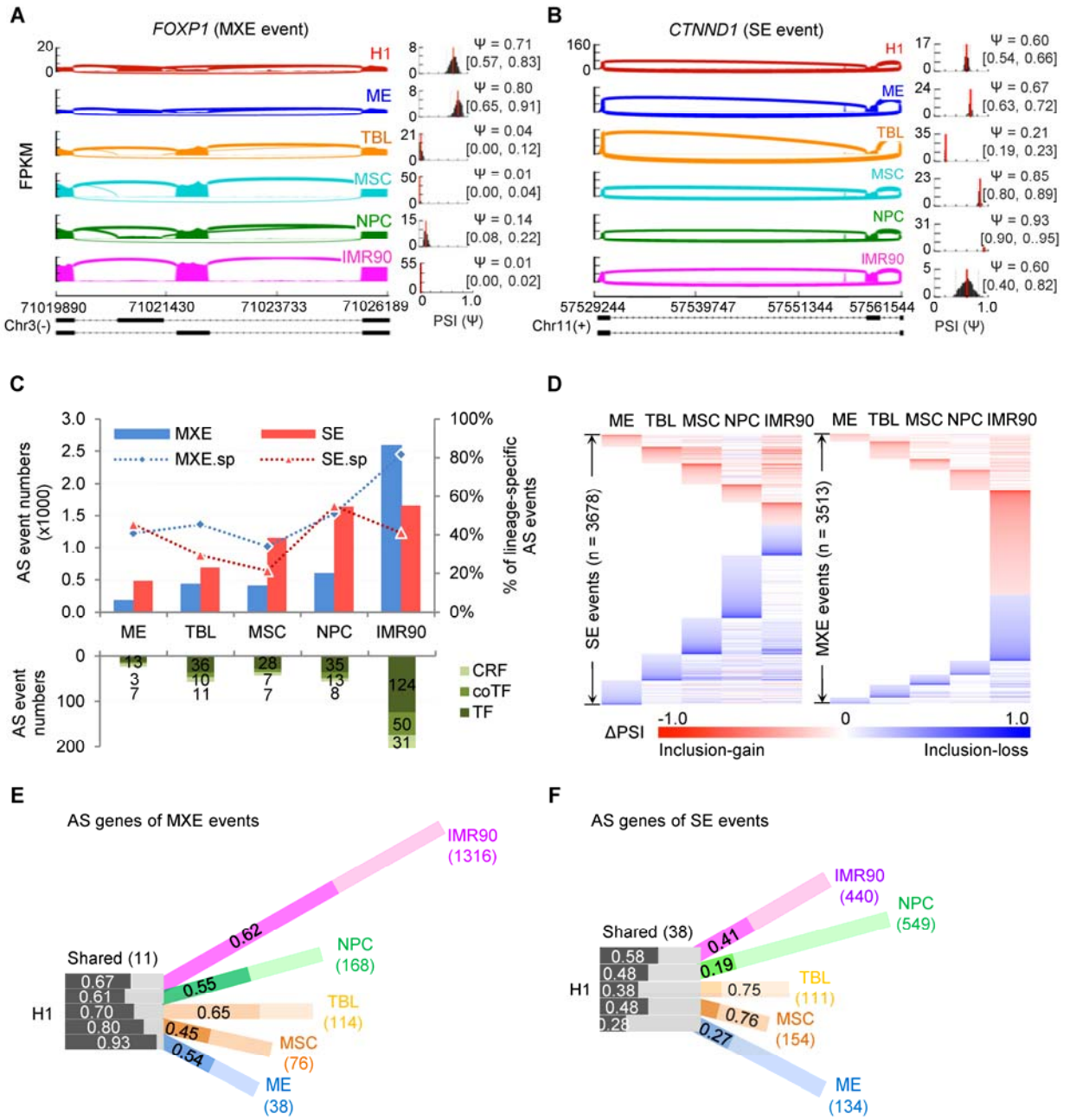

Figure 1 Alternative splicing characterizes the hESC differentiation. (A) and **(B)** Sashimi plots show two AS events of previously known ESC-specific genes, *FOXP1* (A) and *CTNND1* (B). Inset histograms show the PSIs (Ψ) of the AS exons in all cell types based on the MISO estimation. **(C)** Bar graph shows that the number of total AS events and lineage-specific AS events increase coordinately with the developmental levels. Higher developmental level induces more (lineage-specific) AS events. MXE.sp and SE.sp indicate the percentage of lineage-specific AS events. **(D)** Heat maps show the differential 'percent splice in' (ΔPSIs) of SE (left panel) and MXE (right panel) AS events (rows) for each cell lineage (columns). For MXE event, the ΔPSIs are of the upstream exons. **(E)** and **(F)** The hosting genes of MXE (E) and SE (F) AS events characterize cell lineages. Black-and-white bars refer to the common AS genes shared by all cell lineages, while the color bars indicate the lineage-specific AS genes. The length of the color bars is proportional to the percentage of lineage-specific genes. Dark fills indicate the inclusion-gain events, while light fills indicate the

inclusion-loss events. The numbers in the bars are the proportion of corresponding parts; the numbers in the parentheses are the numbers of common AS genes or lineage-specific AS genes of each lineage. Gain or loss for MXE events refers to the upstream exons. Also see **Figures S1–S3**.

During hESC differentiation, about 20% of expressed genes undergo AS (2,257 for SE and 2,489 for MXE), including previously known ESC-specific AS genes, such as the pluripotency factor *FOXP1* (Gabut et al. 2011) (Figure 1A) and the Wnt/β-catenin signaling component *CTNND1* (Salomonis et al. 2010) (Figure 1B). These hESC differentiation-related AS genes contain many transcription factors, transcriptional co-factors, chromatin remodeling factors, housekeeping genes, and bivalent domain genes implicated in ESC pluripotency and development (Sharov and Ko 2007) (Figures 1C and S1C). Enrichment analysis based on a stemness gene set (Pinto et al. 2015) also shows that hESC differentiation-related AS genes are enriched in regulators or markers most significantly associated with stemness of ESCs (Figure S3A, see Methods).

Clustering on AS events across cell lineages shows lineage-dependent splicing patterns (Figure 1D). Upon hESC differentiation, the SE exons tend to lose their inclusion levels (inclusion-loss), while the upstream exons of MXE events are likely to gain their inclusion levels (inclusion-gain) (Fisher's exact test, p = 3.83E-107). The numbers of AS events increase accordingly with the developmental level following hESC differentiation (Figure 1C). For example, the differentiation to ME involves the fewest AS events and ME presents the most stem-cell-like AS profiles, while the IMR90 has most AS events and exhibits AS profiles most similar to adult cells (Figure 1C, D). Inter-lineage comparisons show, on average, that 42.0% of SE and 56.4% of MXE events (Figures 1C, D and S3B, C), involved in 29.6% and 38.6% of AS hosting genes (Figures 1E, F and S3D, E), are lineage-specific. In contrast, only 0.65% of SE and 0.14% of MEX events (Figures S3B, C), involved in 0.49% and 1.52% of AS hosting genes, are shared by all lineages (Figures 1E, F and S3D, E). Similar trends are observed from pairwise comparisons (Figure S3F). Furthermore, one third of AS genes (n = 881) have both MXE and SE events (Figure S3G). Only four genes are common across all cell lineages and AS types; of these, the AS events of *Ctnnd1*

and *Mbd1* have been reported to regulate mESC differentiation (Salomonis et al. 2010). Together, these results demonstrate that alternative splicing depicts lineage- and developmental level-dependent characterizations of hESC differentiation.

Dynamic changes of histone modification predominantly occur in AS exons

In ESCs, epigenetic mechanisms contribute mainly to maintaining the expression of pluripotency genes and the repression of lineage-specific genes in order to avoid exiting from stemness. Upon differentiation, epigenetic mechanisms orchestrate the expression of developmental programs spatiotemporally to ensure heritability of existing or newly acquired phenotypic states. Though epigenetic signatures are mainly found to be enriched in promoters, it has become increasingly clear that they are also present in gene bodies, especially in exon regions, implying a potential link of epigenetic regulation to pre-mRNA splicing (Barski et al. 2007; Kornblihtt et al. 2009). Consistent with previous reports (Kornblihtt et al. 2004; Wang and Cooper 2007), we also observed that few involved SF genes are differentially expressed during H1 cells differentiation (Figure S3H, see Methods), which confirms the existence of additional layer of epigenetic regulations on AS. However, the extents to which the AS is epigenetically regulated and how these AS genes contribute to cell fate decision are poorly understood. We focused on 16 histone modifications, including nine histone acetylation and seven histone methylation that have available data in all six cell types (see Methods), and aimed to reveal their regulated AS genes during hESC differentiation.

To investigate whether the dynamic changes of these HMs upon cell differentiation prefer AS exons consistently (Figures 2A, B), we profiled the differential HM patterns of around the hESC differentiation-associated AS exons and the same number of randomly selected constitutive splicing (CS) exons of the same AS genes for each differentiation lineage. We compared the changes of ChIP-seq reads count (Δ reads number) in ±150bp regions around the splice sites

