Alternative splicing links histone modifications to stem cell fate decision

Yungang Xu[1,2,†], Weiling Zhao[1,2,†], Scott D. Olson[3], Karthik S. Prabhakara[3], Xiaobo Zhou[1,2,*]

[1]Center for Computational Systems Medicine, School of Biomedical Informatics, The 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]Department of Pediatric Surgery, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

[†]These two authors contributed equally to this work.

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

Email addresses:

YX: yungangx.xu@uth.tmc.edu

WZ: weiling.zhao@uth.tmc.edu

SO: scott.d.olson@uth.tmc.edu

KP: Karthik.Prabhakara@uth.tmc.edu

XZ: xiaobo.zhou@uth.tmc.edu

Abstract

Background: Understanding the embryonic stem cell (ESC) fate decision between self-renewal and proper differentiation is important for developmental biology and regenerative medicine. Attention has focused on mechanisms involving histone modifications, alternative pre-mRNA splicing, and cell-cycle progression. However, their intricate interrelations and joint contributions to ESC fate decision remain unclear.

Results: We analyze the transcriptomes and epigenomes of human ESC and five types of differentiated cells. We identify thousands of alternatively spliced exons and reveal their development and lineage-dependent characterizations. Several histone modifications show dynamic changes in alternatively spliced exons and three are strongly associated with 52.8% of alternative splicing events upon hESC differentiation. The histone modification-associated alternatively spliced genes predominantly function in G2/M phases and ATM/ATR-mediated DNA damage response pathway for cell differentiation, whereas other alternatively spliced genes are enriched in the G1 phase and pathways for self-renewal. These results imply a potential epigenetic mechanism by which some histone modifications contribute to ESC fate decision through the regulation of alternative splicing in specific pathways and cell-cycle genes. Supported by experimental validations and extended dataset from Roadmap/ENCODE projects, we exemplify this mechanism by a cell cycle-related transcription factor, PBX1, which regulates the pluripotency regulatory network by binding to NANOG. We suggest that the isoform switch from PBX1a to PBX1b links H3K36me3 to hESC fate determination through the PSIP1/SRSF1 adaptor, which results in the exon skipping of PBX1.

Conclusion: We reveal the mechanism by which alternative splicing links histone modifications to stem cell fate decision.

Keywords

Alternative splicing, histone modification, cell fate decision, cell cycle machinery, embryonic stem cell

Background

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 [1]. The past few years have witnessed remarkable progress in understanding the ESC fate decision, i.e. either pluripotency maintenance (self-renewal) or proper differentiation [2]. The underlying mechanisms have been largely expanded from the core pluripotent transcription factors (TFs) [3], signaling pathways [4-10], specific microRNAs [11, 12], and long non-coding RNAs [13] to alternative pre-mRNA splicing (AS) [14, 15], histone modifications (HMs) [16-20], and cell-cycle machinery [21]. These emerging mechanisms suggest their intricate interrelations and potential joint contributions to ESC pluripotency and differentiation, which, however, remain unknown.

Alternative splicing (AS) is one of the most important pre-mRNA processing to increase the diversity of transcriptome and proteome in tissue- and development-dependent manners [22]. The estimates based on RNA-seq revealed that up to 94%, 60% and 25% of genes in human, *Drosophila melanogaster* and *Caenorhabditis elegans*, respectively, undergo AS [22-26]. AS also provides a powerful mechanism to control the developmental decision in ESCs [27-29]. Specific isoforms are necessary to maintain both the identity and activity of stem cells and switching to different isoforms ensures proper differentiation [30]. Especially, the AS of TFs plays major roles in ESC fate determination, such as *FGF4* [31] and *FOXP1* [14] for hESC, *Tcf3* [15] and *Sall4* [32] for mouse ESCs (mESCs). Understanding the precise regulations on AS would contribute to the elucidation of ESC fate decision and has attracted extensive efforts [33]. For

many years, studies aiming to shed light on this process focused on the RNA level, characterizing the manner by which splicing factors (SFs) and auxiliary proteins interact with splicing signals, thereby enabling, facilitating and regulating RNA splicing. These *cis*-acting RNA elements and *trans*-acting splicing factors have been assembled into splicing code [34], revealing a number of AS regulators critical for ESC differentiation, such as MBNL [35] and SON [29]. However, these genetic controls are far from sufficient to explain the faithful regulation of AS [36]. Especially in some cases that tissue-specific AS patterns exist despite the identity in sequences and ubiquitously expression of involved SFs [37, 38], indicating additional regulatory layers leading to specific AS patterns. As expected, we are increasingly aware 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 [39]. 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 [39].

HMs have long been thought to play crucial roles in ESC 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 [40]. 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 [41]. 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 [42]. 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 [43]. A number of HMs, such as H3K4me3 [44], H3K9me3 [45], H3K36me3 [46, 47], and hyperacetylation of H3 and H4 [48-52], have been proven to regulate AS by either directly recruiting chromatin-associated factors and SFs or indirectly modulating transcriptional elongation rate [39]. Together, these studies reveal that HMs determine not only what parts of the genome are expressed, but also how they are spliced. However, few studies focused on the detailed mechanisms, i.e., epigenetic regulations on AS in the context of cell fate decision.

Additionally, cell-cycle machinery dominates the mechanisms underlying ESC pluripotency and differentiation [21, 53]. Changes of cell fates require going through the cell-cycle progression. Studies in mESCs [54] and hESCs [55, 56] 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 differentiation from the hESCs (H1 cell line) to five differentiated cell types [16]. These cells cover three germ layers for embryogenesis, adult stem cells, and adult

somatic cells, representing multiple lineages of different developmental levels (Additional file 1: Figure S1A). This carefully selected dataset enabled our understanding of AS epigenetic regulations in the context of cell fate decision. First, we identified several thousands of AS events that are differentially spliced between the hESCs and differentiated cells, including 3,513 mutually exclusive exons (MXE) and 3,678 skipped exons (SE) which were used for further analyses. These hESC differentiation-related AS events involve ~20% of expressed genes and characterize the multiple lineage differentiation. 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. Especially, with experimental valuations, we demonstrated an H3K36me3-regulated isoform switch from PBX1a to PBX1b, which is implicated in hESC differentiation by attenuating the activity of the pluripotency regulatory network. Collectively, we presented a mechanism conveying the HM information into cell fate decision through the regulation of AS, which will drive extensive studies on the involvements of HMs in cell fate decision via determining the transcript structure rather than only the transcript abundance.

Results

1. Alternative splicing characterizes hESC differentiation

The role of AS in the regulation of ES cell fates adds another notable regulatory layer to the know mechanisms that govern stemness and differentiation [57]. 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 [16]. 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). We also considered IMR90, a cell line for primary human fetal lung fibroblast, as an example of terminally differentiated cells. These cells represent five cell lineages of different developmental levels (Additional file 1: Figure S1A). We identified thousands of AS events of all types with their changes of 'per spliced in' (ΔPSIs) are larger 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 [58] (Additional file 1: Figure S1B and Table S1, see Methods). We implemented further analyses only on the most common AS events, including 3,513 mutually exclusive exons (MXE) and 3,678 skipped exons (SE), which are referred to as hESC differentiation-associated AS exons (Additional file 1: Figure S1C and Additional file 2: Table S2).

These hESC differentiation-related AS exons possess typical properties as previously described [59, 60] as follows. (1) Most of their hosting genes are not differentially expressed between hESCs and differentiated cells (Additional file 1: 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 (Additional file 1: Figures S2A, B). (3) The arrangement of shorter AS exons surrounded by longer introns is consistent across cell lineages and AS types (Additional file 1: Figures S2C, D). (4) The lengths of AS exons are more often divisible by three to preserve the reading frame (Additional file 1: Figure S2E).

During hESC differentiation, about 20% of expressed genes undergo AS (2,257 genes for SE and 2,489 genes for MXE), including previously known ESC-specific AS genes, such as the pluripotency factor *FOXP1* [14] (Figure 1A) and the Wnt/β-catenin signalling component *CTNND1* [15] (Figure 1B). These hESC differentiation-related AS genes include many transcription factors, transcriptional co-factors, chromatin remodelling factors, housekeeping genes, and bivalent domain genes implicated in ESC pluripotency and development [41] (Figures 1C and Additional file 1: Figure S1C). Enrichment analysis based on a stemness gene set [61] also shows that hESC differentiation-related AS genes are enriched in the regulators or markers that are most significantly associated with stemness signatures of ESCs (Additional file 1: Figure S3A, see Methods).

Clustering on AS events across cell lineages show 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 the most AS events and exhibits the most similar AS profiles 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 Additional file 1: Figure S3B, C), involved in 29.6% and 38.6% of AS hosting genes (Figures 1E, F and Additional file 1: Figure S3D, E), are lineage-specific. In contrast, only 0.65% of SE and 0.14% of MEX events (Additional file 1: Figures S3B, C), involved in 0.49% and 1.52% of AS hosting genes, are shared by all lineages (Figures 1E, F and Additional file 1: Figure S3D, E). Similar trends are observed from pairwise comparisons (Additional file 1: Figure S3F). Furthermore, one-third of AS genes (n = 881) have both MXE and SE events (Additional file 1: Figure S3G). Only four genes are common across all cell lineages and AS types, of which the AS events of *Ctnnd1* and *Mbd1* have been reported to regulate mESC differentiation [15]. Together, these results

demonstrate that alternative splicing depicts lineage- and developmental level-dependent characterizations of hESC differentiation.

2. Dynamic changes of histone modifications 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 the heritability of existing or newly acquired phenotypic states. Though epigenetic signatures are mainly found to be enriched in promoters and enhancers, 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 [62, 63]. Consistent with previous reports [37, 38, 64], we also observed that few involved splicing factors are differentially expressed during H1 cells differentiation (Additional file 1: Figure S3H, see Methods), which confirms the existence of an additional layer of epigenetic regulations on AS. However, the extents to which the AS is epigenetically regulated and how these AS genes contribute to the 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 associations with AS genes during hESC differentiation.

To investigate whether the dynamic changes of these HMs upon cell differentiation prefer to 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 (normalized Δ reads count, see Methods) in ±150bp regions around the splice sites upon hESC differentiation (Figure 2C and Additional file 1: Figure S4, see Methods). Except for a small part of cases

(with black dots or boxes in Figure 2D), most HMs changed more significantly around AS exons than around constitutive exons upon hESC differentiation (Mann-Whitney-Wilcoxon test, $p \leq 0.05$, Figures 2D and Additional file 1: Figure S4). Nevertheless, some HMs displayed strong links to AS, such as H3K79me1 and H3K36me3, while others only had 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 [39, 46, 65], while the latter are the epigenetic marks of repressed regions and heterochromatin [66]. The link strengths are presented 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 sidebar 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.

3. Three histone modifications are significantly associated with alternative splicing upon hESC differentiation

To quantitatively associate the HMs with AS, all ChIP-seq data were processed for narrow peak calling using MACS2 [67]. For each AS exon of each differentiation lineage, we then quantified the differential inclusion levels, i.e., the changes of 'percent splice in' (ΔPSIs, Additional file 1: Figure S1B), and the differential HMs signals, i.e., the changes of normalized narrow peak height of ChIP-seq (ΔHMs, Additional file 1: Figure S5A, see Methods) between H1 and differentiated cells. We observed significant differences in all HM profiles (except H3K27me3, Additional file 1: Figure S5B) between the inclusion-gain and inclusion-loss exons across cell lineages and AS types (Mann-Whitney-Wilcoxon test, $p \leq 0.05$) (Figure 3A and Additional file 1: Figure S5B). However, three independent correlation tests showed only weak global quantitative associations between the ΔPSIs and ΔHMs for some HMs (Figure 3C and Additional file 1: Figure S5C), including eight HMs for MXE AS exons and eight HMs for SE AS exons. The

weak associations may indicate that only subsets of AS exons are strongly associated with HMs, and *vice versa*, which is consistent with a recent report [68].

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 (Figure 3C and Additional file 1: Figure S5D and S6, see Methods). We obtained three subsets of HM-associated SE exons and three subsets of HM-associated MXE exons (Additional file 3: Table S3). The three HM-associated SE subsets include 180, 664, and 1,062 exons, and are negatively associated with H4K8ac (Additional file 1: Figure S6), negatively associated with H3K36me3 (Figure 3C), and positively associated with H3K36me3 (Additional file 1: Figure S6), respectively. The three HM-associated MXE subsets include 99, 821, and 971 exons, and are positively associated with H3K27ac (Figure 3D), negatively associated with H3K36me3 (Additional file 1: Figure S6), and positively associated with H3K36me3 (Additional file 1: Figure S6), respectively. The exons of each subset show significant correlations between their ΔPSIs and ΔHMs upon hESC differentiation (Figure 3D). These HM-associated AS exons account for an average of 52.8% of hESC differentiation-related AS events, on average (Additional file 1: 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 [47] and MRG15/PTBP1 [46]. The former increases the inclusion levels of targeting AS exons, whereas the latter decreases the inclusion levels [39]. As expected, 139 and 11 of our identified H3K36me3-associated AS genes have been reported to be regulated by SRSF1 [69, 70] (Additional file 1: Figure S5F) and PTBP1 [71] (Additional file 1: Figure S5G), respectively. Taken together, our analysis showed that more than half (52.8%) of hESC differentiation-associated AS events are significantly associated with 3 of 16 HMs during hESC differentiation, including H3K36me3, H3K27ac, and H4K8ac.

