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

Yungang Xu\textsuperscript{1,2}, Guangxu Jin\textsuperscript{2,3}, Liang Liu\textsuperscript{2}, Dongmin Guo\textsuperscript{2}, Xiaobo Zhou\textsuperscript{1,2,3,*}

\textsuperscript{1}Center for Systems Medicine, School of Medical Bioinformatics, University of Texas Health Science Center at Houston, TX, 77030, USA
\textsuperscript{2}Center for Bioinformatics and Systems Biology, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA
\textsuperscript{3}Wake Forest Baptist Comprehensive Cancer Center, Winston-Salem, NC 27157, USA

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

Email addresses:
YX: yuxu@wakehealt.edu
LL: lliu@wakehealth.edu
DG: dguo@wakehealth.edu
GJ: gjin@wakehealth.edu
XZ: xizhou@wakehealth.edu

Running title: AS links histone modification to ESC fate decision

Keywords: alternative splicing, histone modification, embryonic stem cell, cell fate decision, cell cycle machinery
Figure S1 Identifying hESC differentiation-related AS exons. (A) The schematic of hESC differentiation and a summary of the histological relations between the six cell types considered in this study. Pair-wised comparing differentiated cells with hESCs represents multiple differentiation lineages. (B) The identification of AS exons upon hESC differentiation. The AS exons were identified if their changes of ‘percent spliced in’ (ΔPSIs) are greater than 0.1 (inclusion-loss) or smaller than -0.1 (inclusion-gain) between H1 and differentiated cells with the FDRs are less than 5%. (C) The statistics of hESC differentiation-related AS events. TF, transcription factor; coTF, transcription co-factor; CRF, chromatin remodeling factor; HK, housekeeping gene. Gain or loss, inclusion-gain or -loss AS events; MXE.sp and SE.sp, lineage-specific MXE and SE events, respectively; %sp, percentage of lineage-specific AS events. (D) The majority of genes hosting the AS exons are not differentially expressed between H1 cells and differentiated cells (fold change < 2); p-values show the significances based on hypergeometric test; N, the number of involved AS events; up or down, the number of AS events that their hosting genes are up- or down-regulated upon hESC differentiation.

Related to Figure 1.
Figure S1

A

hESC (H1)

hESC

Mesenchymal stem cell (MSC)

Trophoblast-like cell (TBL)

Neural progenitor cell (NPC)

Mesendoderm (ME)

Fetal lung fibroblast (IMR90)

Mesenchymal stem cell (MSC)

Embryonic cells

Ectoderm

Mesoderm

Endoderm

Embryo

Embryonic cells

B

hESC

Differentiated cell

Identification of AS exons

Inclusion-loss

ΔPSI ≥ 0.1

FDR ≤ 5%

Inclusion-gain

ΔPSI ≤ -0.1

ΔPSI = PSI_{hESC} - PSI_{diff. cell}

C

<table>
<thead>
<tr>
<th>Cell lineages</th>
<th>#Reads</th>
<th>MXE*</th>
<th>TF</th>
<th>coTF</th>
<th>CRF</th>
<th>HK</th>
<th>Bivalent</th>
<th>Gain</th>
<th>Loss</th>
<th>MXE.sp</th>
<th>%sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ME</td>
<td>770M</td>
<td>177</td>
<td>13</td>
<td>3</td>
<td>76</td>
<td>4</td>
<td>4</td>
<td>102</td>
<td>75</td>
<td>80</td>
<td>0.452</td>
</tr>
<tr>
<td>--TBL</td>
<td>709M</td>
<td>426</td>
<td>36</td>
<td>10</td>
<td>1164</td>
<td>36</td>
<td>4</td>
<td>246</td>
<td>180</td>
<td>217</td>
<td>0.509</td>
</tr>
<tr>
<td>--MSC</td>
<td>775M</td>
<td>398</td>
<td>28</td>
<td>7</td>
<td>148</td>
<td>39</td>
<td>3</td>
<td>209</td>
<td>189</td>
<td>157</td>
<td>0.394</td>
</tr>
<tr>
<td>--NPC</td>
<td>806M</td>
<td>589</td>
<td>35</td>
<td>13</td>
<td>8</td>
<td>210</td>
<td>27</td>
<td>344</td>
<td>245</td>
<td>360</td>
<td>0.611</td>
</tr>
<tr>
<td>IMR90</td>
<td>319M</td>
<td>2583</td>
<td>124</td>
<td>50</td>
<td>31</td>
<td>1133</td>
<td>137</td>
<td>1588</td>
<td>995</td>
<td>2211</td>
<td>0.856</td>
</tr>
<tr>
<td>H1</td>
<td>780M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*AS event numbers, of which MXE involves pairwise exons, but only the upstream exons are considered in following analysis;
#Reads: the sum of reads number for 2 replicates; AS events are subject to |ΔPSI| ≥ 0.1, FDR ≤ 5%.