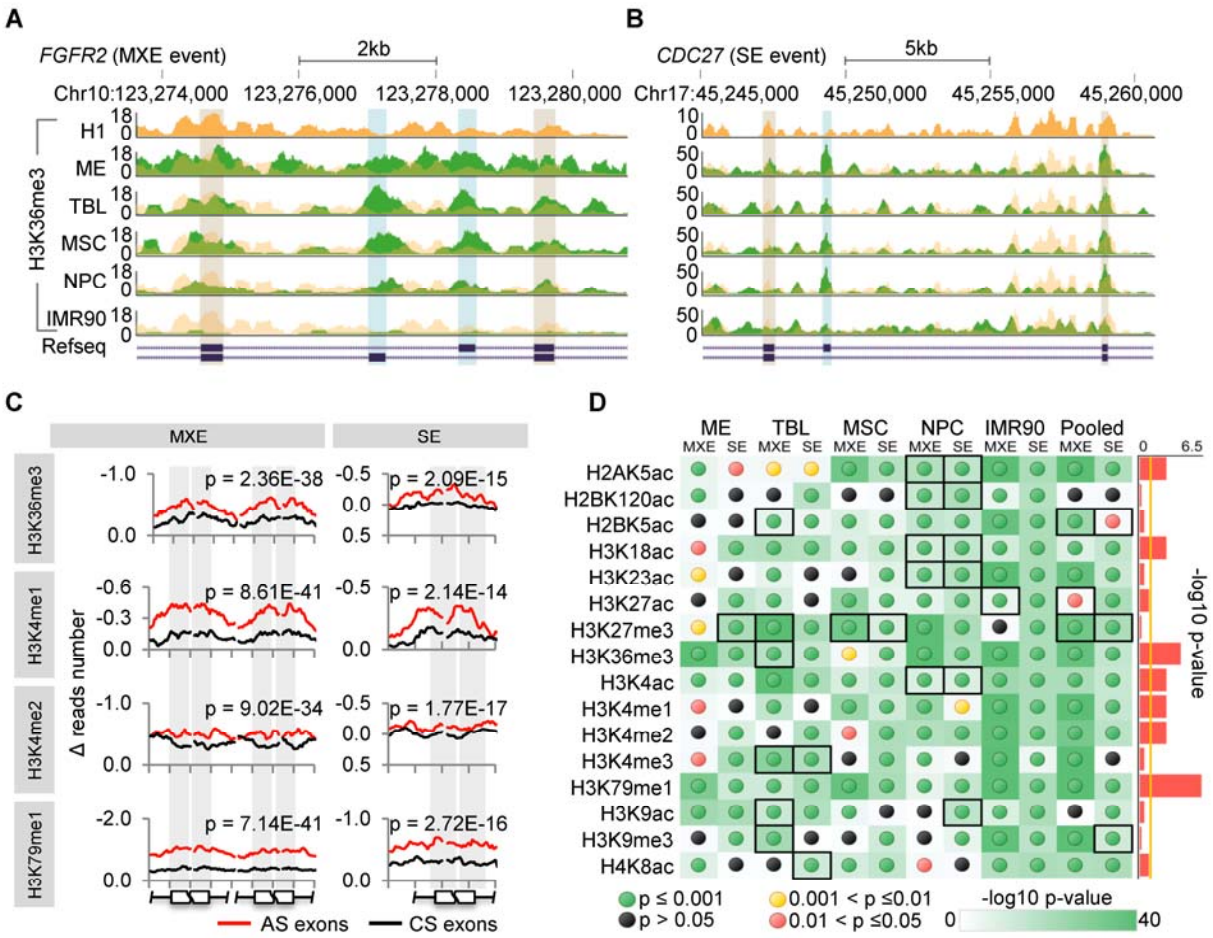

Figure 2 Dynamic changes of histone modifications predominantly occur in AS exons. (A) and **(B)** Genome browser views of representative H3K36me3 changes in MXE (exemplified as *FGFR2*) and SE (exemplified as *CDC27*) events respectively, showing that the changes of H3K36me3 around the AS exons (blue shading) are more significant than around the flanking constitutive exons (gray shading) in 4 H1-derived cells and IMR90. The tracks of H1 are duplicated as shadings (yellow) overlapping with other tracks of the derived cells (green) for a better comparison. **(C)** Representative profiles of HM changes (normalized Δ reads number) around the AS exons and randomly selected constitutive splicing (CS) exons upon hESC differentiation, shown as the average of all cell lineages pooled together. The ±150bp regions (exons and flanking introns) of the splice sites were considered and 15bp-binned to produce the curves. It shows that the changes of HMs are more significant around AS exons than around constitutive exons, especially in exonic regions (gray shading). The p-values, Mann-Whitney-Wilcoxon test. **(D)** The statistic significances for changes of all 16 HMs in all cell lineages and pooling them together (pooled), represented as the -log10 p-values based on Mann-Whitney-Wilcoxon test. The detailed profiles are provided in **Figure S4**. Black boxes indicate the cases that HMs around constitutive exons change more significantly than around AS exons, corresponding to the red-shaded panels in **Figure S4**. Side bars represent the significances whether the changes of HMs are consistently enriched in AS exons across cell lineages, showing the link strength between AS and HMs and represented as the -log10 p-value based on Fisher's exact test. The yellow vertical line indicates the significance cutoff of 0.05. Also see **Figure S4**.

upon hESC differentiation (Figure 2C, see Methods). Except a small part of cases (with black dots or boxes in Figure 2D), most HMs change more significantly around AS exons than around constitutive exons upon hESC differentiation (Mann-Whitney-Wilcoxon test, $p \leq 0.05$, Figures 2D and S4). Nevertheless, some HMs have strong links to AS, such as H3K79me1 and H3K36me3, while others only have weak link strengths, such as H3K27me3 and H3K9me3 (Figure 2D). This result is consistent with the fact that the former are involved in active expression and AS regulation (Kolasinska-Zwierz et al. 2009; Luco et al. 2010; Luco et al. 2011), while the latter are the epigenetic marks of repressed regions and heterochromatins (Mikkelsen et al. 2007). The link strengths are shown as the -log10 p-values to test whether the HM changes consistently prefer the AS exons across different cell lineages and AS types (Figure 2D side bar graph, see Methods). Taken together, these results, from a global view, revealed a potential regulatory link from HMs to RNA splicing, of which some are strong while the others are weak.

Three histone modifications are associated with alternative splicing upon hESC differentiation

To quantitatively associate the HMs with AS (Figure 3A), all ChIP-seq data were processed for narrow peak calling using MACS2 (Feng et al. 2012). For each AS exon of each differentiation lineage, we then quantified the differential inclusion levels, i.e., the changes of 'percent splice in' (ΔPSIs, Figure S1B), and the differential HMs signals, i.e., the changes of normalized narrow peak height of ChIP-seq (ΔHMs, Figure S5A, see Methods) between H1 and differentiated cells. We observed significant differences in all HM profiles (except H3K27me3, Figure S5B) between the inclusion-gain and inclusion-loss exons across cell lineages and AS types (Mann-Whitney-Wilcoxon test, $p \leq 0.05$) (Figures 3B and S5B). However, three independent correlation tests showed only weak quantitative associations between the ΔPSIs and ΔHMs for some HMs (Figures 3C and S5C), including eight HMs for MXE AS exons and eight HMs for SE AS exons.

The weak associations probably indicate that only subsets of AS exons are strongly associated with HMs, and *vice versa*, consistent with a recent report (Curado et al. 2015).

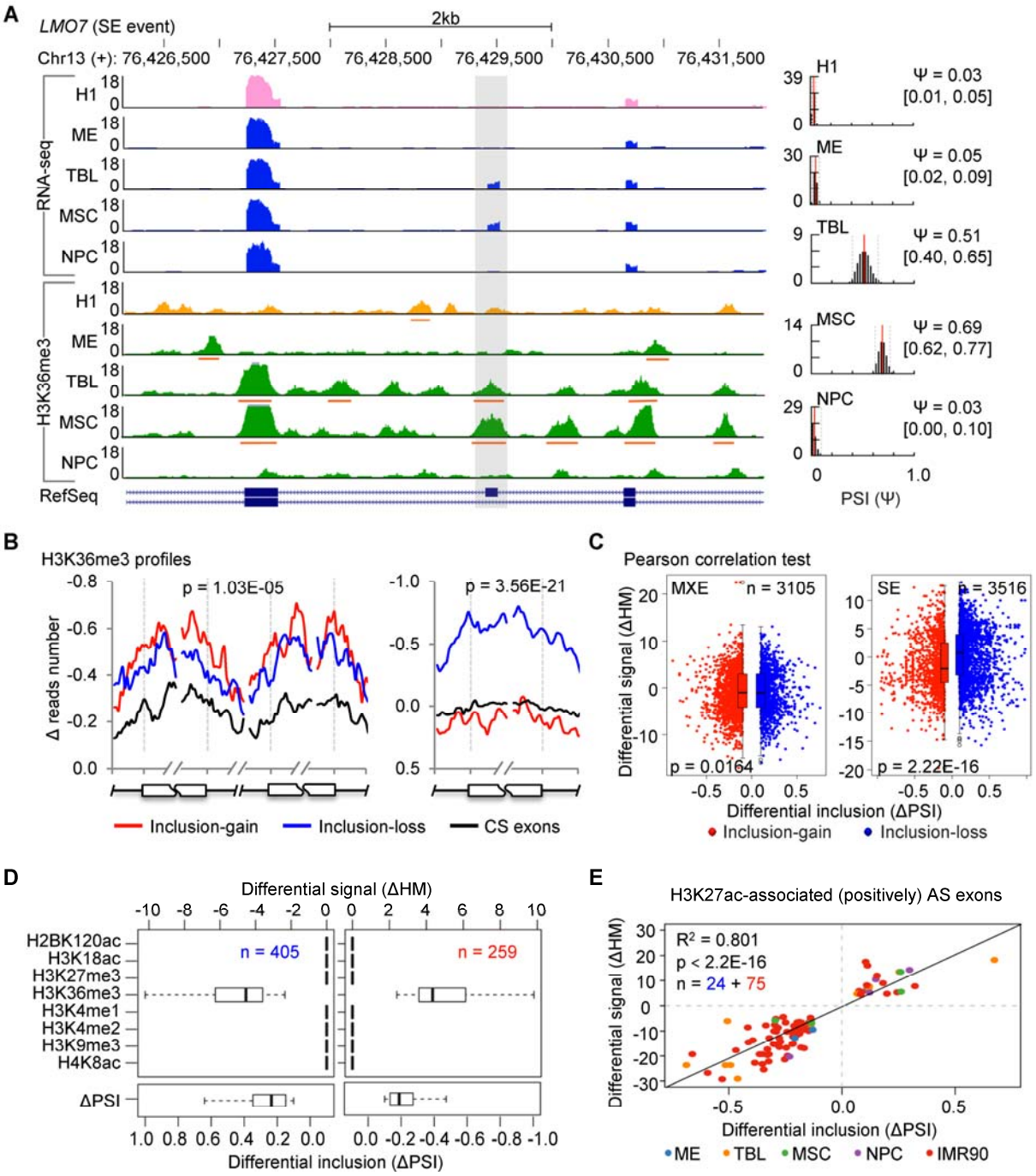

Figure 3 A subset of histone modifications and alternative splicing are strongly associated upon hESC differentiation. (A) Genome browser view of a representative gene (*LMO7*) that has an SE AS event, of which the inclusion level is positively associated with the surrounding H3K36me3 signal across cell types. Inset histograms show the PSIs (Ψ) of the AS exon in all cell types based on the MISO perturbation. Red horizontal bars below the ChIP-seq

tracks indicate the narrow peaks called by MACS2. **(B)** Representative profiles of HM (H3K36me3) changes (normalized Δ reads number) around the inclusion-gain (red lines) and inclusion-loss (blue lines) AS exons, as well as randomly selected constitutive splicing (CS) exons (black lines) for both MXE (left panel) and SE (right panel) AS events. It shows that HM changes are significantly different between inclusion-gain and inclusion-loss AS exons (p-values, Mann-Whitney-Wilcoxon test). **Figure S5B** provides the whole significances of all HMs across AS types and cell lineages. **(C)** Pearson correlation test between differential HM signals (ΔHMs) and differential inclusion levels (ΔPSIs), taking H3k36me3 as an example. **Figure S5C** provides the correlation test results of other HMs based on two more tests. **(D)** A representative k-means cluster shows a subset of SE AS events having negative correlation between the ΔPSIs and the ΔHMs of H3K36me3. **Figures S5D** and **S6** provide all clustering results. **(E)** Scatter plot shows that HM-associated AS events display significant correlations between the ΔPSIs and the ΔHMs upon hESC differentiation, taking H3K27ac–associated (positively) MXE events as an example. Also see **Figures S5, S6**.

To explore the subsets of highly associated AS exons and corresponding HMs, we performed *k*-means clustering on the sets of inclusion-gain and inclusion-loss exons of SE and MXE events, separately, taking the ΔHMs of eight identified HMs as epigenetic features (Figures 3D, S5D, and S6, see Methods). We obtained three subsets of HM-associated SE exons and three subsets of HM-associated MXE exons (Supplemental Table S2). The three HM-associated SE subsets include 180, 664, and 1,062 exons, and are negatively associated with H4K8ac (Figure S6), negatively associated with H3K36me3 (Figure 3D), and positively associated with H3K36me3 (Figure S6), respectively. The three HM-associated MXE subsets include 99, 821, and 971 exons, and are positively associated with H3K27ac (Figure 3E), negatively associated with H3K36me3 (Figure S6), and positively associated with H3K36me3 (Figure S6), respectively. The exons of each subset show significant correlations between their ΔPSIs and ΔHMs upon hESC differentiation (Figure 3E). These HM-associated AS exons account for 52.8% of all hESC differentiation-related AS events, on average (Figure S5E).