4. 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 [16, 17]. 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 even shorter AS exons (Additional file 1: Figure S7A, $p < 0.001$, Student's t-test), though AS exons are shorter than the average length of all exons (Additional file 1: 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 (Additional file 1: 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 (Additional file 1: Figure S7D); the most common shared genes are lineage-independent housekeeping genes (Additional file 1: Figure S7E). These suggest that HM-associated AS genes contribute more to lineage specificity. In addition, 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, D and Additional file 1: Figure 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 (Additional file 1: Figures 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 [21].

Further study of the top enriched cell-cycle AS genes (Figure 4D and Additional file 1: Figure 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 is 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 [21]. The latter play roles in G1 phase, cell cycle arrest, and Wnt/β-catenin signalling (Additional file 1: Figures S8C, D). Since cell fate choices seem to occur or at least be initiated during G1/S transition [55], while cell fate commitment is achieved in G2/M [56], it could be rational for stem cells to change their identity during the G2 phase when HMs are reprogrammed [21].

Intriguingly, the top enriched pathway of HM-associated AS genes is 'ATM/ATR-mediated DNA damage response', which 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 [56]. Together with our previous results [20], it suggests 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. Additionally, many cell-cycle TF genes are involved in the top enriched HM-

associated AS gene set. The pre-B-cell leukaemia 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 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.

5. Splicing of PBX1 links H3K36me3 to hESC fate decision

The past few years have identified key factors required for maintaining the pluripotent state of hESCs [72, 73], 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) [74]. These factors appear to activate a transcriptional network that endows cells with pluripotency [75]. 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 is a member of the TALE (three-amino acid loop extension) family homeodomain transcription factors [76, 77] and well known for its functions in lymphoblastic leukaemia [78-81] and several cancers [82-91]. PBX1 also plays roles in regulating developmental gene expression [92], maintaining stemness and self-renewal [82, 93, 94], and promoting the cell-cycle transition to S phase [95]. Additionally, multiple lines of evidence obtained from *in vivo* and in vitro highlighted its functions as a pioneer factor [88, 96]. However, few studies have distinguished the distinct functions of its different isoforms.

PBX1 has three isoforms [97], including PBX1a, PBX1b, and PBX1c (Figure 5C and Additional file 1: Figure S9A). PBX1a and PBX1b are produced by the alternative splicing of exon 7 (Figure 5A) and attract most of the research attention 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 (Figure 5C and Additional file 1: Figure S9A). This C-terminal alteration of PBX1a has been reported to affect its cooperative interactions with HOX partners [98], which may impart different functions to these two isoforms. We here revealed its H3K36me3-regulated isoform switch between PBX1a and PBX1b, which functions at the upstream of pluripotency transcriptional network to link H3K36me3 with ESC fate decision.

We first observed differential transcript expressions of these two isoforms between the hESCs and differentiated cells, wherein *PBX1a* was predominantly transcribed in hESCs, while *PBX1b* was predominantly induced in differentiated cells (Figure 5A and Additional file 1: Figure S9B). The same trend was also observed in an extended dataset of 56 human cell lines/tissues (Figure 5D) from the Roadmap [99] and ENCODE [100] projects (Additional file 4: Table S4). Additionally, we did not observe significantly different expression of the total *PBX1* and three other *PBX* family members across cell types (Additional file 1: Figure S9C, fold change < 2), indicating that the isoform switch of *PBX1*, rather than the differential expression of its family members play more important roles during hESC differentiation. To further test the possible mechanism by which PBX1b contributes to stem cell differentiation, we investigated the transcription levels of Yamanaka factors. Of these TFs, the *NANOG* is activated by direct promoter binding of PBX1 and KLF4, which is essential for stemness maintenance [93, 101]. Consistently, all these core factors are repressed in differentiated cells where *PBX1b* is highly expressed (Additional file 1: Figures S9D-G), even though the *PBX1a* is expressed. 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* (Figure 5E), as well as positive correlations between these two factors and *PBX1a* (or inclusion level of exon 7, Additional file 1: Figures S10A-B). Consistent with previous reports showing that the PBX1a and PBX1b differ in their ability to activate or repress the expression of reporter genes [102, 103], we hypothesize that PBX1a promotes the activity of the

pluripotent regulatory network by promoting the expression of *NANOG*, whereas PBX1b may attenuate this activity by competitively binding and regulating the same target gene, since PBX1b retains the same DNA-binding domain as PBX1a. These observations are strongly suggestive that the switch from PBX1a to PBX1b is a mechanism by which PBX1 contributes to hESC differentiation via regulating the pluripotency regulatory network.

The exon 7 of PBX1 shows significantly positive correlations between its inclusion levels (PSIs) 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 [46] and PSIP1/SRSF1 [47], which endow dual functions to H3K36me3 in AS regulation [39]. Based on the 56 cell lines/tissues from the Roadmap/ENCODE projects, we first found significant positive correlations between the expressions of *PBX1a* (or inclusion level of exon 7) and *PSIP1/SRSF1* (Figure 5F), but not with *MRG15/PTBP1* (Additional file 1: Figures S10C-D). This result suggests that the alternative splicing of *PBX1* is epigenetic regulated by H3K36me3 through the PSIP1/SRSF1 adaptor system, which was strongly supported by a recent report using the HeLa cell lines [69]. The overexpression of SRSF1 in Hela cells introduces a PSI increase of 0.18 for the exon 7 of PBX1 (chr1: 164789308-164789421 based on NCBI37/hg19 genome assembly) based on the RNA-seq (Table S1 of [69]). Additionally, this exon was one of the 104 AS exons that were further validated using radioactive RT-PCR (Table S2 of [69]). Their results showed that exon 7 of PBX1 is indeed a splicing target of SRSF1, supporting our conclusions.

We then validated the above hypotheses on MSCs and IM90 cells, since these two cells types show the most significant difference from H1 cells regarding our hypotheses (Figure 5B). We cultured H1 cells, IMR90 cells, and induced H1 cells to differentiate into MSCs (H1-MSCs, see Methods for details).

Additionally, we also included other two sources of MSCs, including one derived from human bone marrow (hBM-MSCs) and the other derived from adipose tissue (hAT-MSCs) (see Methods for details). Consistent with the results from RNA-seq, the same expression patterns of Yamanaka factors in H1, MSCs, and IMR90 cells were observed using qRT-PCR and Western blot (Figure 6A), which confirmed the pluripotent state of H1 cells and the differentiated states of other cell types. We then detected the isoform switch from PBX1a to PBX1b in our cultured cells, which are consistent both in mRNA and protein levels (Figure 6B and Additional file 1: Figure S10E) and further confirmed by the western blot using PBX1b-specific antibody (anti-PBX1b) (Figure 6B bottom and Additional file 1: Figure S10E iii). These results have verified that the PBX1b was significantly induced in differentiated cells, where the PBX1a was significantly reduced.

We also validated the mechanism by which the splicing of PBX1 links H3K36me3 to stem cell fate decision. We first confirmed that PBX1b also binds to the promoter of *NANOG* at the same region where PBX1a binds to, and the binding signals (ChIP-PCR) were high in the differentiated cells but very low in H1 stem cells (Figure 6C i and Additional file 1: Figure S10F i). Consistent with the results from ChIP-seq, we also observed reduced H3K36 tri-methylation around the exon 7 of PBX1 based on ChIP-PCR assay (Figure 6C ii and Additional file 1: Figure S10F ii). Furthermore, the chromatin factor PSIP1 only showed high binding signal in H1 stem cells (Figure 6C iii and Additional file 1: Figure S10F iii), which recruit the splicing factor SRSF1 to the PBX1 exclusively in H1 stem cells (Figure 6D and Additional file 1: Figure S10G) even though the physical binding between these two factors were universally detected in all cell types (Figure 6E). All these experimental results suggested that, upon differentiation, stem cells reduced the H3K36 tri-methylation and may attenuate the recruitment of PSIP1/SRSF1 adaptor around the exon 7 of PBX1, leading to the exclusion of exon 7 and highly expressed PBX1b in differentiated cells. High expression of PBX1b may attenuate the activity of PBX1a in promoting the pluripotency regulatory network.

Taken together, we suggested that H3K36me3 regulates the alternative splicing of PBX1 via the PSIP1/SRSF1 adaptor system, leading the isoform switch from PBX1a to PBX1b during hESC differentiation. Subsequently, PBX1b competitively binds to *NANOG* and abolishes the bindings of PBX1a. This competitive binding attenuates the pluripotency regulatory network to repress self-renewal and consequently facilitate differentiation (Figure 6F). These findings revealed how the PBX1 contributes to cell fate decision and exemplify the mechanism by which alternative splicing links histone modifications to stem cell fate decision.

Discussion

ESCs provide a vital tool for studying the regulation of early embryonic development and cell-fate decision.[1]. In addition to the core pluripotency regulatory network, emerging evidence revealed other processes regulating ESC pluripotency and differentiation, including histone modifications (HMs), alternative splicing (AS), cell-cycle machinery, and signalling pathways [56]. 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 analyses implicated the association of HMs, AS regulation, and cell-cycle progression with hESC fate decision. Specifically, HMs orchestrate a subset of AS outcomes that play critical roles in cell-cycle progression via the related pathways, such as ATM/ATR-mediated DNA response [20], and TFs, such as PBX1 (Additional file 1: Figure S10H). In this way, HMs, AS regulation, and signalling pathways are converged into the cell-cycle machinery that has been claimed to rule the pluripotency [21].

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 [46, 47], H3K4me3 [44], H3K9me3 [45], and the

acetylation of H3 and H4 [48-52] 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 [68]. Like that report, we also found a positive association of H3K27ac with a subset of AS events. However, our results differ regarding the other two HMs that we identified to be 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 [65], and it has been reported to have dual roles in AS regulations through two different chromatin-adapter systems, PSIP1/SRSF1 [46] and MRG15/PTBP1 [47]. SRSF1 is a splicing factor which will increase the inclusion of targeted AS exons, and PTBP1 will decrease the inclusion levels of the regulated AS exons. Therefore, the exons that are regulated by the PSIP1/SRSF1 adapter system will show positive correlations with the H3K36me3, while the exons regulated by MRG15/PTBP1 will show negative correlations. Our results are consistent with this fact and show both direction correlations between different sets of AS events and H3K36me3. Many of these AS events in our study have been validated by other studies (Additional file 1: Figures S5F-G).

H4K8ac is associated with the fewest number of AS events in our results. Although rarely studied, H4K8ac is known to act as a transcriptional activator in both the promoters and transcribed regions [104]. Its negative association with AS is supported by the finding that it recruits proteins involved in increasing the Pol II elongation rate [105]. 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 [68] 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 were 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 [39]. We demonstrated that the HMs-associated AS events have a significant impact on 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 pluripotency state dissolution and serves as a gatekeeper of the pluripotency state through the cell cycle [56]. The cell cycle has been considered the hub machinery for cell fate decision [21] since all commitments will go through the 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 (Additional file 1: Figure S10H).

We also exemplified our hypothesized mechanism by an H3K36me3-regulated isoforms switch of PBX1. In addition to its well-known functions in lymphoblastic leukaemia [78-81] and a number of cancers [82-91], PBX1 was also found to promote hESC self-renewal by corporately binding to the regulatory elements of *NANOG* with KLF4 [101]. We found that the transcriptions of two isoforms of *PBX1*, *PBX1a* and *PBX1b*, are 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 the activity of the core pluripotency regulatory network composed of Yamanaka factors. The switch from PBX1a to PBX1b is modulated by H3K36me3 via the PSIP1/SRSF1 adapter system [47]. 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).

A very recent report showed that the switch in Pbx1 isoforms was regulated by Ptbp1 during neuronal differentiation in mice [106], indicating a contradiction that the AS of PBX1 should be negatively regulated by H3K36me3 via the MRG15/PTBP1 [46]. 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 [106] showing that NPCs and mature neurons express increasing levels of PBX1a rather than PBX1b (Additional file 1: Figure S9B). Another recent report showed that PBX1 was a splicing target of SRSF1 in the HeLa cell line [69], which strongly supports our findings. Taken together, these evidence suggests that there are two parallel mechanisms regulating PBX1 isoforms in embryonic development, in which neuronal differentiation adopts a mechanism that is different from other lineages.