D

Expression of AS genes in differentiated cells (log2 FPKM)

Expression of AS genes in H1 cells (log2 FPKM)

Inclusion-loss

Inclusion-gain

Expression of AS genes in differentiated cells (log2 FPKM)

Expression of AS genes in H1 cells (log2 FPKM)

Expression of AS genes in H1 cells (log2 FPKM)

Expression of AS genes in H1 cells (log2 FPKM)

Expression of AS genes in H1 cells (log2 FPKM)

Expression of AS genes in H1 cells (log2 FPKM)

Expression of AS genes in H1 cells (log2 FPKM)
Figure S2 The hESC differentiation-related AS exons possess the typical properties of AS exons. (A) The average length of AS exons is much shorter than that of all exons based on RefSeq annotation. (B) The average length of adjoining introns of AS exons is much longer than that of all introns based on RefSeq annotation. (C) The length of AS exons is much shorter than that of the neighboring up- or down-stream constitutive exons across all cell lineages and AS types. (D) The length of introns between MXE exons is much longer than that of the introns aside the AS exons and that of all introns based on RefSeq annotation. The introns surrounding SE exons are much longer than the average intron length base on RefSeq. (E) The length of AS exons is more often divisible by three than flanking constitutive splicing (CS) exons or all RefSeq exons.

Related to Figure 1.
Figure S2

A. Exon types

B. Intron types

C. Exon length comparison across different cell lines

D. Intron length comparison across different cell lines

E. Fraction of exons divisible by three

* p < 2.2E-16 (Student’s t-test)
Figure S3 AS profiles upon hESC differentiation show lineage-specific splicing pattern. (A) Enrichment analysis of all AS genes based on a manually curated stemness signature gene set, showing hESC differentiation-related AS genes are significantly associated with stemness, especially, of ESCs (-log10 Bonferroni adjusted p-values). (B) and (C) Venn graphs show the overlaps of MXE and SE events between different cell lineages. Numbers indicate the count of events for each part and percentage in parentheses show the lineage-specific percentages. (D) and (E) Venn graphs showing the overlapping of MXE and SE AS genes between different cell lineages. Numbers indicate the number of genes for each part and percentage in parentheses show the lineage-specific percentages. (F) Pairwise overlaps of AS events (uper) or AS genes (lower) between different cell lineages. (G) The overlaps between SE and MXE AS genes, as well as the overlap between the inter-lineage shared MXE and SE AS genes. Genes in red have been reported that their AS events are involved in mouse ESC differentiation. (H) Few SF genes are differentially expressed between H1 cells and differentiated cells (fold change < 2); p-values show the significances based on hypergeometric test; N, the number of selected SF genes; up or down, the number of SF genes that are up- or down-regulated upon hESC differentiation.