Of the three AS-associated HMs, H3K36me3 has both positive and negative correlations with AS exons. This is consistent with the fact that H3K36me3 has dual functions in regulating AS through two different chromatin-adapter systems, PSIP1/SRSF1 (Pradeepa et al. 2012) and MRG15/PTBP1 (Luco et al. 2010). The former increases the inclusion levels of targeting AS exons, whereas the latter decreases the inclusion levels (Luco et al. 2011). As expected, 139 and 11 of

our identified H3K36me3-associated AS genes have been reported to be regulated by SRSF1 (Pandit et al. 2013; Anczukow et al. 2015) (Figure S5F) and PTBP1 (Xue et al. 2009) (Figure S5G), respectively. Taken together, we showed that more than half (52.8%) of hESC differentiation-related AS events are significantly associated with 3 of 16 HMs during hESC differentiation, including H3K36me3, H3K27ac, and H4K8ac.

HM-associated AS genes predominantly function in G2/M phases to facilitate hESC differentiation

Epigenetic mechanisms have been proposed to be dynamic and play crucial roles in human ESC differentiation (Gifford et al. 2013; Xie et al. 2013b). Given the aforementioned associations between HMs and AS, and the well-established links between AS and hESC differentiation, we hypothesized that the three HMs (H3K36me3, H3K27ac, and H4K8ac) may contribute to stem cell differentiation through their associated AS events. To test our hypothesis and gain more insights into the differences between the HM-associated and -unassociated AS events, we performed comparative function analyses between their hosting genes, revealing that HMs are involved in alternatively splicing the core components of cell-cycle machinery and related pathways to regulate stem cell pluripotency and differentiation.

We found that HMs prefer to be associated with much shorter AS exons (Figure S7A, $p < 0.001$, Student's t-test), though AS exons are shorter than average length of all exons (Figure S2A). HM-associated genes ($n = 2125$) show more lineage specificity, i.e., more genes (49.76% versus 29.6% of MXE or 38.6% of SE genes) are lineage-specific (Figures S7B and S3D, E), regardless of whether IMR90 is included or not (Figure S7C). Only a few HM-associated genes are shared by different cell lineages, even in pairwise comparisons (Figure S7D); the most common shared genes are lineage-independent housekeeping genes (Figure S7E). These suggest that HM-associated AS genes contribute more to lineage specificity. Additionally, the HM-associated AS

genes (966 of 2,125) are more enriched in stemness signatures than do unassociated AS genes (429 of 1,057) (Figure 4A). TF binding enrichment analysis shows that HM-associated AS genes are likely to be regulated by TFs involved in cell differentiation, whereas HM-unassociated AS genes are likely to be regulated by TFs involved in cell proliferation and growth (Figure 4B). All these results suggest that HM-associated and -unassociated AS genes function differently during hESC differentiation.

Gene ontology (GO) enrichment analysis shows that more than half of the HM-associated AS genes (1,120 of 2,125) function in cell-cycle progression and exhibit more significant enrichment than do HM-unassociated AS genes (376 of 1,057, Figures 4C and S8A). The significance of the top enriched GO term (GO:0007049, cell cycle) is consistent across cell lineages, although HM-associated AS genes exhibit more lineage specificity and few of them are shared among lineages (Figures 4D, S7B-D, and S8B). These results suggest the involvement of HMs in AS regulation of the cell-cycle machinery that has been reported to be exploited by stem cells to control their cell fate decision (Vallier 2015).

Further study of the top enriched cell-cycle AS genes (Figures 4D and S8A) shows that HM-associated (n = 282) and -unassociated AS genes (n = 150) play roles in different cell-cycle phases and related pathways. The former are prone to function in G2/M phases and DNA damage response (Figures 4E, F). This indicates that HMs contribute to cell differentiation, at least partially, via AS regulations in these phases, which is consistent with the fact that inheritance of HMs in daughter cells occurs during the G2 phases (Vallier 2015). The latter play roles in G1 phase, cell cycle arrest, and Wnt/β-catenin signaling (Figures S8C, D). Since cell fate choices seem to occur or at least be initiated during G1/S transition (Pauklin and Vallier 2013), while cell fate commitment is achieved in G2/M (Gonzales et al. 2015), it could be rational for stem cells to change their identity during G2 phase when HMs are reprogrammed (Vallier 2015).

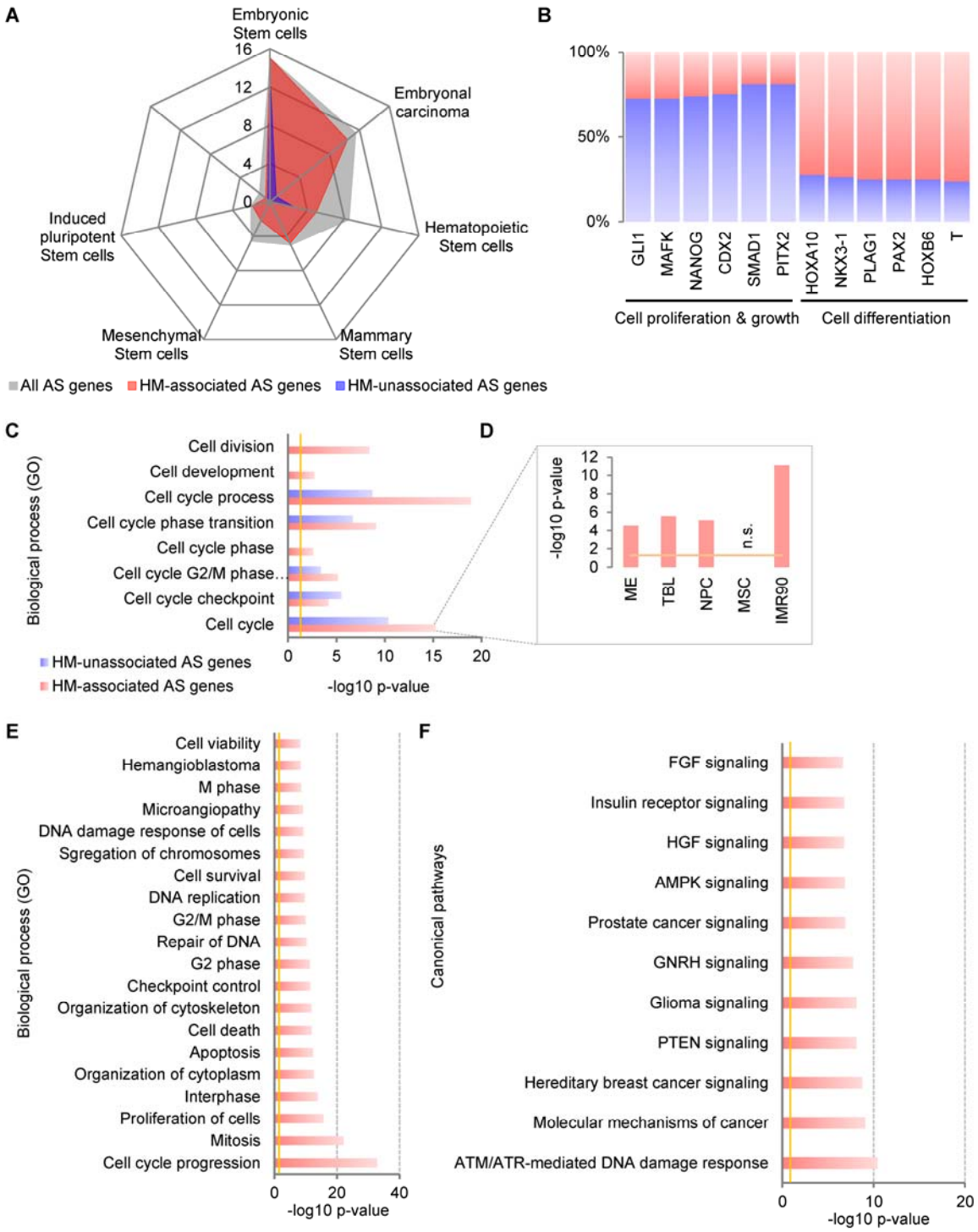

Figure 4 HM-associated AS genes predominantly function in G2/M cell-cycle phases contributing to hESC differentiation. (A) HM-associated AS genes are enriched more significantly in stemness signatures than HM-unassociated AS genes. **(B)** TF binding enrichment shows that HM-associated AS genes prefer to be regulated by TFs involved in cell differentiation, while the HM-unassociated AS genes are prone to be regulated by TFs involved in cell proliferation and growth. **(C)** GO enrichment analysis shows that HM-associated AS genes are enriched more significantly in cell cycle progression than HM-unassociated AS genes, shown as the -log10 p-values after FDR (< %5)

adjustment. **(D)** The significant enrichments of HM-associated AS genes in cell cycle are consistent across cell lineages, with the MSC as an exception that no significant enrichment was observed. **(E)** The top twenty enriched functions show that HM-associated AS genes involved in cell cycle progression prefer to function in G2/M phases and DNA damage response. **(F)** The canonical pathway enrichment shows that AMT/ATR-mediated DNA damage response is the top enriched pathway of HM-associated AS genes. The vertical lines (yellow) indicate the significance cutoff of 0.05. Also see **Figures S7-S10**.

Intriguingly, the top enriched pathway of HM-associated AS genes is 'ATM/ATR-mediated DNA damage response' (Figures 4F, S9A and S10). This pathway is activated in S/G2 phases and has been recently reported as a gatekeeper of the pluripotency state dissolution (PSD) that participates in allowing hESC differentiation (Gonzales et al. 2015). This enrichment is consistent regardless of whether all lineages are pooled together or considered separately (Figure S9C). A number of enriched cell-cycle AS genes are shared by multiple lineages (Figures S9B and S10). Our results suggest the presence of a combinational mechanism involving HMs and AS, wherein HMs facilitate the PSD and cell fate commitment by alternatively splicing the key components of the ATM/ATR pathway (Figure S10). Additionally, many cell-cycle TF genes are involved in the top enriched HM-associated AS gene set. The pre-B-cell leukemia transcription factor 1 (*PBX1*) is one of these genes that contribute to cell cycle progression and is discussed later in next section. Taken together, we suggest that 3 of 16 HMs function in positive or negative ways to affect the AS of subsets of genes, and further contribute to hESC differentiation in a cell-cycle phase-dependent manner. The results suggest a potential mechanistic model connecting the HMs, AS regulations, and cell-cycle progression with the cell fate decision.

H3K36me3-associated isoform switch of PBX1 contributes to hESC fate decision

The past few years have identified key factors required for maintaining the pluripotent state of hESCs (Chen et al. 2008; Silva et al. 2009), including NANOG, OCT4 (POU5F1), SOX2, KLF4, and c-MYC, the combination of which was called Yamanaka factors and sufficient to reprogram somatic cells into induced pluripotent stem cells (iPSCs) (Takahashi et al. 2007). These factors appear to activate a transcriptional network that endows cells with pluripotency (Kim et al. 2008).