Finally, it is worth noting that both our work and other studies [68] reported that HMs cannot explain all AS events identified either during ESC differentiation 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 the recruitment of H3K36 methyltransferase HYPB/Set2, resulting in significant differences in H3K36me3 around the AS exons [68]. 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. Additionally, as we described in

our previous study, the 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' [20]. Taken together, we presented a mechanism conveying the HM information into cell fate decision through the AS of cell-cycle factors or the core components of pathways that controlling cell-cycle progression (Additional file 1: Figure S10H).

Conclusions

We performed integrative analyses on transcriptome and epigenome data of the hESCs, H1 cell line, and five differentiated cell types, demonstrating that three of 16 HMs were strongly associated with half of AS events upon hESC differentiation. We proposed 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 experimentally validated one cell cycle-related transcription factor, PBX1, which demonstrated that alternative splicing provides a mechanism conveying the histone modification information into the regulation of cell fate decisions (Figure 6F). Our study will have a broad impact on the field of epigenetic reprogramming in mammalian development involving splicing regulations and cell cycle machinery.

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 [40]. Two BAM files (replicates) of each cell type were analyzed using rMATS (version 3.0.9) [58] and MISO (version 0.4.6) [107]. The rMATS was used to identify AS exons based on the differential ' per cent splice in' (PSI, Ψ) values between each differentiated cell type and H1 cells (Additional file 1: Figure S1B). The splicing changes (ΔPSIs or ΔΨ) are used to identify the AS events between H1 and other cell types. A higher cutoff is always helpful in reducing the false positive

while compromising the sensitivity. The cutoff, $|\Delta\Psi| \geq 0.1$ or $|\Delta\Psi| \geq 10\%$ is widely accepted and used in AS identification [26, 108-110]. Many other studies even used 0.05 as the cutoff [111-115]. We did additional correlation analyses based on different ΔPSI cutoffs (0.1, 0.2, 0.3, 0.4, and 0.5). With the increase of the cutoffs, the number of AS events was significantly reduced (Additional file 1: Figure S11A), however, the correlations ware only slightly increased between AS and some histone modifications (Additional file 1: Figures S11B-C, upper panels), i.e. no consistently impacts of the cutoffs on the correlations were observed. Similarly, the correlation significances were also not consistently affected (Additional file 1: Figure S11B-C, lower panels). Therefore, in our study, 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.

All identified AS event types are summarized in Additional file 1: Table S1. Finally, two types of AS exons, namely skipped exons (SE) and mutually excluded exons (MXE), which are most common and account for major AS events, were used for subsequent analysis (Additional file 2: Table S2). MISO was used to estimate the confidence interval of each PSI and generate Sashimi graphs [116] (see Figures 1A, B 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.

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) [40]. 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 were 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 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 Additional file 1: Figure 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 (Figure 3A and Additional file 1: Figure S5B) using the same strategy. The statistical results (Figure 2C and Additional file 1: Figure 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/) [67, 117] 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. Because the 3' splice sites (3' ends of the introns) determine the inclusion of the downstream AS exons [118] and the distances from the peaks to their target sites affect their regulatory effects [119], we normalize the peak signals against the distance (in kb) between the peak summits and 3' splice sites (Additional file 1: Figure S5A). Since there is no evidence showing that distal HMs could regulate the AS, we only considered local 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 (MLR), 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'). MLR and LLR were calculated using Weka 3 [120], 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 (Additional file 1: 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 (Additional file 3: Table S3).

K-means clustering was performed separately on inclusion-gain and inclusion-loss AS events of MXE and SE, based on the selected HM signatures (Additional file 1: Figure S5C, checked with a "√"). *K* was set to six for all clustering processes (Additional file 1: 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 (Additional file 1: Figure S6). Finally, we identified six subsets of HM-associated AS events displaying significantly positive or negative correlations with three HMs, respectively.

Gene expression quantification

For each cell type, two aligned BAM files (replicates) were used to estimate the expression level of genes using Cufflinks [121]. 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 [122]. The enrichment significances in Additional file 1: Figures 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 the 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, rather than the direct control on SFs expression.

Genome annotations

Since the RNA-seq reads (BAM) files and ChIP-seq read (BAM) files downloaded from the public sources were mapped to different human genome assemblies, 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 [123]. The hg18 GTF file was used for rMATS and MISO to identify AS during the differentiation from H1 ESCs into 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 2A-C, 3A, and Additional file 1: Figures S4 and S5A). We compared exonic and intronic lengths based on hg18 annotation (Additional file 1: Figure S2).

Gene ontology enrichment analysis

The Gene Ontology (GO) enrichment analysis was performed using ToppGene [124] 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 Additional file 1: Figures S8A-C using either the FDR (0.05) adjusted p-value or the enriched gene numbers (Additional file 1: 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 (Figure 4F and Additional file 1: Figure S8D) analysis through Ingenuity Pathway Analysis (IPA, http://www.ingenuity.com/products/ipa).

Stemness signature and TF binding enrichment analysis

StemChecker [61], 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 were 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. Additional file 1: Figure S3A shows the enrichment significance as the -log10 p-values (Bonferroni adjusted). 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 are shown as -log10 p-values (Bonferroni adjusted) in Figure 4A.

FunRich (v2.1.2) [125, 126], a stand-alone functional enrichment analysis tool, was used for TF binding enrichment analysis to get the enriched TFs that may regulate the query genes. 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

All raw data are available from the GEO accession IDs GSE18927 and GSE16256. The individual sources of RNA-seq data for 56 cell lines/tissues from Roadmap/ENCODE projects were listed in Additional file 4: Table S4. 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 5 and Additional file 1: Figure S10 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, 3A, and Additional file 1: Figures S4 and S5B, with Fisher's exact test for Figures 1D, 2D, and Additional file 1: Figure S5B, with Student's t-test for Figures 5D and Additional file 1: Figure S7A, and with a

Hypergeometric test for Additional file 1: Figures S1D and S3H. Levels of correlation significance were calculated with Pearson correlation (PC), multiple linear regression (MLR), and logistic regression (LLR) for Figures 3D, 3C, and Additional file 1: Figure S5C. MLR and LLR were performed using Weka 3 [120], whereas all other tests were performed using R packages. The p-values for the enrichment analyses (Figure 4, Additional file 1: Figures S3A and S8) were adjusted either by PDR or Bonferroni (refer to the corresponding method section for details). The statistical analyses of the ChIP, RIP and western blotting assays were shown in Additional file 1: Figures S10E-G. Three replicates were conducted for each assay and the quantitatively statistical analyses were performed on the relative band intensities normalized by control genes (ß-Actin) or input signals. ImageJ (https://imagej.nih.gov/ij/) was used to quantify the band intensities. ANOVA was used for statistical significance tests.

Cell culture and MSCs induction

Human embryonic stem cell (H1cell line) line was purchased from the WiCell Research Institute (product ID: WA01). H1 cells were cultured on the matrigel-coated 6-well culture plates (BD Bioscience) in the defined mTeSR1 culture medium (STEMCELL Technologies). The culture medium was refreshed daily. The H1-derived mesenchymal stem cells (H1-MSC) were differentiated from H1 cells as described previously [127]. Briefly, small H1 cell aggregates were cultured on a monolayer of mouse OP9 bone-marrow stromal cell line (ATCC) for 9 days. After depleting the OP9 cell layer, the cells were then cultured in semisolid colony-forming serum-free medium supplemented with fibroblast growth factor 2 and platelet-derived growth factor BB for 2 weeks. The mesodermal colonies were selected and cultured in mesenchymal serum-free expansion medium with FGF2 to generate and expand H1-MSCs. hAT-MSCs were derived from the subcutaneous fats provided by the National Disease Research Interchange (NDRI) using the protocol described previously [128]. Briefly, the adipose tissue was mechanically minced and digested with collagenase Type II (Worthington Bio, Lakewood, NJ). The resulted single cell suspension

was cultured in α-minimal essential medium with 5% human platelet lysate (Cook Regentec, Indianapolis, IN, USA), 10 μg/ml gentamicin (Life Technologies, CA, USA), and 1x Glutamax (Life Technologies). After reaching ~70% confluence, the adherent cells were harvested at the passage 1 (P1) hAT-MSC/AEY239. hBM-MSCs were isolated from commercially available fresh human bone marrow (hBM) aspirates (AllCells, Emeryville, CA) and expanded following a standard protocol [129]. Briefly, hBM-MSC were cultured in α-minimal essential medium supplemented with 17% fetal bovine serum (FBS; lot-selected for rapid growth of MSC; Atlanta Biologicals, Norcross, GA), 100 units/ml penicillin (Life Technologies, CA, USA), 100 μg/ml streptomycin (Life Technologies), and 2 mM L-glutamine (Life Technologies). After reaching ~70% confluence, the adherent cells were harvested at the passage 1 (P1) hBM-MSC-5204(G). The human IMR-90 cell line was purchased from the American Tissue Type Culture Collection (Manassas, VA) and cultured in Eagle's Minimum Essential Medium (Thermo Scientific, Logan, UT) supplemented 10% fetal calf serum, 2 mM L-glutamine, 100U/mL penicillin, and 100ug/mL streptomycin at 37°C in humidified 5% CO_2.

RNA isolation and reverse transcription PCR (RT-PCR)

Total RNA was extracted from cells using the RNeasy Mini plus kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. RT reactions for first-strand cDNA synthesis were performed with 2μg of total RNA in 25uL reaction mixture containing 5 μl of 5X first-strand synthesis buffer, 0.5 mM dNTP, 0.5 μg oligo(dT)12-18mer (Invitrogen), 200 units of M-MLV reverse transcriptase (Promega, Madison, WI) and 25 units of RNase inhibitor (Invitrogen). The mixture was then incubated at 42°C for 50 min and 52°C for 10 min. The reaction was stopped by heating at 70°C for 15 min. The PCR amplifications were carried out in 50 μl reaction solution containing 1 μl of RT product, 5 μl of 10X PCR buffer, 0.15 mM $MgCl_2$, 0.2 mM dNTP, 0.2 mM sense and antisense primers, and 2.5 U Taq polymerase (Bohringer, Mannheim, Germany). The sequences for the upstream and downstream primers of PBX1 and β-actin

are listed in Additional file 1: Table S5. PCR reaction solution was denatured initially at 95°C for 3 min, followed by 35 cycles at 95°C for 1 min, 55˚C for 40 s, 72˚C 40 s. The final extension step was at 72°C for 10 min. The PCR products were resolved in a 2% ethidium bromide-containing agarose gel and visualized using ChemiDoc MP Imager (Bio-Rad).

Quantitative Real-time PCR (qRT-PCR)

The qPCR amplification was done in a 20uL reaction mixture containing 100 ng of cDNA, 10 uL 2 X All-in-One™ qPCR mix (GeneCopoeia, Rockville, MD), 0.3 mM of upstream and downstream primers and nuclear-free water. The PCR reaction was conducted with 1 cycle at 95 °C for 10 min, 40 cycles at 95 °C for 15 s, 40 °C for 30 s and 60 °C for 1 min, followed by dissociation curve analysis distinguishing PCR products. The expression level of a gene was normalized with the endogenous control gene β-actin. The relative expressions of genes were calculated using the 2-ΔΔCT method, normalized by β-actin and presented as mean ± SD (n = 3) (Figure 6A). The sequences of the paired sense and antisense primers for human SOX2, NANOG, OCT4, c-MYC, KLF4 and β-actin are listed in the Additional file 1: Table S5.

Western blotting

Cells were lysed with 1 X RIPA buffer supplemented with protease and phosphatase inhibitor cocktail (Roche Applied Science, Indianapolis, IN) and stored in aliquots at – 20 °C until use. Twenty micrograms of cell lysates were mixed with an equal volume of Laemmli sample buffer, denatured by boiling and separated by SDS-PAGE. The separated proteins were then transferred to PVDF membranes (Bio-Rad, Hercules, CA). The membranes were blocked using 5% BSA for 1 h at room temperature and incubated with the 1st antibodies against in 1% BSA overnight. SOX2, OCT4, NANOG, KLF4, c-MYC and β-actin antibodies were from Cell signalling technology (Beverly, MA). PBX1 antibody was from Abcam (Cambridge, MA) and PBX1b antibody from Santa Cruz technology (Dallas, Texas). After incubated with

IgG horseradish peroxidase-conjugated secondary antibodies (Cell signalling) for 2 h at room temperature, the immunoblots were developed using SuperSignal West Pico PLUS Chemiluminescent reagent (Thermo Fisher Scientific, Waltham, MA) and imaged using ChemiDoc MP Imager (Bio-Rad).