Related to Figure 1.
Figure S3

A

Expression of SF genes in H1 cells (log2 FPKM)

B

Pairwise overlaps (AS events)

C

Pairwise overlaps (AS genes)

D

MXE: n = 2257

E

SE: n = 2489

F

Pairwise overlaps (AS events)

G

Pairwise overlaps (AS genes)

H

Expression of SF genes in differentiated cells

Expression of SF genes in H1 cells (log2 FPKM)

**Differentially expressed SF genes**

**Non-differentially expressed SF genes**

**Other genes (background)**
Figure S4 Histone modifications (HMs) change significantly around the alternatively spliced (AS) exons upon hESC differentiation. (A-F) The global profiles of HM changes (normalized Δ reads number) around the AS exons and randomly selected constitutive splicing (CS) exons during the differentiation from H1 ESCs to ME (A), TBL (B), MSC (C), NPC (D), or compared to IMR90 (E), as well as pooled them together (F); ±150bp regions of the splice sites (exons-intron boundaries) were considered and 15bp-binned to produce the curves. The p-values with red shading indicate no significant difference between AS and constitutive exons; Red-shaded panels indicate the cases that HMs change more significantly around constitutive exons rather than around the AS exons. p-values, Mann-Whitney-Wilcoxon test.

Related to Figure 2.
Figure S4

A  ME

<table>
<thead>
<tr>
<th></th>
<th>MXE</th>
<th>SE</th>
<th></th>
<th>MXE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3KAc</td>
<td>p = 2.53E-05</td>
<td>p = 0.0126</td>
<td></td>
<td>p = 7.47E-08</td>
<td>p = 0.0036</td>
</tr>
<tr>
<td>H3K36me3</td>
<td>p = 0.00116</td>
<td>p = 0.294</td>
<td>p = 0.0016</td>
<td>p = 1.33E-05</td>
<td>p = 0.0126</td>
</tr>
<tr>
<td>Δ reads number</td>
<td>p = 7.29E-41</td>
<td>p = 3.55E-21</td>
<td></td>
<td>p = 0.0722</td>
<td>p = 2.53E-05</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>p = 0.417</td>
<td>p = 0.35</td>
<td>p = 5.68E-08</td>
<td>p = 0.0126</td>
<td>p = 0.0126</td>
</tr>
<tr>
<td>Δ reads number</td>
<td>p = 2.97E-15</td>
<td>p = 3.71E-21</td>
<td></td>
<td>p = 0.103</td>
<td>p = 0.0126</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>p = 0.727</td>
<td>p = 0.0036</td>
<td></td>
<td>p = 5.68E-08</td>
<td>p = 0.0126</td>
</tr>
</tbody>
</table>

B  TBL

<table>
<thead>
<tr>
<th></th>
<th>MXE</th>
<th>SE</th>
<th></th>
<th>MXE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3KAc</td>
<td>p = 0.00663</td>
<td>p = 0.00192</td>
<td></td>
<td>p = 1.68E-05</td>
<td>p = 3.55E-21</td>
</tr>
<tr>
<td>H3K36me3</td>
<td>p = 3.98E-19</td>
<td>p = 1.18E-15</td>
<td></td>
<td>p = 2.35E-18</td>
<td>p = 7.95E-20</td>
</tr>
<tr>
<td>Δ reads number</td>
<td>p = 3.58E-06</td>
<td>p = 1.48E-14</td>
<td></td>
<td>p = 0.44</td>
<td>p = 0.954</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>p = 2.64E-18</td>
<td>p = 4.32E-17</td>
<td></td>
<td>p = 1.18E-28</td>
<td>p = 1.06E-19</td>
</tr>
<tr>
<td>Δ reads number</td>
<td>p = 1.89E-06</td>
<td>p = 8.27E-05</td>
<td></td>
<td>p = 1.75E-13</td>
<td>p = 1.97E-19</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>p = 0.143</td>
<td>p = 0.469</td>
<td>p = 1.75E-13</td>
<td>p = 0.274</td>
<td>p = 0.883</td>
</tr>
<tr>
<td>Δ reads number</td>
<td>p = 5.29E-11</td>
<td>p = 7.85E-21</td>
<td></td>
<td>p = 5.35E-22</td>
<td>p = 7.64E-13</td>
</tr>
</tbody>
</table>

AS exons - black
CS exons - red
Figure S4 continue.