We identified a putatively H3K36me3-regulated AS “switch” at the upstream of pluripotency transcriptional network, strongly suggesting a potential mechanistic connection between HMs and ESC fate decision through the AS,

The above integrative analyses showed strong links between three HMs and RNA splicing, revealing a group of epigenetic regulated AS genes involved in cell cycle machinery. *PBX1* was one of the genes that their ASs are positively associated with H3K36me3 (Figures 5A, B). Its protein, PBX1, is a member of the TALE (three-amino acid loop extension) family homeodomain transcription factors (Chang et al. 1997; Merabet and Mann 2016) and well known for its functions in lymphoblastic leukemia (Kamps and Baltimore 1993; Foa et al. 2000; Tesanovic et al. 2014; Melendez et al. 2015) and several cancers (Berge et al. 2006; Qiu et al. 2007; Park et al. 2008; Magnani et al. 2011; Thiaville et al. 2012; Li et al. 2014; Feng et al. 2015; Jung et al. 2015; Magnani et al. 2015; Jung et al. 2016). PBX1 also plays roles in regulating developmental gene expression (Kim et al. 2002) and maintaining stemness and self-renewal (Ficara et al. 2008; Jung et al. 2016), and in cell-cycle progression by repressing the cell-cycle inhibitor CDKN2B to promote transition to S phase (Koss et al. 2012). However, few studies have distinguished its different isoforms, and we suggested that the splicing variation of PBX1 is likely implicated in ESC fate decision through a distinct mechanism.

PBX1 has three isoforms (Pruitt et al. 2014), PBX1a, PBX1b and PBX1c (Figures 5C and S11A). The first two are produced by the alternative splicing of exon 7 (Figure 5A) and attract most of the research attentions of PBX1. PBX1b retains the same DNA-binding homeodomain as PBX1a, but changes 14 amino acids (from 334 to 347) and truncates the remaining 83 amino acids at the C-terminus of PBX1a (Figures 5C and S11A). This C-terminal alteration of PBX1a has been reported to affect its cooperative interactions with HOX partners (Burglin and Ruvkun 1992), which may impart different functions to these two isoforms. We found that the switch from PBX1a to PBX1b

is H3K36me3-associated, and these two isoforms may function competitively during ESC differentiation.

We first observed differential transcript expressions of these two isoforms between the hESCs and differentiated cells. Specifically, *PBX1a* was predominantly transcribed in hESCs, while *PBX1b* was exclusively induced in differentiated cells (Figure 5D) except ME and NPC. The exception of ME is due to its low differentiation level, demonstrated as the most stem-cell-like splicing patterns (Figures 1A-D) and expression profiles of pluripotent factors (Figures 5E-G). Previous studies showed that neuronal development in mice required high expression of *Pbx1a* rather than *Pbx1b* (Paraguison et al. 2007; Golonzhka et al. 2015; Linares et al. 2015), supporting the finding of no *PBX1b* expression in NPCs. Additionally, we did not observe significantly different expression of the total *PBX1* and three other *PBX* family members across cell types (Figure 5H, fold change < 2). Together with previous reports showing that the PBX1a and PBX1b differ in their ability to activate or repress the expression of reporter genes (DiRocco et al. 1997; Asahara et al. 1999), these observations suggest that the isoform switch of *PBX1*, rather than the differential expression of its family members play roles that are more critical for hESC differentiation.

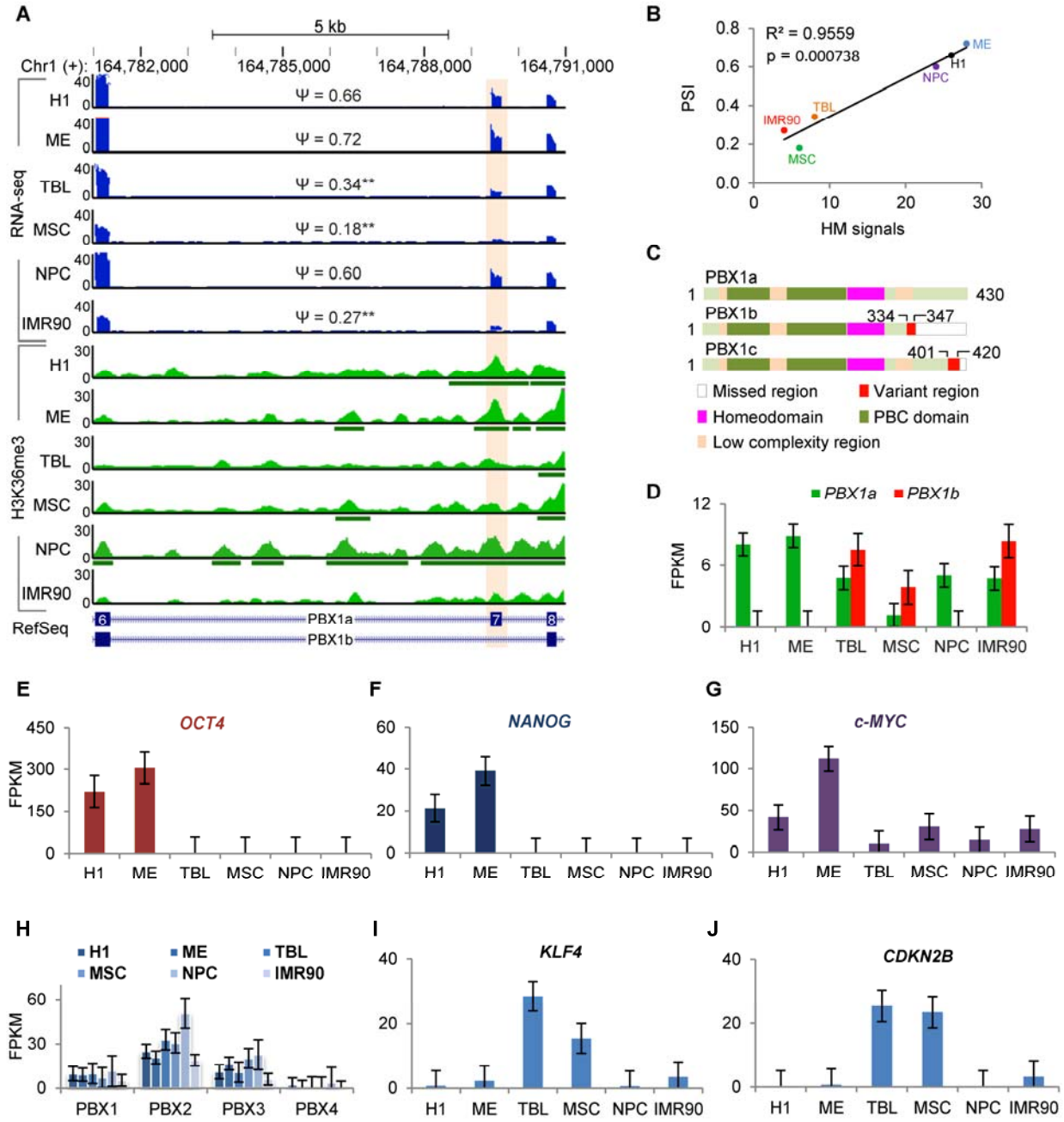

Figure 5 Isoform switch from PBX1a to PBX1b contributes to hESC differentiation. (A) Genome browser view shows the AS events and H3K36me3 signals of PBX1 upon hESC differentiation. The green horizontal bars below the ChIP-seq tracks indicate the narrow peaks called by MACS2. **(B)** The inclusion level for exon 7 of PBX1 is significantly correlated to the H3K36me3 signals over this exon across cell lineages. **(C)** The sequence difference of three protein isoforms of PBX1 and the main functional domains. **(D)** The expression levels of isoforms *PBX1a* and *PBX1b* in hESCs and differentiated cells. The error bars represent standard error of the mean (SEM, n = 2). **(E)** The expression levels of *OCT4* in six studied cell types. **(F)** The expression levels of *NANOG* in six studied cell types. **(G)** The expression levels of *c-MYC* in six studied cell types. **(H)** The expression levels of PBX family genes show no significant difference

across cell types (2-fold change as cutoff). **(I)** The expression levels of *KLF4* in six studied cell types. **(J)** The expression levels of *CDKN2B* in six studied cell types. Also see **Figure S11**.

To further test whether and how PBX1b contributes to stem cell differentiation, we investigated the transcription levels of core pluripotent TFs, including *c-MYC*, *NANOG*, and *OCT4*. Of these TFs, the *NANOG* is activated by direct promoter binding of PBX1 and KLF4, which is essential for stem cell maintenance (Ficara et al. 2008; Chan et al. 2009). In our observations, all these core factors are repressed in differentiated cells where *PBX1b* is highly expressed (Figures 5E-G), even though the *PBX1a* is expressed. As aforementioned, NPC is an exception, in which these stemness factors are also not expressed, although very low *PBX1b* expression is observed. The reason for this exception has been suggested that the expression of stemness factors, especially for *Nanog*, was co-regulated by Pbx1a and Klf4 (Chan et al. 2009), while *KLF4* was not expressed in NPCs (Figure 5I), leading to the repression of stemness regulators in the absence of PBX1b (Figures 5E-G). Additionally, a PBX1a-repressed gene, *CDKN2B* (Koss et al. 2012), is expressed in cells (TBL, MSC, and IMR90) where *PBX1b* is expressed (Figure 5J), regardless of whether or not *PBX1a* is expressed. The presence of CDKN2B will arrest the cell cycle in G1 phase, which is important for cell differentiation (Pauklin and Vallier 2013). It suggests that PBX1b may promote hESC differentiation by abolishing the PBX1a-induced repression of CDKN2B. Collectively, we hypothesize that PBX1b abolishes both the active and repressive functions of PBX1a by competitively binding to the same target genes during hESC differentiation.

The transcript switches between *PBX1a* and *PBX1b* were also observed in an extended dataset of 56 human cell lines/tissues from the Roadmap (Bernstein et al. 2010) and ENCODE (Thomas et al. 2007) projects (supplemental Table S3). Their relative expressions are apparently reversed from stem cells to the differentiated cells and adult tissues (Figure 6A), which confirms that the isoform switch from PBX1a to PBX1b plays critical roles in hESC differentiation. Based on the 56 human cell lines/tissues, we also observed significant negative correlations between expression

of most important pluripotent factors (*NANOG* and *OCT4*) and *PBX1b* (Figures 6B, C), as well as positive correlations between the two factors and *PBX1a* (Figures S11B, C). These results suggest distinct functions of PBX1a and PBX1b, wherein the former promotes the activity of the pluripotent regulatory network, whereas the latter attenuates this activity by competitively binding and regulating the same target genes, such as *NANOG*, since PBX1b retains the same DNA-binding domain as PBX1a. These observations are strongly suggestive that the switch from PBX1a to PBX1b is an additional mechanism by which PBX1 contributes to hESC differentiation via regulating the pluripotency regulatory network.