Co-immunoprecipitation (Co-IP)

Co-IP was used to validate the PSIP1-SRSF1 interaction (Figure 6E). Cells were lysed with ice-cold non-denaturing lysis buffer containing 20 mM Tris HCl (pH 8), 137 mM NaCl, 1% Nonidet P-40, 2 mM EDTA and proteinase inhibitor cocktails. 200 µg protein were pre-cleared with 30uL protein G magnetic beads (Cell signalling) for 30 minutes at 4 °C on a rotator to reduce non-specific protein binding. The pre-cleared lysate was immunoprecipitated with the 5µL anti-SRSF1 antibody (Invitrogen) overnight and then incubated with protein G magnetic beads for 4 hours at 4 °C. Beads without anti-SRSF1 antibody were used as an IP control. The protein G magnetic beads were washed 5 times with lysis buffer and the precipitated protein collected. The SRSF1-bound PSIP1 protein (also known as LEDGF) level was determined with the PSIP1 antibody (Human LEDGF Antibody, R&D systems) using Western blot as described above. Three specific bands (one for p75 and two for p52) were detected (Figure 6E) as indicated in R&D systems website (https://www.rndsystems.com/products/human-ledgf-antibody-762815_mab3468).

Chromatin immunoprecipitation (ChIP)

ChIP assay was performed using a SimpleChIP Enzymatic Chromatin IP Kit (Cell Signaling Technology) according to the manufacturer's instruction. Briefly, 2 X 10^7 cells were cross-linked with 1% formaldehyde and lysed with lysis buffer. Chromatin was fragmented by partial digestion with Micrococcal Nuclease Chromatin. The protein-DNA complex was precipitated with ChIP-Grade Protein G Magnetic Beads (Cell signalling) and ChIP-validated antibodies against H3K36me3 (Abcam), PSIP1

(Novus), and PBX1b (Santa Cruz). Normal mouse IgG and normal rabbit IgG were used as negative controls. After reversal of protein-DNA cross-links, the DNA was purified using DNA purification spin columns. The purified DNA fragments were then amplified with the appropriate primers on T100 thermal cycler (Bio-Rad). The primer pairs used for PCR are listed in the Additional file 1: Table S5. The H3K36me3- and PSIP1-immunoprecipitated DNA fragments surrounding the exon 7 of PBX1b and the promoter region of NANOG in the PBX1b-immunoprecipitated DNA fragments were PCR-amplified. The ChIP-PCR products were revealed by electrophoresis on a 2% agarose gel (Figure 6C).

RNA immunoprecipitation (RIP)

RIP assay was performed using Magna RIP™ RNA-Binding Protein Immunoprecipitation Kit (Millipore Sigma) following the manufacturer's instruction. Briefly, the cells were lysed with RIP lysis buffer with RNase and protease inhibitors. 500 µg of total protein was precleared with protein G magnetic beads for 30 minutes. The protein G magnetic beads were preincubated with 5 µg of mouse monoclonal anti-SRSP1 antibody or normal mouse IgG for 2 hours at 4° C. The antibody-coated beads were then incubated with precleared cell lysates at 4° C overnight with rotation. The RNA/protein/beads conjugates were washed 5 times with RIP was buffer and the RNA-protein complexes were eluted from the protein G magnetic beads on the magnetic separator. The SRSP1-bound RNA was extracted using acid phenol-chloroform and precipitated with ethanol. The RNA was then reverse-transcribed and the expression levels of PBX1a and PBX1b in the immunoprecipitated and non-immunoprecipitated (input) samples were analyzed using RT-PCR (Figure 6D).

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.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All RNA-seq and 16 HMs ChIP-seq data of H1 and five other differentiated cells are available in Gene Expression Omnibus (GEO) under accession number GSE16256 [40]. The BAM files of the RNA-seq data (two replicates for each, aligned to human genome hg18) are alternatively available at http://renlab.sdsc.edu/differentiation/download.html. Both RNA-seq and ChIP-seq data of 56 cell lines/tissues from the Roadmap/ENCODE projects [99, 100] are available on their official website (RoadMap: ftp://ftp.ncbi.nlm.nih.gov/pub/geo/DATA/roadmapepigenomics/by_sample/; ENCODE: ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/) and all raw files are also available at GEO from the accession IDs GSE18927 and GSE16256. Additional file 4: Table S4 provides the detailed information of these data.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the National Institutes of Health (NIH) [1R01GM123037 and AR069395].

Author's Contributions

YX, WZ, and XZ conceived the study. YX carried out the sequencing data analysis and interpreted the results. WZ conducted the experimental validations. WZ, SO, and KP worked on growing and characterizing the MSCs. YX wrote the first draft of the manuscript and all authors revised it. All authors read and approved the final version of the manuscript.

Acknowledgements

We would like to give our special thanks to Drs Guangxu Jin, Liang Liu, and Dongmin Guo from Wake Forest School of Medicine, who gave a lot of comments and suggestions on the data analyses and interpretations. We also acknowledge the editorial assistance of Karen Klein, MA, in the Wake Forest Clinical and Translational Science Institute (UL1 TR001420; PI: McClain).

Additional files

Additional file 1: Figure S1. Identifying hESC differentiation-related AS exons. **Figure S2.** The hESC differentiation-related AS exons possess the typical properties of AS exons. **Figure S3.** AS profiles upon hESC differentiation show lineage-specific splicing pattern. **Figure S4.** Histone modifications (HMs) change significantly around the alternatively spliced (AS) exons upon hESC differentiation. **Figure S5.** A subset of AS events are significantly associated with some HMs upon hESC differentiation. **Figure S6.** K-means clustering based on selected epigenetic features of eight HMs for MXE and SE AS exons. **Figure**

S7. HM-associated AS genes are more lineage-specific. **Figure S8.** HM-unassociated AS genes are enriched in G1 cell-cycle phase and pathways for self-renewal. **Figure S9.** Isoform switch from PBX1a and PBX1b during hESC differentiation. **Figure S10.** Isoform switch of PBX1 links H3K36me3 to hESC fate decision. **Figure S11.** The effect of ΔPSI cutoffs for AS-HM correlations. **Table S1.** The number of all AS events identified during hESC differentiation. **Table S5.** The PCR primers used in this study. (PDF 2.1MB)

Additional file 2: Table S2. Alternative splicing events (AS exons) during the differentiation from H1 cells to differentiated cells. (XLSX 1.8MB)

Additional file 3: Table S3. HM-associated AS exons based on k-means clustering. (XLSX 1.1MB)

Additional file 4: Table S4. 56 cell lines/tissues and their corresponding RNA-seq data sources from ENCODE and Roadmap projects. (XLSX 14.5KB)

References

1. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM: **Embryonic stem cell lines derived from human blastocysts.** *Science* 1998, **282:**1145-1147.
2. Graveley BR: **Splicing up Pluripotency.** *Cell* 2011, **147:**22-24.
3. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JR, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, **122:**947-956.
4. Lian XJ, Zhang JH, Azarin SM, Zhu KX, Hazeltine LB, Bao XP, Hsiao C, Kamp TJ, Palecek SP: **Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions.** *Nature Protocols* 2013, **8:**162-175.
5. Ogaki S, Shiraki N, Kume K, Kume S: **Wnt and Notch Signals Guide Embryonic Stem Cell Differentiation into the Intestinal Lineages.** *Stem Cells* 2013, **31:**1086-1096.
6. Clevers H, Loh KM, Nusse R: **An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control.** *Science* 2014, **346:**1248012.
7. Clevers H, Loh KM, Nusse R: **An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control.** *Science* 2014, **346:**54-+.
8. Itoh F, Watabe T, Miyazono K: **Roles of TGF-beta family signals in the fate determination of pluripotent stem cells.** *Seminars in Cell & Developmental Biology* 2014, **32:**98-106.
9. Sakaki-Yumoto M, Katsuno Y, Derynck R: **TGF-beta family signaling in stem cells.** *Biochimica Et Biophysica Acta-General Subjects* 2013, **1830:**2280-2296.
10. Watabe T, Miyazono K: **Roles of TGF-beta family signaling in stem cell renewal and differentiation.** *Cell Research* 2009, **19:**103-115.

11. Tay Y, Zhang JQ, Thomson AM, Lim B, Rigoutsos I: **MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation.** *Nature* 2008, **455:**1124-1128.
12. Shenoy A, Blelloch RH: **Regulation of microRNA function in somatic stem cell proliferation and differentiation.** *Nature Reviews Molecular Cell Biology* 2014, **15:**565-576.
13. Fatica A, Bozzoni I: **Long non-coding RNAs: new players in cell differentiation and development.** *Nature Reviews Genetics* 2014, **15:**7-21.
14. Gabut M, Samavarchi-Tehrani P, Wang XC, Slobodeniuc V, O'Hanlon D, Sung HK, Alvarez M, Talukder S, Pan Q, Mazzoni EO, et al: **An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming.** *Cell* 2011, **147:**132-146.
15. Salomonis N, Schlieve CR, Pereira L, Wahlquist C, Colas A, Zambon AC, Vranizan K, Spindler MJ, Pico AR, Cline MS, et al: **Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107:**10514-10519.
16. Xie W, Schultz MD, Lister R, Hou ZG, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, et al: **Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells.** *Cell* 2013, **153:**1134-1148.
17. Gifford CA, Ziller MJ, Gu HC, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R, et al: **Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells.** *Cell* 2013, **153:**1149-1163.
18. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, et al: **Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells.** *Cell Stem Cell* 2010, **6:**479-491.
19. Boland MJ, Nazor KL, Loring JF: **Epigenetic Regulation of Pluripotency and Differentiation.** *Circulation Research* 2014, **115:**311-324.
20. Xu Y, Wang Y, Luo J, Zhao W, Zhou X: **Deep learning of the splicing (epi)genetic code reveals a novel candidate mechanism linking histone modifications to ESC fate decision.** *Nucleic Acids Res* 2017, **45:**12100-12112.
21. Vallier L: **Cell Cycle Rules Pluripotency.** *Cell Stem Cell* 2015, **17:**131-132.
22. Pan Q, Shai O, Lee LJ, Frey J, Blencowe BJ: **Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.** *Nature Genetics* 2008, **40:**1413-1415.
23. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al: **Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project.** *Science* 2010, **330:**1775-1787.
24. Graveley BR, Brooks AN, Carlson J, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al: **The developmental transcriptome of Drosophila melanogaster.** *Nature* 2011, **471:**473-479.
25. Ramani AK, Calarco JA, Pan Q, Mavandadi S, Wang Y, Nelson AC, Lee LJ, Morris Q, Blencowe BJ, Zhen M, Fraser AG: **Genome-wide analysis of alternative splicing in Caenorhabditis elegans.** *Genome Research* 2011, **21:**342-348.
26. Wang ET, Sandberg R, Luo SJ, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes.** *Nature* 2008, **456:**470-476.
27. Yeo GW, Xu XD, Liang TY, Muotri AR, Carson CT, Coufal NG, Gage FH: **Alternative splicing events identified in human embryonic stem cells and neural progenitors.** *Plos Computational Biology* 2007, **3:**1951-1967.
28. Salomonis N, Nelson B, Vranizan K, Pico AR, Hanspers K, Kuchinsky A, Ta L, Mercola M, Conklin BR: **Alternative Splicing in the Differentiation of Human Embryonic Stem Cells into Cardiac Precursors.** *Plos Computational Biology* 2009, **5:**e1000553.