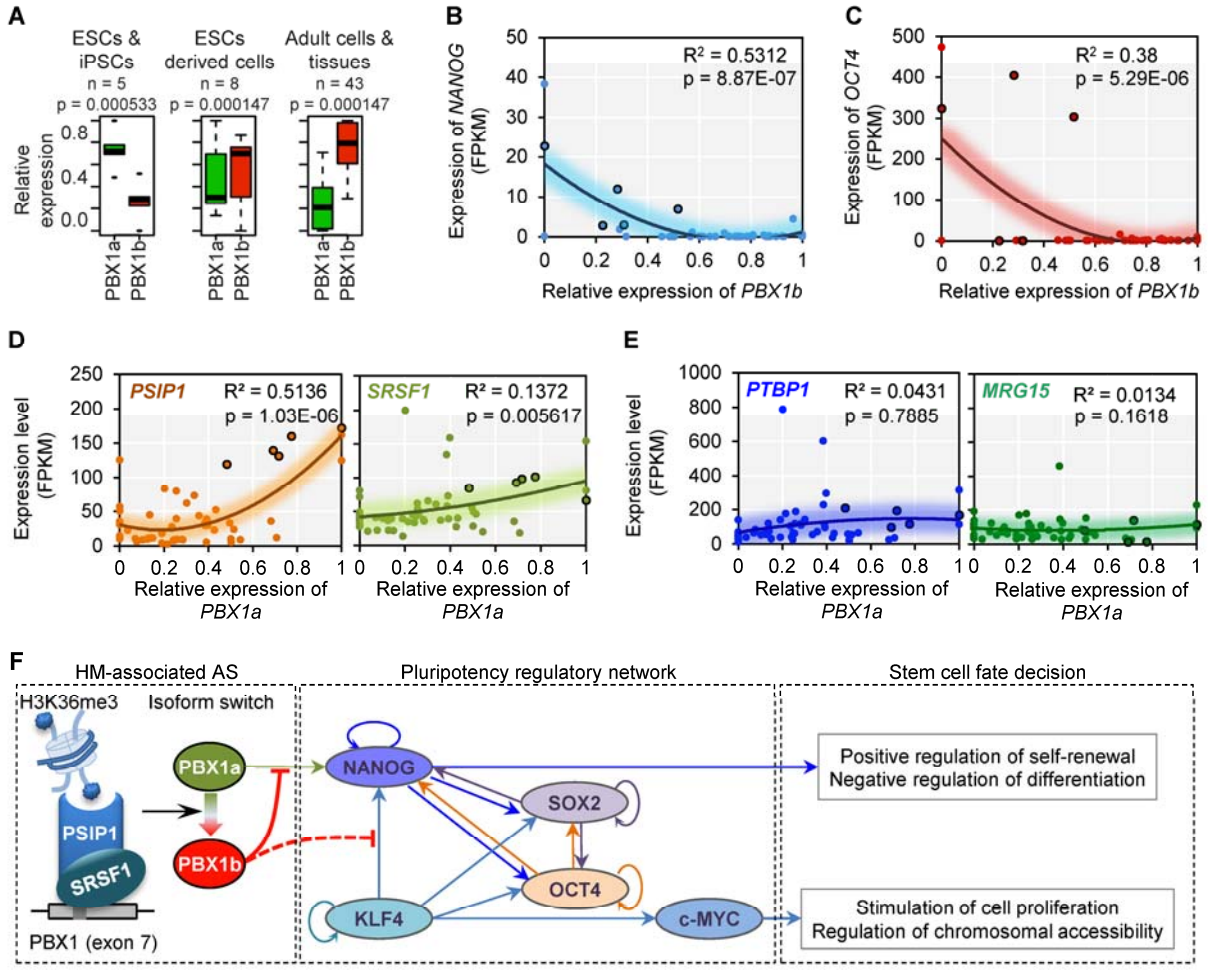

Figure 6 Isoform switch of PBX1 is a novel candidate linking H3K36me3 to ESC fate decision. (A) The relative expressions of PBX1a and PBX1b in 56 cells/tissues, representing the differential expressions of two isoforms in three groups based on their developmental states. **(B)** and **(C)** The expression levels of NANOG and OCT4 genes are negatively correlated with the expression of PBX1b. **(D)** The expression levels of PSIP1 and SRSF1 show significantly

positive correlations with the expression level of PBX1a. **(E)** No significant correlations between the expression levels of PTB/MRG15 and PBX1a are observed. Shapes with black outlines in Figures b-e indicate the ESCs or iPSCs. **(F)** The mechanism by which H3K36me3 is involved in cell fate decision by expanding the transcriptional regulatory network for pluripotency, wherein PBX1b may attenuate the bindings of KLF4 and PBX1a to NANOG by competitively bindings to the same sites. Also see **Figure S11**.

The exon 7 of PBX1 shows significantly positive correlations between its inclusion levels and the surrounding epigenetic signals of H3K36me3 in hESCs and differentiated cells (Figure 5B). It suggests a potential role of H3K36me3 in regulating the isoform switch between *PBX1a* and *PBX1b*. To investigate the regulatory role of H3K36me3, we focused on two previously proved chromatin-adaptor complexes, MRG15/PTBP1 (Luco et al. 2010) and PSIP1/SRSF1 (Pradeepa et al. 2012), which endow dual functions to H3K36me3 in AS regulation (Luco et al. 2011). Based on the 56 cell lines/tissues from the Roadmap/ENCODE projects, we found significantly positive correlations between the expressions of *PBX1a* and *PSIP1/SRSF1* (Figures 6D and S11D), but not with *MRG15/PTBP1* (Figures 6E and S11D). This result suggests that the alternative splicing of *PBX1* is epigenetic regulated by H3K36me3 through the PSIP1/SRSF1 adaptor system. This conclusion is strongly supported by a recent report using the HeLa cell line, which showed that PBX1 is indeed a splicing target of SRSF1 (Anczukow et al. 2015).

Taken together, we strongly suggested that H3K36me3 modulates the alternative splicing of PBX1 via the PSIP1/SRSF1 adaptor system, leading the isoform switch between PBX1a and PBX1b during hESC differentiation. Subsequently, PBX1b competitively binds to *NANOG* and abolishes the bindings of PBX1a and KLF4. This competitive binding attenuates the pluripotency regulatory network to repress self-renewal and consequently facilitate differentiation (Figure 6F). These findings reveal an alternative mechanism by which PBX1 contribute to the ESC fate decision through modulating the pluripotency regulatory network in addition to controlling the cell-cycle progression. We also exemplified a potential mechanism by which HMs contribute to stem

cell fate decision by regulating the alternative splicing of specific genes and adding a new regulatory layer upstream from the canonic pluripotency regulatory network.

Discussion

ESCs provide a vital tool for studying the regulation of early embryonic development and cell-fate decision.(Thomson et al. 1998). In addition to the core pluripotency regulatory TF network, emerging evidences revealed other processes regulating ESC pluripotency and differentiation, including HMs, AS, cell-cycle machinery, and signaling pathways (Gonzales et al. 2015). Here, we connected these previously separate avenues of investigations, beginning with the discovery that 3 of 16 HMs are significantly associated with more than half of AS events upon hESC differentiation. Further findings have profound implications linking HMs, AS regulation, and cell-cycle progression to hESC fate decision. Specifically, HMs orchestrate a subset of AS outcomes that play critical roles in cell-cycle progression via related pathways, such as ATM/ATR-mediated DNA response, and TFs, such as PBX1 (Figure S11E). In this way, HMs, AS regulation, and signaling pathways are converged into the cell-cycle machinery that has been claimed to rule the pluripotency (Vallier 2015).

Although epigenetic signatures, such as HMs, are mainly enriched in promoters and other intergenic regions, it has become increasingly clear that they are also present in the gene body, especially in exon regions. This indicates a potential link between epigenetic regulation and the predominantly co-transcriptional occurrence of AS. Thus far, H3K36me3 (Luco et al. 2010; Pradeepa et al. 2012), H3K4me3 (Sims et al. 2007), H3K9me3 (Piacentini et al. 2009), and H3acetyl (Gunderson and Johnson 2009) have been revealed to regulate AS, either via the chromatin-adapter systems or by altering Pol II elongation rate. Here, we investigated the extent to which the HMs could associate with AS by integrative analyses on both transcriptome and epigenome data during hESC differentiation. We found that three HMs are significantly associated

with about half of AS events. By contrast, a recent report showed that only about 4% of differentially regulated exons among five human cell lines are positively associated with three promoter-like epigenetic signatures, including H3K9ac, H3K27ac and H3K4m3 (Curado et al. 2015). Like that report, we also found a positive association between H3K27ac and a subset of AS events. However, our results differ regarding the other two HMs that we identified as associated with AS.

In our study, H3K36me3 is associated with the most identified HM-associated AS events, either positively or negatively. It is reasonable since H3K36me3 is a mark for actively expressed genomes (Kolasinska-Zwierz et al. 2009), and it has been reported to regulate AS via two chromatin-adaptor systems (Luco et al. 2010; Pradeepa et al. 2012). H4K8ac is associated with the fewest number of AS events in our results. Although little studied, H4K8ac is known to act as a transcriptional activator in both the promoters and transcribed regions (Wang et al. 2008b). Its negative association with AS is supported by the finding that it recruits proteins involved in increasing the Pol II elongation rate (Sim et al. 2004). This suggests that H4K8ac may function in AS regulation by altering the Pol II elongation rate, rather than via the chromatin-adaptor systems. However, further studies are required. Collectively, possible reasons for the different results between this study and others (Curado et al. 2015) could be that (1) we considered differentiation from hESCs to five different cell types, which covered more inter-lineage AS events than the previous report; or (2) different sets of epigenetic signatures were considered, which may lead to relatively biased results in both studies. Obviously, the inclusion of more cell lineages and epigenetic signatures may reduce this bias. Therefore, an extended dataset of 56 cell lines/tissues was included in our study and the observations support our results.

Our study also extended the understanding that HMs contribute to cell fate decision via determining not only what parts of the genome are expressed, but also how they are spliced (Luco et al. 2011). We demonstrated that the strongly associated HMs and AS events jointly affect cell

fate decision in a cell-cycle-dependent manner. The most intriguing discovery is that the HM-associated genes are enriched in G2/M phases and predominantly function in ATM/ATR-mediated DNA response. Evidentially, the ATM/ATR-mediated checkpoint has been recently revealed to attenuate PSD and serves as a gatekeeper of the pluripotency state through the cell cycle (Gonzales et al. 2015). The cell cycle has been considered the hub machinery for cell fate decision (Vallier 2015), since all commitments will go through cell-cycle progression. Our study expanded such machinery by linking the HMs and AS regulation to cell-cycle pathways and TFs, which together, contribute to cell fate decision (Figure 11E).

We also exemplified our hypothesized mechanism by an H3K36me3-associated isoforms switch of PBX1. In addition to its well-known functions in lymphoblastic leukemia (Kamps and Baltimore 1993; Foa et al. 2000; Tesanovic et al. 2014; Melendez et al. 2015) and a number of cancers (Berge et al. 2006; Qiu et al. 2007; Park et al. 2008; Magnani et al. 2011; Thiaville et al. 2012; Li et al. 2014; Feng et al. 2015; Jung et al. 2015; Magnani et al. 2015; Jung et al. 2016), PBX1 was also found to promote hESC self-renewal by corporately binding to the regulatory elements of *NANOG* with KLF4 (Chan et al. 2009). We found that the transcriptions of two isoforms of *PBX1*, *PBX1a* and *PBX1b*, are putatively regulated by H3K36me3 during hESC differentiation. Their protein isoforms competitively bind *NANOG* and the binding of PBX1b will abolish the binding of PBX1a, which further attenuates activity of the core pluripotency regulatory network composed of NAOG, OCT4, and SOX2. The switch from PBX1a to PBX1b is modulated by H3K36me3 via the PSIP1/SRSF1 adapter system (Pradeepa et al. 2012). Our results were also supported by an extended dataset of 56 cell lines/tissues from the Roadmap/ENCODE projects. Collectively, our findings expanded understanding of the core transcriptional network by adding a regulatory layer of HM-associated AS (Figure 6F).

However, a very recent report showed that the switch in Pbx1 isoforms was regulated by Ptbp1 during neuronal differentiation in mice (Linares et al. 2015), indicating a contradiction that the AS

of PBX1 should be negatively regulated by H3K36me3 via the MRG15/PTBP1 (Luco et al. 2010). Our study also included the neuronal lineage and showed that differentiation to NPC is an exception, distinct from other lineages. If NPC is considered separately, the results are consistent with the recent report (Linares et al. 2015) showing that NPCs and mature neurons express increasing levels of PBX1a rather than PBX1b. Another very recent report showed that PBX1 was a splicing target of SRSF1 in the HeLa cell line (Anczukow et al. 2015), which strongly supports our findings. Taken together, these evidences suggest the presence of two parallel mechanisms regulating PBX1 isoforms in embryonic development, of which neuronal differentiation takes the one different from other lineages.

Finally, it is worth noting that both our work and other studies (Curado et al. 2015) reported that HMs cannot explain all AS events identified either during ESC differentiations or based on pairwise comparisons between cell types. Moreover, bidirectional communication between HMs and AS has been widely reported. For instance, the AS can enhance recruitment of H3K36 methyltransferase HYPB/Set2, resulting in significant differences in H3K36me3 around the AS exons (Curado et al. 2015). These findings increased the complexity of defining the cause and effect between HMs and AS. Nevertheless, our findings suggest that at least a subset of AS events are regulated epigenetically, similar to the way that epigenetic states around the transcription start sites define what parts of the genome are expressed. Taken together, AS outcomes may be estimated more precisely by combining splicing *cis*-elements and *trans*-factors (i.e., genetic splicing code) and HMs (i.e., epigenetic splicing code), as an 'extended splicing code'.

Methods

Identification of alternative splicing exons upon hESC differentiation

Aligned BAM files (hg18) for all six cell types (H1, ME, TBL, MSC, NPC, and IMR90) were downloaded from the source provided in reference (Xie et al. 2013a). Two BAM files (replicates)

of each cell type were analyzed using rMATS (version 3.0.9) (Shen et al. 2014) and MISO (version 0.4.6) (Katz et al. 2010). The rMATS was used to identify AS exons based on the differential 'percent splice in' (PSI, Ψ) values between each differentiated cell type and H1 cells (Figure S1B). Only AS exons, which hold the $|\Delta PSI| \geq 0.1$, p-value ≤ 0.01, and FDR $\leq 5\%$, were considered as final hESC differentiation-related AS exons. Finally, two types of AS exons, namely skipped exons (SE) and mutually excluded exons (MXE), which account for most AS events, were used for subsequent analysis. MISO was used to estimate the confidence interval of each PSI and generate Sashimi graphs (Katz et al. 2013) (see Figures 1A, B, 3A, and 5A). To match the ChIP-seq analysis, genomic coordinates of identified AS events were lifted over to hg19 using LiftOver tool downloaded from UCSC. Supplemental Table S1 provides the full list of hESC differentiation-related AS exons for all cell lineages.

ChIP-seq data process and histone modification profiling

ChIP-seq data (aligned BAM files, hg19) were downloaded from Gene Expression Omnibus (GEO, accession ID: GSE16256) (Xie et al. 2013a). This dataset includes the ChIP-seq reads of up to 24 types of HMs for six cell types (H1, ME, TBL, MSC, NPC, and IMR90). Among these, nine histone acetylation modifications (H2AK5ac, H2BK120ac, H2BK5ac, H3K18ac, H3K23ac, H3K27ac, H3K4ac, H3K9ac, and H4K8ac) and seven histone methylation modifications (H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K79me1, and H3K9me3) are available for all six cell types and therefore are used for our analyses.

To generate global differential profiles of histone modification changes between AS exons and constitutive exons upon hESC differentiation, for each MXE and SE AS events, we first randomly selected the constitutive splicing (CS) exons from the same genes, composing a set of CS exons. We then considered the histone modification changes in a ±150bp region flanking the both splice sites of each AS and CS exon, i.e. a 300bp exon-intron boundary region. Each region was 15bp-

binned. Alternatively, for a few cases where the exon or intron is shorter than 150 bps, the entire exonic or intronic region was evenly divided into 10 bins. This scaling allows combining all regions of different lengths to generate uniform profiles of histone modification changes around the splice sites (see Figures 2C and S4). To this end, we calculated the sequencing depth-normalized Δ reads number for each binned region between H1 cells and differentiated cells as follows:

$$\Delta\ reads\ number = \frac{\text{reads number of H1 cells} - \text{reads number of differentiated cells}}{\text{bin size in bps}}$$

Each region is assigned a value representing the average Δ reads number between H1 cells and differentiated cells for each histone modification. We also compared histone modification profiles between the inclusion-gain and –loss exons (Figures 3B and S5B) using the same strategy. The statistical results (Figures 2C and S5B) are presented as the p-values based on Mann-Whitney-Wilcoxon tests (using the R package).

To quantitatively link HMs to alternative splicing upon hESC differentiation, the ChIP-seq data were further processed by narrow peak calling. For each histone ChIP-seq dataset, the MACS v2.0.10 peak caller (https://github.com/taoliu/MACS/) (Zhang et al. 2008; Feng et al. 2012) was used to compare ChIP-seq signal to a corresponding whole cell extract (WCE) sequenced control to identify narrow regions of enrichment (narrow peak) that pass a Poisson p-value threshold of 0.01. All other parameters (options) were set as default. We then compared the histone modification signals between H1 cells and differentiated cells. We defined the 'differential HM signals (ΔHMs)' as the difference of the normalized peak signals (i.e., the heights of the narrow peaks) between H1 and the differentiated cells. The normalization was performed based on the distance (in kb) between the peak summits and 3' splice sites (upstream) of the AS exons (Figure S5A). Since there is no evidence showing that distal HMs could regulate the AS, we considered only peaks with at least one bp overlapping on either side of the AS exon. For exons without overlapping peaks, peak signals of these exons were set to zero. For the exons there are more

than one overlapping peaks, the peak signals of these exons were set to the greater ones. For MXE events, only the upstream AS exons were considered due to their exclusiveness in inclusion level between these two exons, i.e., the sum of the PSIs for two exons of an MXE event is always one.

Association studies and k-means clustering

To quantitatively estimate associations between HMs and AS, we first used three independent correlation tests, including Pearson correlation (PC), multiple linear regression (MLL), and logistic regression (LLR), to test global correlations between AS events and each of 16 HMs based on differential inclusions (ΔPSIs) and differential HM signals (ΔHMs). PC was performed using the R package (stats, cor.test(),method = 'pearson'). MLL and LLR were calculated using Weka 3 (Hall et al. 2009), wherein the ΔHMs are independent variables and ΔPSIs are response variables. The results show that only some HMs correlate with AS, and most correlations are weak (Figure S5C). HMs that have significant correlations ($p \leq 0.05$) with AS were used for further clustering analysis, through which we identified six subsets of HM-associated AS events (Supplemental Table S2).

K-means clustering was performed separately on inclusion-gain and inclusion-loss AS events of MXE and SE, based on the selected HM signatures (Figure S5C). *K* was set to six for all clustering processes (Figure S5D), which produced the minimal RSME (root mean square error) for most cases based on a series of clustering with *k* varying from two to eight (data not shown). Then the two clusters that generate mirror patterns, of which one from inclusion-gain events and one from inclusion-loss events, were combined to be considered as a subset of HM-associated AS events (Figure S6). Finally, we identified six subsets of HM-associated AS events displaying significantly positive or negative correlations with three HMs.

Gene expression quantification

For each cell type, two aligned BAM files (replicates) were used to estimate the expression level of genes using Cufflinks (Trapnell et al. 2012). Default parameters (options) were used. The expression level of each gene was presented as FPKM for each cell type. Differentially expressed genes (DEGs) were defined as those genes whose fold changes between the differentiated cells and hESCs are greater than 2. Specifically for DEG analysis of SF genes, we collected a list of 47 ubiquitously expressed SFs with “RNA splicing” in their GO annotations from AmiGO 2 (Carbon et al. 2009). The enrichment significances in Figure S1D and S3H are shown as the p-values based on hypergeometric tests, using the DEGs of all expressed genes (with the FPKM ≥ 1 in at least hESCs or one of differentiated cell types) as background. We found that the AS genes are generally not differentially expressed between hESCs and differentiated cells, indicating that they function in hESC differentiation via isoform switches rather than expression changes. Few SF genes show differential expression between hESCs and differentiated cells, indicating the existence of epigenetic control of AS.

Genome annotations

Since the RNA-seq reads (BAM) files and ChIP-seq read (BAM) files downloaded from the public sources are mapped to different human genome assembles, NCBI36/hg18 (Mar. 2006) and GRCh37/hg19 (Feb. 2009), respectively, we downloaded two version of gene annotations (in GTF formats) from the UCSC Table Browser (Karolchik et al. 2004). The hg18 GTF file was used for rMATS and MISO to identify AS during the differentiation from H1 ESCs to five differentiated cells. The hg19 GTF file was used to define the genome coordinates of AS exons and further for ChIP-seq profile analysis (Figures 2C, 3B, S4, and S5A). We compared exonic and intronic lengths based on hg18 annotation (Figure S2).

Gene ontology enrichment analysis

The Gene ontology (GO) enrichment analysis was performed using ToppGene (Chen et al. 2009) by searching the HGNC Symbol database under default parameters (p-value method: probability density function). Overrepresented GO terms for the GO domain 'biological process' were used to generate data shown in Figures 4C-E and Figures S8A-C using either the FDR (0.05) adjusted p-value or the enriched gene numbers (Figure S8A).

Canonic pathway enrichment analysis

Both the HM-associated (n = 282) and –unassociated (150) AS genes from the top enriched GO term (GO:0007049) were used to perform canonic pathway enrichment (Figures 4F and S8D) analysis through Ingenuity Pathway Analysis (IPA, http://www.ingenuity.com/products/ipa).

Stemness signature and TF binding enrichment analysis

StemChecker (Pinto et al. 2015), a web-based tool to discover and explore stemness signatures in gene sets, was used to calculate enrichment of AS genes in stemness signatures. Significances was tested (hypergeometric test) using the gene set from human (*Homo sapiens*) of this database as the background. For all AS genes (n = 3,865), 2,979 genes were found in this data set. 1,395 of 2,979 genes were found in the stemness signature gene set, most of which (n = 813) are ESC signature genes. Figure S3A shows the enrichment significance as the -log10 p-values. For HM-associated AS genes (n = 2,125), 1,992 genes were found in this data set. 966 of 1,992 genes were found in the stemness signature gene set, most of which (n = 562) are ESC signature genes. For HM-unassociated genes (n = 1,057), 987 genes were found in this data set. 429 of 987 genes were found in the stemness signature gene set, most of which (n = 251) are ESC signature genes. The significances (-log10 p-values) are shown in Figure 4A.