29. Lu XY, Goke J, Sachs F, Jacques PE, Liang HQ, Feng B, Bourque G, Bubulya PA, Ng HH: **SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells.** *Nature Cell Biology* 2013, **15:**1141-1152.
30. Kalsotra A, Cooper TA: **Functional consequences of developmentally regulated alternative splicing.** *Nature Reviews Genetics* 2011, **12:**715-729.
31. Mayshar Y, Rom E, Chumakov I, Kronman A, Yayon A, Benvenisty N: **Fibroblast growth factor 4 and its novel splice isoform have opposing effects on the maintenance of human embryonic stem cell self-renewal.** *Stem Cells* 2008, **26:**767-774.
32. Rao S, Zhen S, Roumiantsev S, McDonald LT, Yuan GC, Orkin SH: **Differential Roles of Sall4 Isoforms in Embryonic Stem Cell Pluripotency.** *Molecular and Cellular Biology* 2010, **30:**5364-5380.
33. Fiszbein A, Kornblihtt AR: **Alternative splicing switches: Important players in cell differentiation.** *Bioessays* 2017, **39**.
34. Barash Y, Calarco JA, Gao WJ, Pan Q, Wang XC, Shai O, Blencowe BJ, Frey BJ: **Deciphering the splicing code.** *Nature* 2010, **465:**53-59.
35. Han H, Irimia M, Ross PJ, Sung HK, Alipanahi B, David L, Golipour A, Gabut M, Michael IP, Nachman EN, et al: **MBNL proteins repress ES-cell-specific alternative splicing and reprogramming.** *Nature* 2013, **498:**241-245.
36. Schwartz S, Meshorer E, Ast G: **Chromatin organization marks exon-intron structure.** *Nat Struct Mol Biol* 2009, **16:**990-995.
37. Kornblihtt AR, De la Mata M, Fededa JP, Munoz MJ, Nogues G: **Multiple links between transcription and splicing.** *Rna-a Publication of the Rna Society* 2004, **10:**1489-1498.
38. Wang GS, Cooper TA: **Splicing in disease: disruption of the splicing code and the decoding machinery.** *Nature Reviews Genetics* 2007, **8:**749-761.
39. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T: **Epigenetics in Alternative Pre-mRNA Splicing.** *Cell* 2011, **144:**16-26.
40. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, et al: **Epigenomic analysis of multilineage differentiation of human embryonic stem cells.** *Cell* 2013, **153:**1134-1148.
41. Sharov AA, Ko MSH: **Human ES cell profiling broadens the reach of bivalent domains.** *Cell Stem Cell* 2007, **1:**237-238.
42. Singh AM, Sun YH, Li L, Zhang WJ, Wu TM, Zhao SY, Qin ZH, Dalton S: **Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency.** *Stem Cell Reports* 2015, **5:**323-336.
43. Kouzarides T: **Chromatin modifications and their function.** *Cell* 2007, **128:**693-705.
44. Sims RJ, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D: **Recognition of trimethylated histone h3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing.** *Molecular Cell* 2007, **28:**665-676.
45. Piacentini L, Fanti L, Negri R, Del Vescovo V, Fatica A, Altieri F, Pimpinelli S: **Heterochromatin Protein 1 (HP1a) Positively Regulates Euchromatic Gene Expression through RNA Transcript Association and Interaction with hnRNPs in Drosophila.** *Plos Genetics* 2009, **5:**e1000670.
46. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T: **Regulation of alternative splicing by histone modifications.** *Science* 2010, **327:**996-1000.
47. Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA: **Psip1/Ledgf p52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing.** *Plos Genetics* 2012, **8:**e1002717.
48. Gunderson FQ, Johnson TL: **Acetylation by the Transcriptional Coactivator Gcn5 Plays a Novel Role in Co-Transcriptional Spliceosome Assembly.** *Plos Genetics* 2009, **5:**e1000682.

49. Schor IE, Rascovan N, Pelisch F, Allo M, Kornblihtt AR: **Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing.** *Proc Natl Acad Sci U S A* 2009, **106:**4325-4330.
50. Schor IE, Kornblihtt AR: **Playing inside the genes: Intragenic histone acetylation after membrane depolarization of neural cells opens a path for alternative splicing regulation.** *Commun Integr Biol* 2009, **2:**341-343.
51. Zhou HL, Hinman MN, Barron VA, Geng C, Zhou G, Luo G, Siegel RE, Lou H: **Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner.** *Proc Natl Acad Sci U S A* 2011, **108:**E627-635.
52. Sharma A, Nguyen H, Cai L, Lou H: **Histone hyperacetylation and exon skipping: a calcium-mediated dynamic regulation in cardiomyocytes.** *Nucleus* 2015, **6:**273-278.
53. Dalton S: **Linking the Cell Cycle to Cell Fate Decisions.** *Trends in Cell Biology* 2015, **25:**592-600.
54. Coronado D, Godet M, Bourillot PY, Tapponnier Y, Bernat A, Petit M, Afanassieff M, Markossian S, Malashicheva A, Iacone R, et al: **A short G1 phase is an intrinsic determinant of naive embryonic stem cell pluripotency.** *Stem Cell Research* 2013, **10:**118-131.
55. Pauklin S, Vallier L: **The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity.** *Cell* 2013, **155:**135-147.
56. Gonzales KAU, Liang H, Lim YS, Chan YS, Yeo JC, Tan CP, Gao B, Le BL, Tan ZY, Low KY, et al: **Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways.** *Cell* 2015, **162:**564-579.
57. Aaronson Y, Meshorer E: **STEM CELLS Regulation by alternative splicing.** *Nature* 2013, **498:**176-177.
58. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y: **rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data.** *Proc Natl Acad Sci U S A* 2014, **111:**E5593-5601.
59. Magen A, Ast G: **The importance of being divisible by three in alternative splicing.** *Nucleic Acids Research* 2005, **33:**5574-5582.
60. Zheng CL, Fu XD, Gribskov M: **Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse.** *Rna* 2005, **11:**1777-1787.
61. Pinto JP, Kalathur RK, Oliveira DV, Barata T, Machado RSR, Machado S, Pacheco-Leyva I, Duarte I, Futschik ME: **StemChecker: a web-based tool to discover and explore stemness signatures in gene sets.** *Nucleic Acids Research* 2015, **43:**W72-W77.
62. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129:**823-837.
63. Kornblihtt AR, Schor IE, Allo M, Blencowe BJ: **When chromatin meets splicing.** *Nature Structural & Molecular Biology* 2009, **16:**902-903.
64. Mabon SA, Misteli T: **Differential recruitment of pre-mRNA splicing factors to alternatively spliced transcripts in vivo.** *PLoS Biol* 2005, **3:**e374.
65. Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J: **Differential chromatin marking of introns and expressed exons by H3K36me3.** *Nature Genetics* 2009, **41:**376-381.
66. Mikkelsen TS, Ku MC, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448:**553-U552.
67. Feng JX, Liu T, Qin B, Zhang Y, Liu XS: **Identifying ChIP-seq enrichment using MACS.** *Nature Protocols* 2012, **7:**1728-1740.
68. Curado J, Iannone C, Tilgner H, Valcarcel J, Guigo R: **Promoter-like epigenetic signatures in exons displaying cell type-specific splicing.** *Genome Biol* 2015, **16:**236.

69. Anczukow O, Akerman M, Clery A, Wu J, Shen C, Shirole NH, Raimer A, Sun SY, Jensen MA, Hua YM, et al: **SRSF1-Regulated Alternative Splicing in Breast Cancer.** *Molecular Cell* 2015, **60:**105-117.
70. Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li HR, Qiu JS, Huang J, Yeo GW, Ares M, Fu XD: **Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing.** *Molecular Cell* 2013, **50:**223-235.
71. Xue YC, Zhou Y, Wu TB, Zhu T, Ji X, Kwon YS, Zhang C, Yeo G, Black DL, Sun H, et al: **Genome-wide Analysis of PTB-RNA Interactions Reveals a Strategy Used by the General Splicing Repressor to Modulate Exon Inclusion or Skipping.** *Molecular Cell* 2009, **36:**996-1006.
72. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang WW, Jiang JM, et al: **Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.** *Cell* 2008, **133:**1106-1117.
73. Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, Wray J, Yamanaka S, Chambers I, Smith A: **Nanog Is the Gateway to the Pluripotent Ground State.** *Cell* 2009, **138:**722-737.
74. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors.** *Cell* 2007, **131:**861-872.
75. Kim JW, Chu JL, Shen XH, Wang JL, Orkin SH: **An extended transcriptional network for pluripotency of embryonic stem cells.** *Cell* 2008, **132:**1049-1061.
76. Chang CP, Jacobs Y, Nakamura T, Jenkins NA, Copeland NG, Cleary ML: **Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins.** *Molecular and Cellular Biology* 1997, **17:**5679-5687.
77. Merabet S, Mann RS: **To Be Specific or Not: The Critical Relationship Between Hox And TALE Proteins.** *Trends in Genetics* 2016, **32:**334-347.
78. Melendez CLC, Rosales L, Herrera M, Granados L, Tinti D, Villagran S, Jo W, Villegas M, Antillon F, Torselli S, Alvarez L: **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.** *Clinical Lymphoma Myeloma & Leukemia* 2015, **15:**S170-S170.
79. Tesanovic T, Badura S, Oellerich T, Doering C, Ruthardt M, Ottmann OG: **Mechanisms of Antileukemic Activity of the Multikinase Inhibitors Dasatinib and Ponatinib in Acute Lymphoblastic Leukemia (ALL) Harboring the E2A-PBX1 Fusion Gene.** *Blood* 2014, **124**.
80. Foa R, Vitale A, Cuneo A, Mecucci C, Mancini H, Cimino G, Cutinelli F, Saglio G, Torelli G, Quarta G, et al: **E2A-PBX1 fusion in adult acute lymphoblastic leukemia (ALL) with t(1;19) translocation: Biologic and clinical features.** *Blood* 2000, **96:**189b-189b.
81. Kamps MP, Baltimore D: **E2a-Pbx1, the T(1, 19) Translocation Protein of Human Pre-B-Cell Acute Lymphocytic-Leukemia, Causes Acute Myeloid-Leukemia in Mice.** *Molecular and Cellular Biology* 1993, **13:**351-357.
82. Jung JG, Kim TH, Gerry E, Kuan JC, Ayhan A, Davidson B, Shih IM, Wang TL: **PBX1, a transcriptional regulator, promotes stemness and chemoresistance in ovarian cancer.** *Clinical Cancer Research* 2016, **22**.
83. Magnani L, Patten DK, Nguyen VTM, Hong SP, Steel JH, Patel N, Lombardo Y, Faronato M, Gomes AR, Woodley L, et al: **The pioneer factor PBX1 is a novel driver of metastatic progression in ER'-positive breast cancer.** *Oncotarget* 2015, **6:**21878-21891.
84. Jung JG, Park JT, Stoeck A, Hussain T, Wu RC, Shih IM, Wang TL: **Targeting PBX1 signaling to sensitize carboplatin cytotoxicity in ovarian cancer.** *Clinical Cancer Research* 2015, **21**.
85. Feng Y, Li L, Zhang X, Zhang Y, Liang Y, Lv J, Fan Z, Guo J, Hong T, Ji B, et al: **Hematopoietic pre-B cell leukemia transcription factor interacting protein is overexpressed in gastric cancer and**

promotes gastric cancer cell proliferation, migration, and invasion. *Cancer Sci* 2015, **106**:1313-1322.

86. Li WH, Huang K, Guo HZ, Cui GH, Zhao S: **Inhibition of non-small-cell lung cancer cell proliferation by Pbx1.** *Chinese Journal of Cancer Research* 2014, **26**:573-578.

87. Thiaville MM, Stoeck A, Chen L, Wu RC, Magnani L, Oidtman J, Shih IM, Lupien M, Wang TL: **Identification of PBX1 Target Genes in Cancer Cells by Global Mapping of PBX1 Binding Sites.** *Plos One* 2012, **7**.

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

89. Park JT, Shih Ie M, Wang TL: **Identification of Pbx1, a potential oncogene, as a Notch3 target gene in ovarian cancer.** *Cancer Res* 2008, **68**:8852-8860.

90. Qiu Y, Tomita Y, Zhang B, Nakamichi I, Morii E, Aozasa K: **Pre-B-cell leukemia transcription factor 1 regulates expression of valosin-containing protein, a gene involved in cancer growth.** *Am J Pathol* 2007, **170**:152-159.

91. Berge V, Ramberg H, Eide T, Svindland A, Tasken KA: **Expression of Pbx1 and UGT2B7 in prostate cancer cells.** *European Urology Supplements* 2006, **5**:794-794.

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

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

94. Xu B, Cai L, Butler JM, Chen D, Lu X, Allison DF, Lu R, Rafii S, Parker JS, Zheng D, Wang GG: **The Chromatin Remodeler BPTF Activates a Stemness Gene-Expression Program Essential for the Maintenance of Adult Hematopoietic Stem Cells.** *Stem Cell Reports* 2018.

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

96. Grebbin BM, Schulte D: **PBX1 as Pioneer Factor: A Case Still Open.** *Front Cell Dev Biol* 2017, **5**:9.

97. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM, et al: **RefSeq: an update on mammalian reference sequences.** *Nucleic Acids Research* 2014, **42**:D756-D763.

98. Burglin TR, Ruvkun G: **New Motif in Pbx Genes.** *Nature Genetics* 1992, **1**:319-320.

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

100. Thomas DJ, Rosenbloom KR, Clawson H, Hinrichs AS, Trumbower H, Raney BJ, Karolchik D, Barber GP, Harte RA, Hillman-Jackson J, et al: **The ENCODE project at UC Santa Cruz.** *Nucleic Acids Research* 2007, **35**:D663-D667.

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

102. Asahara H, Dutta S, Kao HY, Evans RM, Montminy M: **Pbx-hox heterodimers recruit coactivator-corepressor complexes in an isoform-specific manner.** *Molecular and Cellular Biology* 1999, **19**:8219-8225.

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

104. Wang ZB, Zang CZ, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui KR, Roh TY, Peng WQ, Zhang MQ, Zhao KJ: **Combinatorial patterns of histone acetylations and methylations in the human genome.** *Nature Genetics* 2008, **40:**897-903.
105. Sim RJ, Belotserkovskaya R, Reinberg D: **Elongation by RNA polymerase II: the short and long of it.** *Genes & Development* 2004, **18:**2437-2468.
106. Linares AJ, Lin CH, Damianov A, Adams KL, Novitch BG, Black DL: **The splicing regulator PTBP1 controls the activity of the transcription factor Pbx1 during neuronal differentiation.** *Elife* 2015, **4**.
107. Katz Y, Wang ET, Airoldi EM, Burge CB: **Analysis and design of RNA sequencing experiments for identifying isoform regulation.** *Nature Methods* 2010, **7:**1009-U1101.
108. Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, et al: **Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins.** *Cell* 2012, **150:**710-724.
109. Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, Quesnel-Vallieres M, Permanyer J, Sodaei R, Marquez Y, et al: **An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms.** *Genome Res* 2017, **27:**1759-1768.
110. Singh B, Trincado JL, Tatlow PJ, Piccolo SR, Eyras E: **Genome Sequencing and RNA-Motif Analysis Reveal Novel Damaging Noncoding Mutations in Human Tumors.** *Mol Cancer Res* 2018.
111. Lin L, Park JW, Ramachandran S, Zhang Y, Tseng YT, Shen S, Waldvogel HJ, Curtis MA, Faull RL, Troncoso JC, et al: **Transcriptome sequencing reveals aberrant alternative splicing in Huntington's disease.** *Hum Mol Genet* 2016, **25:**3454-3466.
112. Ji X, Park JW, Bahrami-Samani E, Lin L, Duncan-Lewis C, Pherribo G, Xing Y, Liebhaber SA: **alphaCP binding to a cytosine-rich subset of polypyrimidine tracts drives a novel pathway of cassette exon splicing in the mammalian transcriptome.** *Nucleic Acids Res* 2016, **44:**2283-2297.
113. Bebee TW, Park JW, Sheridan KI, Warzecha CC, Cieply BW, Rohacek AM, Xing Y, Carstens RP: **The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development.** *Elife* 2015, **4**.
114. Melikishvili M, Chariker JH, Rouchka EC, Fondufe-Mittendorf YN: **Transcriptome-wide identification of the RNA-binding landscape of the chromatin-associated protein PARP1 reveals functions in RNA biogenesis.** *Cell Discov* 2017, **3:**17043.
115. Humphrey J, Emmett W, Fratta P, Isaacs AM, Plagnol V: **Quantitative analysis of cryptic splicing associated with TDP-43 depletion.** *BMC Med Genomics* 2017, **10:**38.
116. Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, Mesirov JP, Airoldi EM, Burge CB: **Sashimi plots: quantitative visualization of RNA sequencing read alignments.** *arXiv preprint arXiv:13063466* 2013.
117. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based Analysis of ChIP-Seq (MACS).** *Genome Biology* 2008, **9**.
118. Chiara MD, Reed R: **A two-step mechanism for 5' and 3' splice-site pairing.** *Nature* 1995, **375:**510-513.
119. Liu L, Jin G, Zhou X: **Modeling the relationship of epigenetic modifications to transcription factor binding.** *Nucleic Acids Res* 2015, **43:**3873-3885.
120. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH: **The WEKA data mining software: an update.** *ACM SIGKDD explorations newsletter* 2009, **11:**10-18.
121. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nature Protocols* 2012, **7:**562-578.

122. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, Hub A, Grp WPW: **AmiGO: online access to ontology and annotation data.** *Bioinformatics* 2009, **25:**288-289.
123. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ: **The UCSC Table Browser data retrieval tool.** *Nucleic Acids Research* 2004, **32:**D493-D496.
124. Chen J, Bardes EE, Aronow BJ, Jegga AG: **ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.** *Nucleic Acids Research* 2009, **37:**W305-W311.
125. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, Mouradov D, Sieber OM, Simpson RJ, Salim A, et al: **FunRich: An open access standalone functional enrichment and interaction network analysis tool.** *Proteomics* 2015, **15:**2597-2601.
126. Pathan M, Keerthikumar S, Chisanga D, Alessandro R, Ang CS, Askenase P, Batagov AO, Benito-Martin A, Camussi G, Clayton A, et al: **A novel community driven software for functional enrichment analysis of extracellular vesicles data.** *J Extracell Vesicles* 2017, **6:**1321455.
127. Vodyanik MA, Yu J, Zhang X, Tian S, Stewart R, Thomson JA, Slukvin, II: **A mesoderm-derived precursor for mesenchymal stem and endothelial cells.** *Cell Stem Cell* 2010, **7:**718-729.
128. Meyerrose TE, De Ugarte DA, Hofling AA, Herrbrich PE, Cordonnier TD, Shultz LD, Eagon JC, Wirthlin L, Sands MS, Hedrick MA, Nolta JA: **In vivo distribution of human adipose-derived mesenchymal stem cells in novel xenotransplantation models.** *Stem Cells* 2007, **25:**220-227.
129. Sekiya I, Larson BL, Smith JR, Pochampally R, Cui JG, Prockop DJ: **Expansion of human adult stem cells from bone marrow stroma: conditions that maximize the yields of early progenitors and evaluate their quality.** *Stem Cells* 2002, **20:**530-541.

Figure legends

Figure 1. Alternative splicing characterizes the hESC differentiation. (A) and **(B)** Sashimi plots show two AS events of previously known ESC-specific AS events, *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)** The 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 ' per cent 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 colour bars indicate the lineage-specific AS genes. The length of the colour 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**.

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 cell types and IMR90. The tracks of H1 are duplicated as yellow shadings 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 (grey 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**. Sidebars 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**.

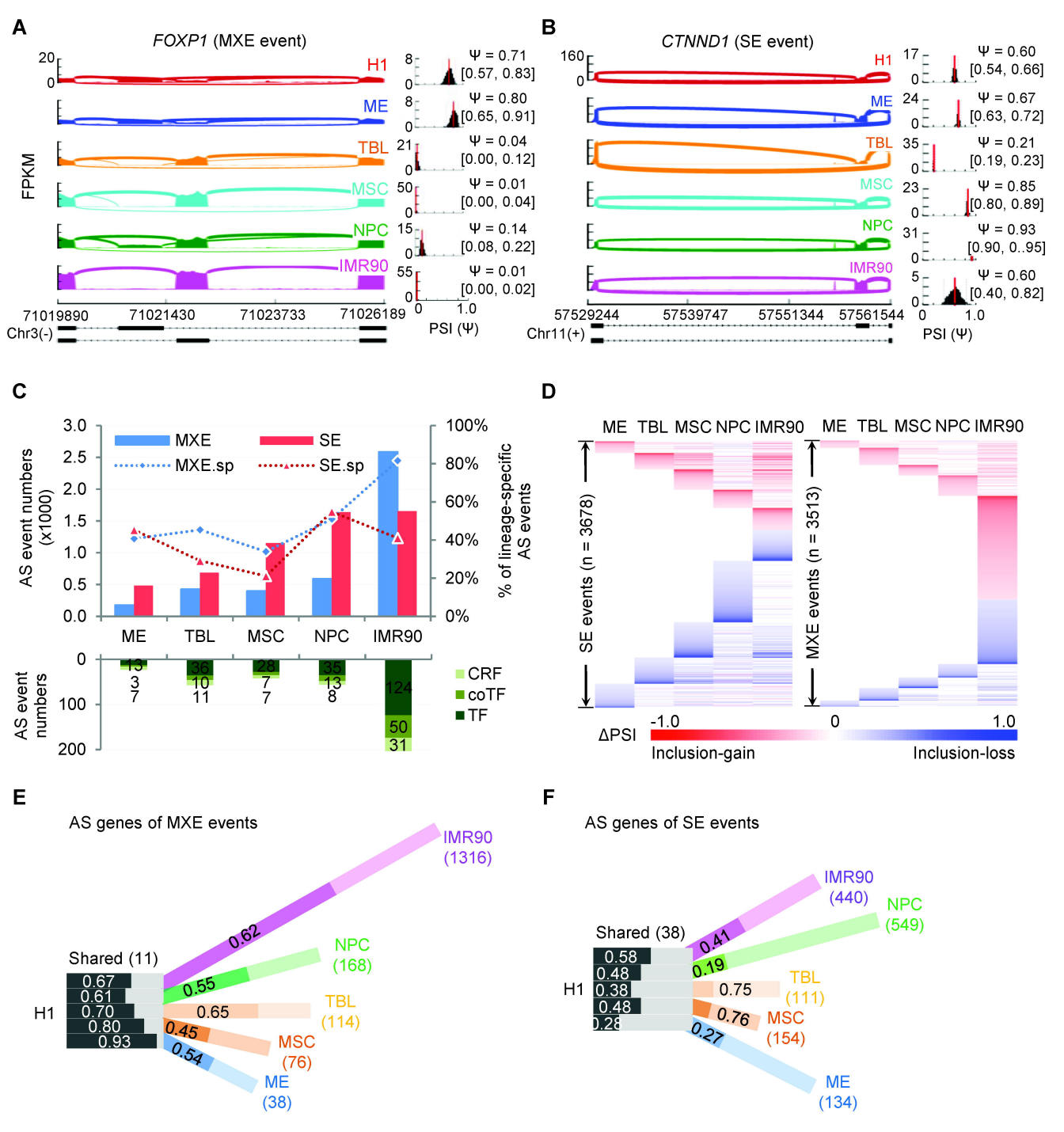
Figure 3. A subset of histone modifications and alternative splicing are strongly associated upon hESC differentiation. (A) Representative profiles of HM (H3K36me3) changes (normalized Δ reads number) around the inclusion-gain (red lines) and inclusion-loss (blue lines) AS exons, as well as randomly 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. **(B)** 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. **(C)** A representative k-means cluster shows a subset of SE AS events having a negative correlation between the ΔPSIs and the ΔHMs of H3K36me3. **Figures S5D** and **S6** provide all the clustering results. **(D)** 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**.

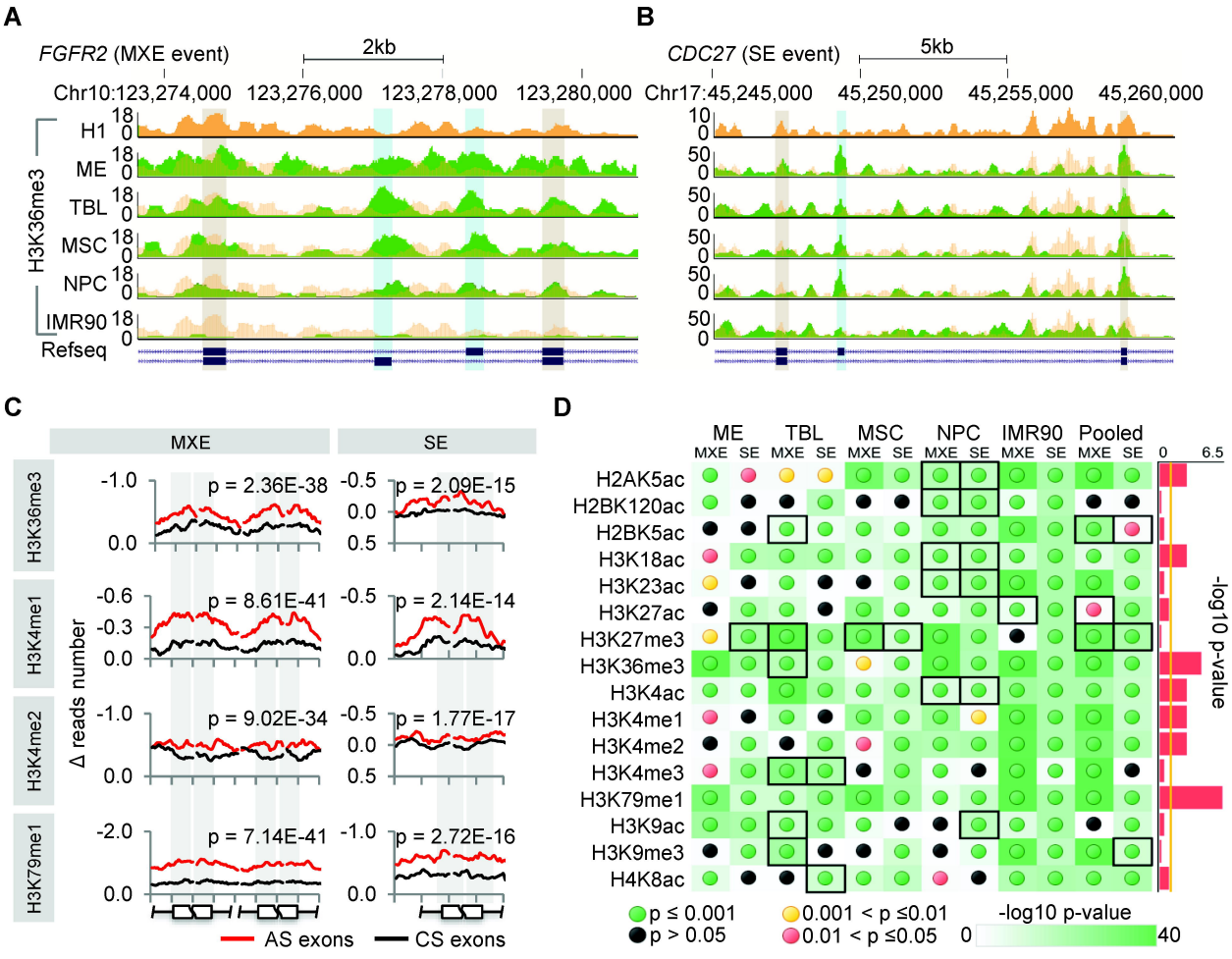
Figure 4. HM-associated AS genes predominantly function in G2/M cell-cycle phases contributing to hESC differentiation. (A) HM-associated AS genes are enriched more significantly in stemness signatures than HM-unassociated AS genes. **(B)** TF binding enrichment shows that HM-associated AS genes prefer 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 (≤ 0.05) adjustment. **(D)** The significant enrichment of HM-associated AS genes in the 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