FunRich (v2.1.2), a stand-alone functional enrichment analysis tool, was used for TF binding enrichment analysis. The top six enriched TFs of HM-associated and –unassociated AS genes are presented and shown as the proportion of enriched AS genes. It shows that HM-associated AS genes are more likely to be regulated by TFs involved in cell differentiation and development, while the HM-unassociated AS genes are more likely to be regulated by TFs involved in cell proliferation and renewal (Figure 4B).

Roadmap and ENCODE data analysis

The sources of RNA-seq data for 56 cell lines/tissues from Roadmap/ENCODE projects were listed in supplemental Table S3. The RNA-seq data (BAM files) were used to calculate the PSI of the exon 7 for PBX1 in each cell line/tissue and to estimate the expression levels of all gene (FPKM), based on aforementioned strategies. The relative expression levels of PBX1a and PBX1b shown in Figure 6 and Figure S11 were calculated as the individual PFKM value of each divided by their total FPKM values.

Statistical analyses and tests

Levels of significance were calculated with Mann-Whitney-Wilcoxon test for Figures 2C, 2D, 3B, S4, and S5B, with Fisher's exact test for Figures 1D, 2D, and S5B, with Student's t test for Figures 2A-D and S7A, and with Hypergeometric test for Figures S1D and S3H. Levels of correlation significance were calculated with Pearson correlation (PC), multiple linear regression (MLR), and logistic regression (LLR) for Figures 3C, 3E, and S5C. MLR and LLR were performed using Weka 3 (Hall et al. 2009), whereas all other tests were performed using R packages.

Data availability

The RNA-seq data of H1 and five other differentiated data (two replicates for each, BAM files) are available at http://epigenome.ucsd.edu/differentiation/download.html. All ChIP-seq data of 16

HMs are available in Gene Expression Omnibus (GEO) under accession number GSE16256. Both RNA-seq and ChIP-seq data of 56 cell lines/tissues from the Roadmap/ENCODE projects are available in their official website. Supplemental Table 3 provides the details on accession to these data.

Abbreviations

ESC: embryonic stem cell; AS: alternative (pre-mRNA) splicing; CS: constitutive splicing; HM: histone modification; TF: transcription factor; SF: splicing factor; MXE: mutually excluded exon; SE: skipped exon; PSI: percent splice in; H2AK5ac: H2A acetylated lysine 5; H2BK120ac: H2B acetylated lysine120; H2BK5ac: H2B acetylated lysine 5; H3K18ac: H3 acetylated lysine 18; H3K23ac: H3 acetylated lysine 23; H3K27ac: H3 acetylated lysine 27; H3K4ac: H3 acetylated lysine 4; H3K9ac: H3 acetylated lysine 9; H4K8ac: H4 acetylated lysine 8; H3K27me3: H3 tri-methylated lysine 27; H3K36me3: H3 tri-methylated lysine 36; H3K4me1: H3 mono-methylated lysine 4; H3K4me2: H3 di-methylated lysine 4; H3K4me3: H3 tri-methylated lysine 4; H3K79me1: H3 mono-methylated lysine 79; H3K9me3: H3 tri-methylated lysine 9.

Acknowledgements

We acknowledge the editorial assistance of Karen Klein, MA, in the Wake Forest Clinical and Translational Science Institute (UL1 TR001420; PI: McClain). This work was funded by the National Institutes of Health (NIH) [1U01CA166886 and AR069395].

Author's Contributions

All authors conceived the study. YX carried out the data analysis and interpreted the results. YX wrote the first draft of the manuscript and all authors revised it. All authors read and approved the final version of the manuscript.

Supplemental Materials

Supplemental file 1: Supplementary Figures. This PDF file contain supplementary Figures S1-S11.

Supplemental File 2: Table S1. This Excel file contains all AS events identified during hESC differentiation. Each sheet provide the AS events of MXE or SE for each cell lineage.

Supplemental File 3: Table S2. This Excel file contains the identified HM-associated AS exons. Six subsets of AS exons are listed separately in six sheets.

Supplemental File 4: Table S3. This Excel file contains the 56 cell lines/tissues from Roadmap/ENCODE projects and their sources for RNA-seq data accession.

Disclosure Declaration

Competing financial interests - the authors declare no competing financial interests.

References

Aaronson Y, Meshorer E. 2013. STEM CELLS Regulation by alternative splicing. *Nature* **498**(7453): 176-177.

Anczukow O, Akerman M, Clery A, Wu J, Shen C, Shirole NH, Raimer A, Sun SY, Jensen MA, Hua YM et al. 2015. SRSF1-Regulated Alternative Splicing in Breast Cancer. *Molecular cell* **60**(1): 105-117.

Asahara H, Dutta S, Kao HY, Evans RM, Montminy M. 1999. Pbx-hox heterodimers recruit coactivator-corepressor complexes in an isoform-specific manner. *Mol Cell Biol* **19**(12): 8219-8225.

Barash Y, Calarco JA, Gao WJ, Pan Q, Wang XC, Shai O, Blencowe BJ, Frey BJ. 2010. Deciphering the splicing code. *Nature* **465**(7294): 53-59.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**(4): 823-837.

Berge V, Ramberg H, Eide T, Svindland A, Tasken KA. 2006. Expression of Pbx1 and UGT2B7 in prostate cancer cells. *Eur Urol Suppl* **5**(14): 794-794.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR et al. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**(10): 1045-1048.

Boland MJ, Nazor KL, Loring JF. 2014. Epigenetic Regulation of Pluripotency and Differentiation. *Circ Res* **115**(2): 311-324.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JR, Guenther MG, Kumar RM, Murray HL, Jenner RG et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**(6): 947-956.

Burglin TR, Ruvkun G. 1992. New Motif in Pbx Genes. *Nat Genet* **1**(5): 319-320.

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, Hub A, Grp WPW. 2009. AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**(2): 288-289.

Chan KKK, Zhang J, Chia NY, Chan YS, Sim HS, Tan KS, Oh SKW, Ng HH, Choo ABH. 2009. KLF4 and PBX1 Directly Regulate NANOG Expression in Human Embryonic Stem Cells. *Stem Cells* **27**(9): 2114-2125.

Chang CP, Jacobs Y, Nakamura T, Jenkins NA, Copeland NG, Cleary ML. 1997. Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. *Mol Cell Biol* **17**(10): 5679-5687.

Chen J, Bardes EE, Aronow BJ, Jegga AG. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**: W305-W311.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang WW, Jiang JM et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**(6): 1106-1117.

Clevers H, Loh KM, Nusse R. 2014. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* **346**(6205): 54-+.

Coronado D, Godet M, Bourillot PY, Tapponnier Y, Bernat A, Petit M, Afanassieff M, Markossian S, Malashicheva A, Iacone R et al. 2013. A short G1 phase is an intrinsic determinant of naive embryonic stem cell pluripotency. *Stem Cell Res* **10**(1): 118-131.

Curado J, Iannone C, Tilgner H, Valcarcel J, Guigo R. 2015. Promoter-like epigenetic signatures in exons displaying cell type-specific splicing. *Genome Biol* **16**: 236.

Dalton S. 2015. Linking the Cell Cycle to Cell Fate Decisions. *Trends Cell Biol* **25**(10): 592-600.

DiRocco G, Mavilio F, Zappavigna V. 1997. Functional dissection of a transcriptionally active, target-specific Hox-Pbx complex. *Embo J* **16**(12): 3644-3654.

Fatica A, Bozzoni I. 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* **15**(1): 7-21.

Feng JX, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**(9): 1728-1740.

Feng Y, Li L, Zhang X, Zhang Y, Liang Y, Lv J, Fan Z, Guo J, Hong T, Ji B et al. 2015. Hematopoietic pre-B cell leukemia transcription factor interacting protein is overexpressed in gastric cancer and promotes gastric cancer cell proliferation, migration, and invasion. *Cancer science* **106**(10): 1313-1322.

Ficara F, Murphy MJ, Lin M, Cleary ML. 2008. Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. *Cell Stem Cell* **2**(5): 484-496.

Foa R, Vitale A, Cuneo A, Mecucci C, Mancini H, Cimino G, Cutinelli F, Saglio G, Torelli G, Quarta G et al. 2000. E2A-PBX1 fusion in adult acute lymphoblastic leukemia (ALL) with t(1;19) translocation: Biologic and clinical features. *Blood* **96**(11): 189b-189b.

Gabut M, Samavarchi-Tehrani P, Wang XC, Slobodeniuc V, O'Hanlon D, Sung HK, Alvarez M, Talukder S, Pan Q, Mazzoni EO et al. 2011. An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming. *Cell* **147**(1): 132-146.

Gerstein MB Lu ZJ Van Nostrand EL Cheng C Arshinoff BI Liu T Yip KY Robilotto R Rechtsteiner A Ikegami K et al. 2010. Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. *Science* **330**(6012): 1775-1787.

Gifford CA, Ziller MJ, Gu HC, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R et al. 2013. Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. *Cell* **153**(5): 1149-1163.

Golonzhka O, Nord A, Tang PL, Lindtner S, Ypsilanti AR, Ferretti E, Visel A, Selleri L, Rubenstein JL. 2015. Pbx Regulates Patterning of the Cerebral Cortex in Progenitors and Postmitotic Neurons. *Neuron* **88**(6): 1192-1207.

Gonzales KAU, Liang H, Lim YS, Chan YS, Yeo JC, Tan CP, Gao B, Le BL, Tan ZY, Low KY et al. 2015. Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell* **162**(3): 564-579.

Graveley BR. 2011. Splicing up Pluripotency. *Cell* **147**(1): 22-24.

Graveley BR, Brooks AN, Carlson J, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW et al. 2011. The developmental transcriptome of Drosophila melanogaster. *Nature* **471**(7339): 473-479.

Gunderson FQ, Johnson TL. 2009. Acetylation by the Transcriptional Coactivator Gcn5 Plays a Novel Role in Co-Transcriptional Spliceosome Assembly. *Plos Genet* **5**(10).

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* **11**(1): 10-18.

Han H, Irimia M, Ross PJ, Sung HK, Alipanahi B, David L, Golipour A, Gabut M, Michael IP, Nachman EN et al. 2013. MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* **498**(7453): 241-+.

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S et al. 2010. Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. *Cell Stem Cell* **6**(5): 479-491.

Itoh F, Watabe T, Miyazono K. 2014. Roles of TGF-beta family signals in the fate determination of pluripotent stem cells. *Semin Cell Dev Biol* **32**: 98-106.

Jung JG, Kim TH, Gerry E, Kuan JC, Ayhan A, Davidson B, Shih IM, Wang TL. 2016. PBX1, a transcriptional regulator, promotes stemness and chemoresistance in ovarian cancer. *Clin Cancer Res* **22**.

Jung JG, Park JT, Stoeck A, Hussain T, Wu RC, Shih IM, Wang TL. 2015. Targeting PBX1 signaling to sensitize carboplatin cytotoxicity in ovarian cancer. *Clin Cancer Res* **21**.

Kalsotra A, Cooper TA. 2011. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* **12**(10): 715-729.

Kamps MP, Baltimore D. 1993. E2a-Pbx1, the T(1, 19) Translocation Protein of Human Pre-B-Cell Acute Lymphocytic-Leukemia, Causes Acute Myeloid-Leukemia in Mice. *Mol Cell Biol* **13**(1): 351-357.

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**: D493-D496.

Katz Y, Wang ET, Airoldi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**(12): 1009-U1101.

Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, Mesirov JP, Airoldi EM, Burge CB. 2013. Sashimi plots: quantitative visualization of RNA sequencing read alignments. *arXiv preprint arXiv:13063466*.

Kim JW, Chu JL, Shen XH, Wang JL, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**(6): 1049-1061.

Kim SK, Selleri L, Lee JS, Zhang AY, Gu XY, Jacobs Y, Cleary ML. 2002. Pbx1 inactivation disrupts pancreas development and in Ipf1-deficient mice promotes diabetes mellitus. *Nat Genet* **30**(4): 430-435.

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* **41**(3): 376-381.

Kornblihtt AR, De la Mata M, Fededa JP, Munoz MJ, Nogues G. 2004. Multiple links between transcription and splicing. *Rna-a Publication of the Rna Society* **10**(10): 1489-1498.

Kornblihtt AR, Schor IE, Allo M, Blencowe BJ. 2009. When chromatin meets splicing. *Nat Struct Mol Biol* **16**(9): 902-903.

Koss M, Bolze A, Brendolan A, Saggese M, Capellini TD, Bojilova E, Boisson B, Prall OWJ, Elliott DA, Solloway M et al. 2012. Congenital Asplenia in Mice and Humans with Mutations in a Pbx/Nkx2-5/p15 Module. *Dev Cell* **22**(5): 913-926.

Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**(4): 693-705.

Li WH, Huang K, Guo HZ, Cui GH, Zhao S. 2014. Inhibition of non-small-cell lung cancer cell proliferation by Pbx1. *Chinese J Cancer Res* **26**(5): 573-578.

Lian XJ, Zhang JH, Azarin SM, Zhu KX, Hazeltine LB, Bao XP, Hsiao C, Kamp TJ, Palecek SP. 2013. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions. *Nat Protoc* **8**(1): 162-175.

Linares AJ, Lin CH, Damianov A, Adams KL, Novitch BG, Black DL. 2015. The splicing regulator PTBP1 controls the activity of the transcription factor Pbx1 during neuronal differentiation. *Elife* **4**.

Lu XY, Goke J, Sachs F, Jacques PE, Liang HQ, Feng B, Bourque G, Bubulya PA, Ng HH. 2013. SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nat Cell Biol* **15**(10): 1141-U1145.

Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. 2011. Epigenetics in Alternative Pre-mRNA Splicing. *Cell* **144**(1): 16-26.

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* **327**(5968): 996-1000.

Magen A, Ast G. 2005. The importance of being divisible by three in alternative splicing. *Nucleic Acids Res* **33**(17): 5574-5582.

Magnani L, Ballantyne EB, Zhang X, Lupien M. 2011. PBX1 genomic pioneer function drives ERalpha signaling underlying progression in breast cancer. *Plos Genet* **7**(11): e1002368.

Magnani L, Patten DK, Nguyen VTM, Hong SP, Steel JH, Patel N, Lombardo Y, Faronato M, Gomes AR, Woodley L et al. 2015. The pioneer factor PBX1 is a novel driver of metastatic progression in ER'-positive breast cancer. *Oncotarget* **6**(26): 21878-21891.

Mayshar Y, Rom E, Chumakov I, Kronman A, Yayon A, Benvenisty N. 2008. Fibroblast growth factor 4 and its novel splice isoform have opposing effects on the maintenance of human embryonic stem cell self-renewal. *Stem Cells* **26**(3): 767-774.

Melendez CLC, Rosales L, Herrera M, Granados L, Tinti D, Villagran S, Jo W, Villegas M, Antillon F, Torselli S et al. 2015. Frequency of the ETV6-RUNX1, BCR-ABL1, TCF3-PBX1 and MLL-AFF1 fusion genes in Guatemalan Pediatric Acute Lymphoblastic Leukemia Patients and Its Ethnic Associations. *Cl Lymph Myelom Leuk* **15**: S170-S170.

Merabet S, Mann RS. 2016. To Be Specific or Not: The Critical Relationship Between Hox And TALE Proteins. *Trends Genet* **32**(6): 334-347.

Mikkelsen TS, Ku MC, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153): 553-U552.

Ogaki S, Shiraki N, Kume K, Kume S. 2013. Wnt and Notch Signals Guide Embryonic Stem Cell Differentiation into the Intestinal Lineages. *Stem Cells* **31**(6): 1086-1096.

Pan Q, Shai O, Lee LJ, Frey J, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**(12): 1413-1415.

Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li HR, Qiu JS, Huang J, Yeo GW, Ares M, Fu XD. 2013. Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing. *Molecular cell* **50**(2): 223-235.

Paraguison RC, Yamamoto K, Yamamoto K, Matsumoto H, Sasaki T, Kato N, Nanba E. 2007. Enhanced autophagic cell death in expanded polyhistidine variants of HOXA1 reduces PBX1-coupled transcriptional activity and inhibits neuronal differentiation. *J Neurosci Res* **85**(3): 479-487.

Park JT, Shih Ie M, Wang TL. 2008. Identification of Pbx1, a potential oncogene, as a Notch3 target gene in ovarian cancer. *Cancer Res* **68**(21): 8852-8860.
Pauklin S, Vallier L. 2013. The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity. *Cell* **155**(1): 135-147.
Piacentini L, Fanti L, Negri R, Del Vescovo V, Fatica A, Altieri F, Pimpinelli S. 2009. Heterochromatin Protein 1 (HP1a) Positively Regulates Euchromatic Gene Expression through RNA Transcript Association and Interaction with hnRNPs in Drosophila. *Plos Genet* **5**(10).
Pinto JP, Kalathur RK, Oliveira DV, Barata T, Machado RSR, Machado S, Pacheco-Leyva I, Duarte I, Futschik ME. 2015. StemChecker: a web-based tool to discover and explore stemness signatures in gene sets. *Nucleic Acids Res* **43**(W1): W72-W77.
Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA. 2012. Psip1/Ledgf p52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing. *Plos Genet* **8**(5).
Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM et al. 2014. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42**(D1): D756-D763.
Qiu Y, Tomita Y, Zhang B, Nakamichi I, Morii E, Aozasa K. 2007. Pre-B-cell leukemia transcription factor 1 regulates expression of valosin-containing protein, a gene involved in cancer growth. *The American journal of pathology* **170**(1): 152-159.
Ramani AK, Calarco JA, Pan Q, Mavandadi S, Wang Y, Nelson AC, Lee LJ, Morris Q, Blencowe BJ, Zhen M et al. 2011. Genome-wide analysis of alternative splicing in Caenorhabditis elegans. *Genome Res* **21**(2): 342-348.
Rao S, Zhen S, Roumiantsev S, McDonald LT, Yuan GC, Orkin SH. 2010. Differential Roles of Sall4 Isoforms in Embryonic Stem Cell Pluripotency. *Mol Cell Biol* **30**(22): 5364-5380.
Sakaki-Yumoto M, Katsuno Y, Derynck R. 2013. TGF-beta family signaling in stem cells. *Bba-Gen Subjects* **1830**(2): 2280-2296.
Salomonis N, Nelson B, Vranizan K, Pico AR, Hanspers K, Kuchinsky A, Ta L, Mercola M, Conklin BR. 2009. Alternative Splicing in the Differentiation of Human Embryonic Stem Cells into Cardiac Precursors. *Plos Comput Biol* **5**(11).
Salomonis N, Schlieve CR, Pereira L, Wahlquist C, Colas A, Zambon AC, Vranizan K, Spindler MJ, Pico AR, Cline MS et al. 2010. Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **107**(23): 10514-10519.
Schwartz S, Meshorer E, Ast G. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* **16**(9): 990-995.
Sharov AA, Ko MSH. 2007. Human ES cell profiling broadens the reach of bivalent domains. *Cell Stem Cell* **1**(3): 237-238.
Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America* **111**(51): E5593-5601.
Shenoy A, Blelloch RH. 2014. Regulation of microRNA function in somatic stem cell proliferation and differentiation. *Nat Rev Mol Cell Bio* **15**(9): 565-576.
Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, Wray J, Yamanaka S, Chambers I, Smith A. 2009. Nanog Is the Gateway to the Pluripotent Ground State. *Cell* **138**(4): 722-737.
Sim RJ, Belotserkovskaya R, Reinberg D. 2004. Elongation by RNA polymerase II: the short and long of it. *Gene Dev* **18**(20): 2437-2468.
Sims RJ, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D. 2007. Recognition of trimethylated histone h3 lysine 4 facilitates the

recruitment of transcription postinitiation factors and pre-mRNA splicing. *Molecular cell* **28**(4): 665-676.
Singh AM, Sun YH, Li L, Zhang WJ, Wu TM, Zhao SY, Qin ZH, Dalton S. 2015. Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. *Stem Cell Rep* **5**(3): 323-336.
Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**(5): 861-872.
Tay Y, Zhang JQ, Thomson AM, Lim B, Rigoutsos I. 2008. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**(7216): 1124-U1112.
Tesanovic T, Badura S, Oellerich T, Doering C, Ruthardt M, Ottmann OG. 2014. Mechanisms of Antileukemic Activity of the Multikinase Inhibitors Dasatinib and Ponatinib in Acute Lymphoblastic Leukemia (ALL) Harboring the E2A-PBX1 Fusion Gene. *Blood* **124**(21).
Thiaville MM, Stoeck A, Chen L, Wu RC, Magnani L, Oidtman J, Shih IM, Lupien M, Wang TL. 2012. Identification of PBX1 Target Genes in Cancer Cells by Global Mapping of PBX1 Binding Sites. *Plos One* **7**(5).
Thomas DJ, Rosenbloom KR, Clawson H, Hinrichs AS, Trumbower H, Raney BJ, Karolchik D, Barber GP, Harte RA, Hillman-Jackson J et al. 2007. The ENCODE project at UC Santa Cruz. *Nucleic Acids Res* **35**: D663-D667.
Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* **282**(5391): 1145-1147.
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**(3): 562-578.
Vallier L. 2015. Cell Cycle Rules Pluripotency. *Cell Stem Cell* **17**(2): 131-132.
Wang ET, Sandberg R, Luo SJ, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008a. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**(7221): 470-476.
Wang GS, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8**(10): 749-761.
Wang ZB, Zang CZ, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui KR, Roh TY, Peng WQ, Zhang MQ et al. 2008b. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**(7): 897-903.
Watabe T, Miyazono K. 2009. Roles of TGF-beta family signaling in stem cell renewal and differentiation. *Cell Res* **19**(1): 103-115.
Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D et al. 2013a. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**(5): 1134-1148.
Xie W, Schultz MD, Lister R, Hou ZG, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D et al. 2013b. Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells. *Cell* **153**(5): 1134-1148.
Xue YC, Zhou Y, Wu TB, Zhu T, Ji X, Kwon YS, Zhang C, Yeo G, Black DL, Sun H et al. 2009. Genome-wide Analysis of PTB-RNA Interactions Reveals a Strategy Used by the General Splicing Repressor to Modulate Exon Inclusion or Skipping. *Molecular cell* **36**(6): 996-1006.
Yeo GW, Xu XD, Liang TY, Muotri AR, Carson CT, Coufal NG, Gage FH. 2007. Alternative splicing events identified in human embryonic stem cells and neural progenitors. *Plos Comput Biol* **3**(10): 1951-1967.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**(9).
Zheng CL, Fu XD, Gribskov M. 2005. Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *Rna* **11**(12): 1777-1787.