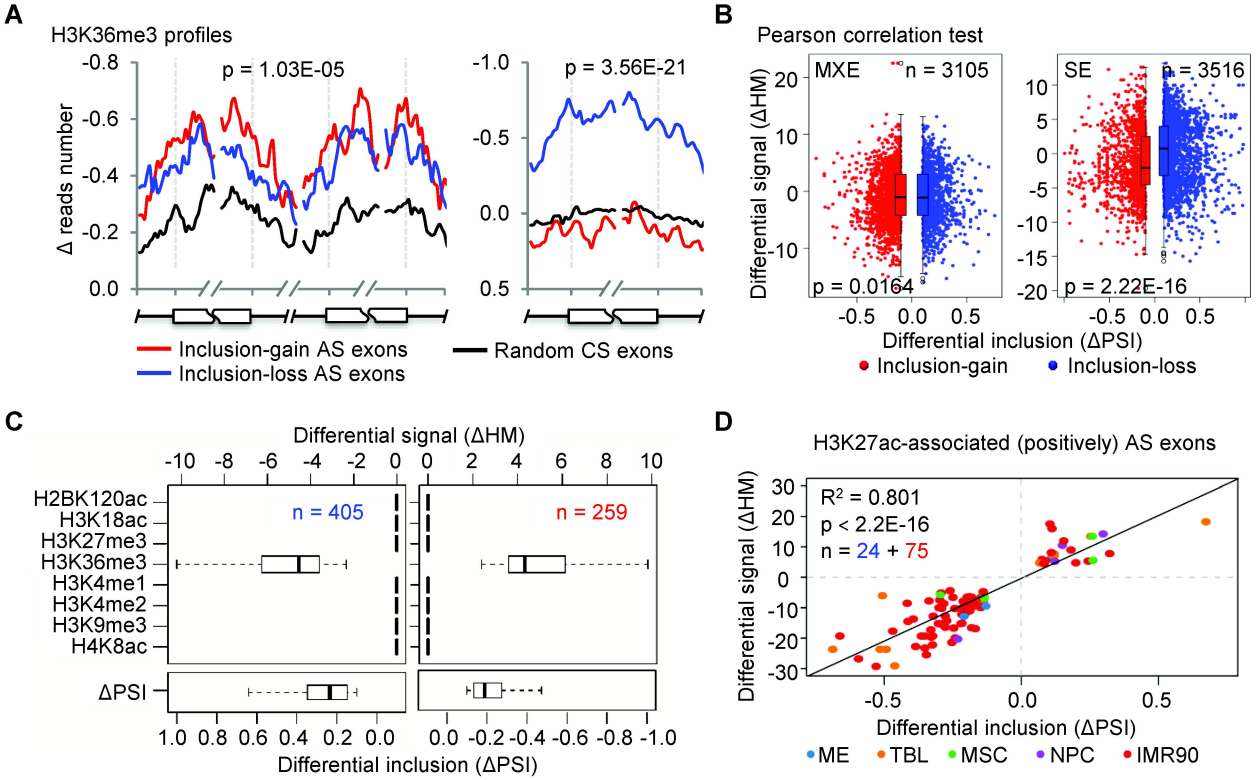
damage response. **(F)** The canonical pathway enrichment shows that AMT/ATR-mediated DNA damage response is the top enriched pathway of HM-associated AS genes. The vertical lines (yellow) indicate the significance cutoff of 0.05. Also, see **Figures S7-S8**.

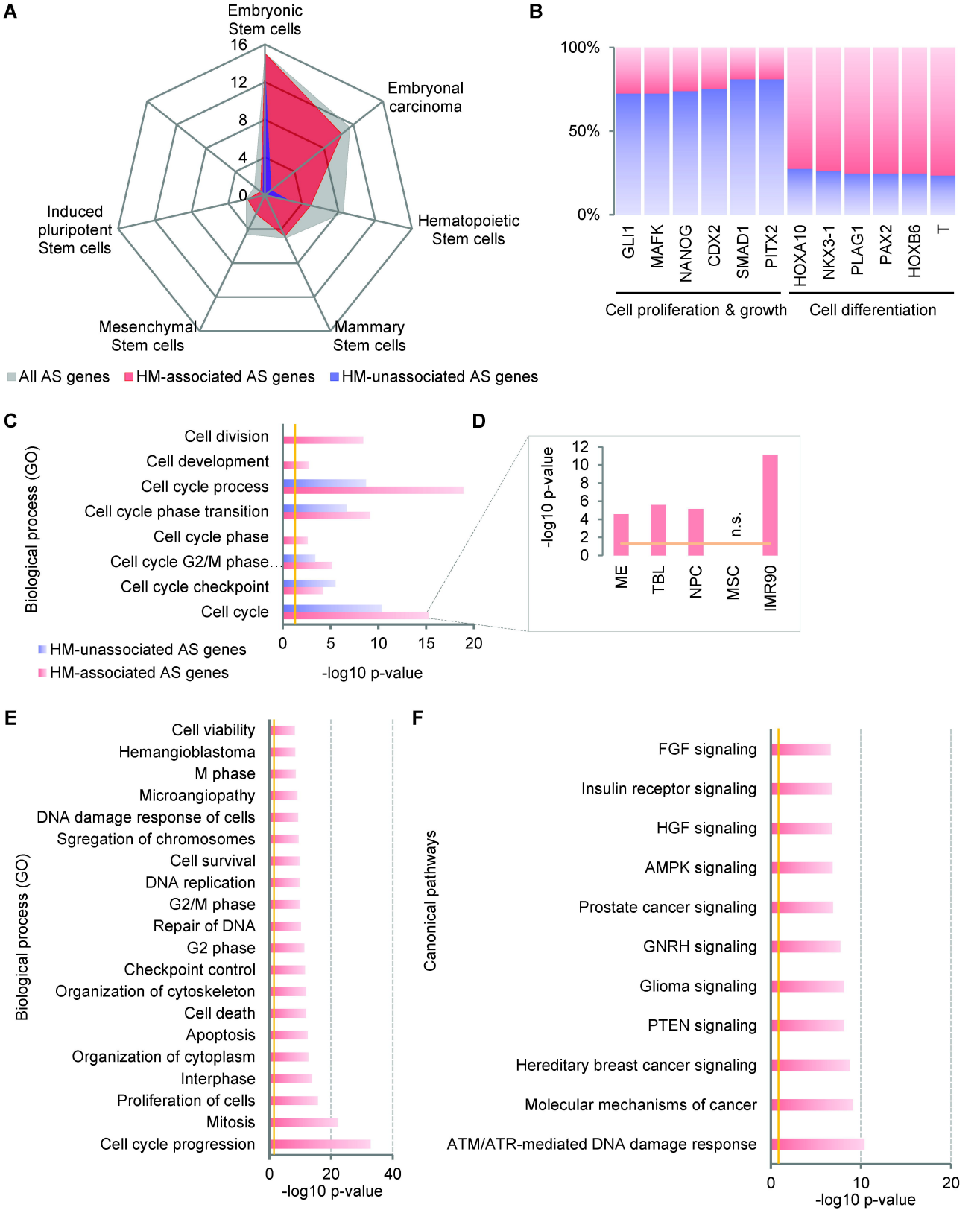
Figure 5. Isoform switch from PBX1a and PBX1b during hESC differentiation. (A) Genome browser view shows the AS event and H3K36me3 signals of PBX1 upon hESC differentiation. The green horizontal bars below the ChIP-seq tracks indicate the narrow peaks called by MACS2. **(B)** The inclusion level for 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 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. **(E)** The expression levels of NANOG and OCT4 genes are negatively correlated with the expression of PBX1b. **(F)** The expression levels of PSIP1 and SRSF1 show significant positive correlations with the expression level of PBX1a. Also, see **Figure S9-S10**.

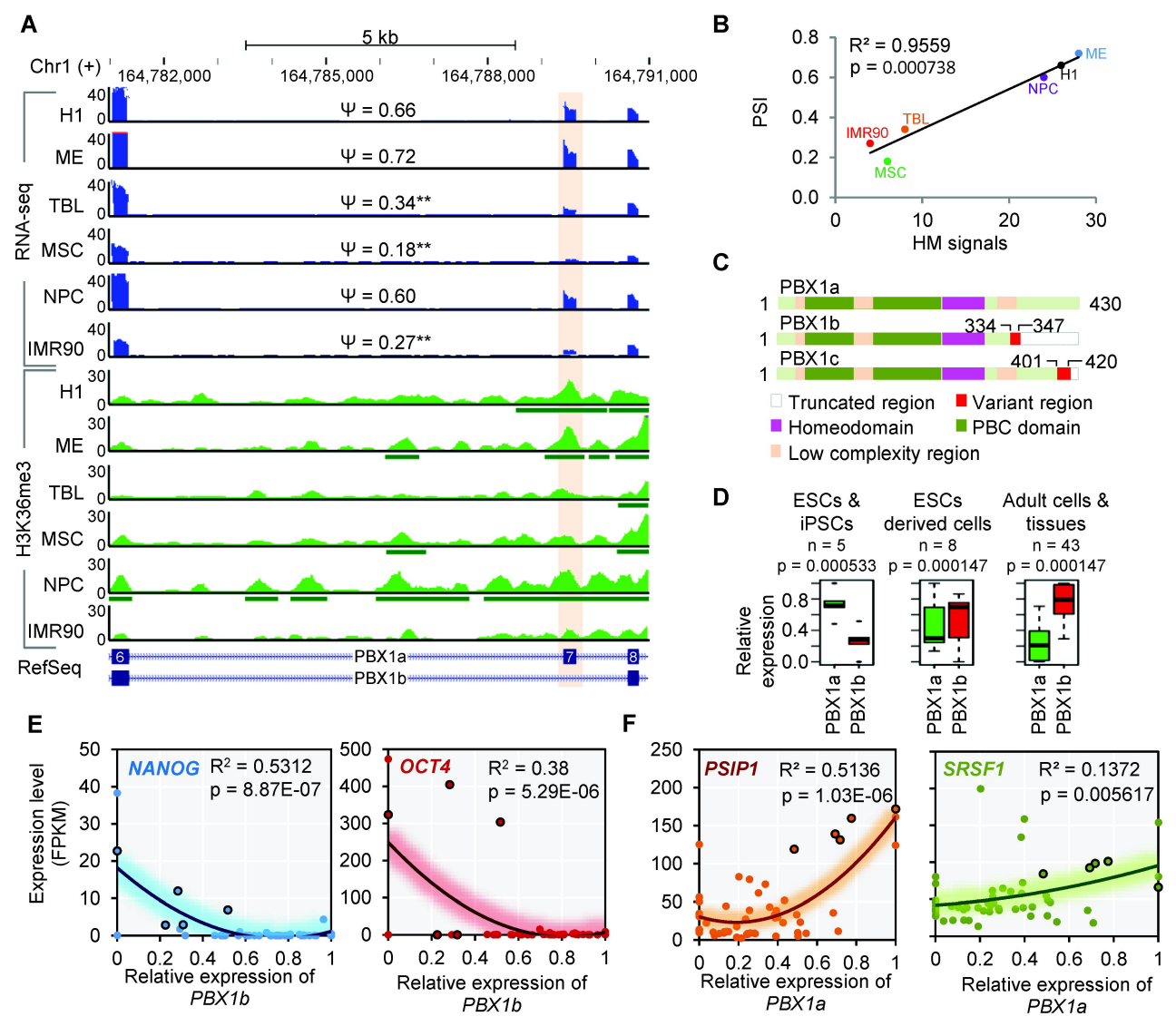
Figure 6. Isoform switch of PBX1 links H3K36me3 to hESC fate decision. (A) qRT-PCR and western blot show the expression levels of Yamanaka factors in H1, MSC and IMR90 cells. Whiskers denote the standard deviations of three replicates. **(B)** RT-PCR and western blot show the isoform switches between PBX1a and PBX1b from H1 cells to differentiated cells. **(C) i.** ChIP-PCR shows the differential binding of PBX1b to NANOG promoter in H1 cells and differentiated cells; **ii.** ChIP-PCR shows the reduced H3K36me3 signal in differentiated cells; iii. ChIP-PCR shows the differential recruitment of PSIP1 to the exon 7 of PBX1. **(D)** RIP-PCR show the differential recruitment of SRSF1 around the exon 7 of PBX1. **(E)** Co-IP shows the overall physical interaction between PSIP1 and SRSF1 in all studied cell types. **(F)** The mechanism by which H3K36me3 is linked to cell fate decision by regulating the isoform switch of PBX1, which functions upstream of the pluripotency regulatory network. Also, see **Figures S9-S10**.

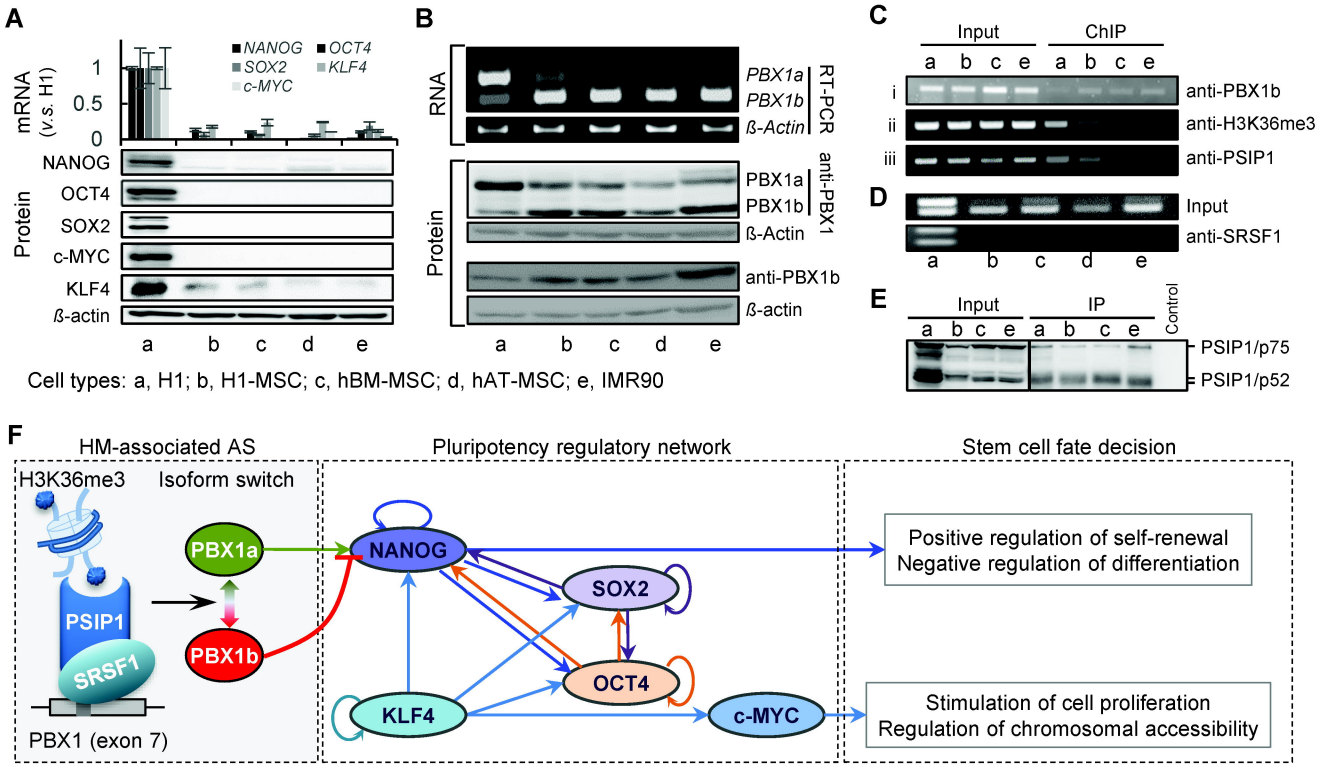

A
FOXP1 (MXE event)
FPKM
H1
ME
TBL
MSC
NPC
IMR90
Ψ = 0.71 [0.57, 0.83]
Ψ = 0.80 [0.65, 0.91]
Ψ = 0.04 [0.00, 0.12]
Ψ = 0.01 [0.00, 0.04]
Ψ = 0.14 [0.08, 0.22]
Ψ = 0.01 [0.00, 0.02]
71019890
71021430
71023733
71026189
Chr3(-)
PSI (Ψ)
B
CTNND1 (SE event)
Ψ = 0.60 [0.54, 0.66]
Ψ = 0.67 [0.63, 0.72]
Ψ = 0.21 [0.19, 0.23]
Ψ = 0.85 [0.80, 0.89]
Ψ = 0.93 [0.90, 0..95]
Ψ = 0.60 [0.40, 0.82]
57529244
57539747
57551344
57561544
Chr11(+)
PSI (Ψ)
C
MXE
SE
MXE.sp
SE.sp
AS event numbers (x1000)
% of lineage-specific AS events
ME
TBL
MSC
NPC
IMR90
AS event numbers
CRF
coTF
TF
D
ME TBL MSC NPC IMR90
SE events (n = 3678)
MXE events (n = 3513)
ΔPSI
Inclusion-gain
Inclusion-loss
E
AS genes of MXE events
Shared (11)
H1
IMR90 (1316)
NPC (168)
TBL (114)
MSC (76)
ME (38)
F
AS genes of SE events
Shared (38)
H1
IMR90 (440)
NPC (549)
TBL (111)
MSC (154)
ME (134)


A
FGFR2 (MXE event)
2kb
Chr10:123,274,000
123,276,000
123,278,000
123,280,000
H3K36me3
H1
ME
TBL
MSC
NPC
IMR90
Refseq
B
CDC27 (SE event)
5kb
Chr17:45,245,000
45,250,000
45,255,000
45,260,000
C
MXE
SE
H3K36me3
H3K4me1
H3K4me2
H3K79me1
p = 2.36E-38
p = 2.09E-15
p = 8.61E-41
p = 2.14E-14
p = 9.02E-34
p = 1.77E-17
p = 7.14E-41
p = 2.72E-16
Δ reads number
AS exons
CS exons
D
ME
TBL
MSC
NPC
IMR90
Pooled
MXE
SE
H2AK5ac
H2BK120ac
H2BK5ac
H3K18ac
H3K23ac
H3K27ac
H3K27me3
H3K36me3
H3K4ac
H3K4me1
H3K4me2
H3K4me3
H3K79me1
H3K9ac
H3K9me3
H4K8ac
-log10 p-value
p ≤ 0.001
0.001 < p ≤0.01
p > 0.05
0.01 < p ≤0.05
-log10 p-value
0
40


A
H3K36me3 profiles
p = 1.03E-05
p = 3.56E-21
Δ reads number
Inclusion-gain AS exons
Inclusion-loss AS exons
Random CS exons
B
Pearson correlation test
MXE
n = 3105
SE
n = 3516
Differential signal (ΔHM)
p = 0.0164
p = 2.22E-16
Differential inclusion (ΔPSI)
Inclusion-gain
Inclusion-loss
C
Differential signal (ΔHM)
H2BK120ac
H3K18ac
H3K27me3
H3K36me3
H3K4me1
H3K4me2
H3K9me3
H4K8ac
n = 405
n = 259
ΔPSI
Differential inclusion (ΔPSI)
D
H3K27ac-associated (positively) AS exons
R² = 0.801
p < 2.2E-16
n = 24 + 75
Differential signal (ΔHM)
Differential inclusion (ΔPSI)
ME
TBL
MSC
NPC
IMR90


A
Embryonic Stem cells
Embryonal carcinoma
Hematopoietic Stem cells
Mammary Stem cells
Mesenchymal Stem cells
Induced pluripotent Stem cells
16
12
8
4
0
All AS genes
HM-associated AS genes
HM-unassociated AS genes
B
100%
50%
0%
GLI1
MAFK
NANOG
CDX2
SMAD1
PITX2
HOXA10
NKX3-1
PLAG1
PAX2
HOXB6
T
Cell proliferation & growth
Cell differentiation
C
Biological process (GO)
Cell division
Cell development
Cell cycle process
Cell cycle phase transition
Cell cycle phase
Cell cycle G2/M phase...
Cell cycle checkpoint
Cell cycle
0
5
10
15
20
-log10 p-value
HM-unassociated AS genes
HM-associated AS genes
D
-log10 p-value
12
10
8
6
4
2
0
ME
TBL
NPC
MSC
n.s.
IMR90
E
Biological process (GO)
Cell viability
Hemangioblastoma
M phase
Microangiopathy
DNA damage response of cells
Sgregation of chromosomes
Cell survival
DNA replication
G2/M phase
Repair of DNA
G2 phase
Checkpoint control
Organization of cytoskeleton
Cell death
Apoptosis
Organization of cytoplasm
Interphase
Proliferation of cells
Mitosis
Cell cycle progression
0
20
40
-log10 p-value
F
Canonical pathways
FGF signaling
Insulin receptor signaling
HGF signaling
AMPK signaling
Prostate cancer signaling
GNRH signaling
Glioma signaling
PTEN signaling
Hereditary breast cancer signaling
Molecular mechanisms of cancer
ATM/ATR-mediated DNA damage response
0
10
20
-log10 p-value


A
5 kb
Chr1 (+)
164,782,000
164,785,000
164,788,000
164,791,000
RNA-seq
H1
ME
TBL
MSC
NPC
IMR90
Ψ = 0.66
Ψ = 0.72
Ψ = 0.34**
Ψ = 0.18**
Ψ = 0.60
Ψ = 0.27**
H3K36me3
RefSeq
PBX1a
PBX1b
B
R² = 0.9559
p = 0.000738
PSI
HM signals
ME
H1
NPC
TBL
IMR90
MSC
C
PBX1a
PBX1b
PBX1c
430
334
347
401
420
Truncated region
Variant region
Homeodomain
PBC domain
Low complexity region
D
ESCs & iPSCs
n = 5
p = 0.000533
ESCs derived cells
n = 8
p = 0.000147
Adult cells & tissues
n = 43
p = 0.000147
Relative expression
PBX1a
PBX1b
E
NANOG
R² = 0.5312
p = 8.87E-07
OCT4
R² = 0.38
p = 5.29E-06
Expression level (FPKM)
Relative expression of PBX1b
F
PSIP1
R² = 0.5136
p = 1.03E-06
SRSF1
R² = 0.1372
p = 0.005617
Relative expression of PBX1a


A
mRNA (v.s. H1)
NANOG
OCT4
SOX2
KLF4
c-MYC
Protein
NANOG
OCT4
SOX2
c-MYC
KLF4
ß-actin
a b c d e
B
RNA
PBX1a
PBX1b
ß-Actin
RT-PCR
Protein
PBX1a
PBX1b
ß-Actin
anti-PBX1
anti-PBX1b
ß-actin
a b c d e
C
Input
ChIP
a b c e a b c e
i
ii
iii
anti-PBX1b
anti-H3K36me3
anti-PSIP1
D
Input
anti-SRSF1
a b c d e
E
Input
IP
Control
a b c e a b c e
PSIP1/p75
PSIP1/p52
Cell types: a, H1; b, H1-MSC; c, hBM-MSC; d, hAT-MSC; e, IMR90
F
HM-associated AS
Pluripotency regulatory network
Stem cell fate decision
H3K36me3
Isoform switch
PSIP1
SRSF1
PBX1 (exon 7)
PBX1a
PBX1b
NANOG
SOX2
OCT4
KLF4
c-MYC
Positive regulation of self-renewal
Negative regulation of differentiation
Stimulation of cell proliferation
Regulation of chromosomal accessibility