C

MSC

<table>
<thead>
<tr>
<th></th>
<th>MXE</th>
<th>SE</th>
<th></th>
<th>MXE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K4ac</td>
<td>p = 1.3E-35</td>
<td>p = 9.18E-09</td>
<td></td>
<td>p = 1.44E-18</td>
<td>p = 5.11E-13</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>p = 0.188</td>
<td></td>
<td></td>
<td>p = 1.09E-07</td>
<td>p = 5.4E-08</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>p = 0.00396</td>
<td></td>
<td></td>
<td>p = 1.17E-05</td>
<td>p = 3.35E-06</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>p = 0.694</td>
<td></td>
<td></td>
<td>p = 1.17E-05</td>
<td>p = 3.35E-06</td>
</tr>
<tr>
<td>H3K4ac</td>
<td>p = 2.77E-25</td>
<td>p = 6.6E-15</td>
<td></td>
<td>p = 2.77E-25</td>
<td>p = 6.6E-15</td>
</tr>
<tr>
<td></td>
<td>p = 2.13E-34</td>
<td></td>
<td></td>
<td>p = 2.13E-34</td>
<td></td>
</tr>
</tbody>
</table>

D

NPC

<table>
<thead>
<tr>
<th></th>
<th>MXE</th>
<th>SE</th>
<th></th>
<th>MXE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K4ac</td>
<td>p = 3.34E-19</td>
<td>p = 1.12E-15</td>
<td></td>
<td>p = 2.34E-09</td>
<td>p = 1.2E-09</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>p = 3.34E-19</td>
<td>p = 1.12E-15</td>
<td></td>
<td>p = 3.34E-19</td>
<td>p = 1.2E-09</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>p = 3.34E-19</td>
<td>p = 1.12E-15</td>
<td></td>
<td>p = 3.34E-19</td>
<td>p = 1.2E-09</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>p = 3.34E-19</td>
<td>p = 1.12E-15</td>
<td></td>
<td>p = 3.34E-19</td>
<td>p = 1.2E-09</td>
</tr>
<tr>
<td></td>
<td>p = 2.77E-25</td>
<td>p = 6.6E-15</td>
<td></td>
<td>p = 2.77E-25</td>
<td>p = 6.6E-15</td>
</tr>
<tr>
<td></td>
<td>p = 2.13E-34</td>
<td></td>
<td></td>
<td>p = 2.13E-34</td>
<td></td>
</tr>
</tbody>
</table>

AS exons  CS exons
Figure S4 continue.

E  IMR90

<table>
<thead>
<tr>
<th></th>
<th>MXE</th>
<th>SE</th>
<th></th>
<th>MXE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2AK5ac</td>
<td>p = 7.49E-41</td>
<td>p = 3.55E-21</td>
<td>H3K18ac</td>
<td>p = 2.73E-24</td>
<td>p = 3.55E-21</td>
</tr>
<tr>
<td>H2BK5ac</td>
<td>p = 7.14E-41</td>
<td>p = 3.55E-21</td>
<td>H2BK120ac</td>
<td>p = 5.75E-24</td>
<td>p = 3.55E-21</td>
</tr>
<tr>
<td>H3K4ac</td>
<td>p = 2.92E-19</td>
<td>p = 3.55E-21</td>
<td>H3K4me1</td>
<td>p = 1.53E-39</td>
<td>p = 3.55E-21</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>p = 7.14E-41</td>
<td>p = 3.55E-21</td>
<td>H3K4me3</td>
<td>p = 7.55E-17</td>
<td>p = 3.55E-21</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>p = 9.02E-34</td>
<td>p = 3.55E-21</td>
<td>H4K8ac</td>
<td>p = 1.61E-14</td>
<td>p = 3.55E-21</td>
</tr>
<tr>
<td></td>
<td>p = 2.71E-16</td>
<td>p = 3.55E-21</td>
<td></td>
<td>p = 3.67E-12</td>
<td>p = 3.55E-21</td>
</tr>
</tbody>
</table>

F  Pooled

<table>
<thead>
<tr>
<th></th>
<th>MXE</th>
<th>SE</th>
<th></th>
<th>MXE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2AK5ac</td>
<td>p = 9.87E-41</td>
<td>p = 1.08E-19</td>
<td>H3K18ac</td>
<td>p = 4.49E-21</td>
<td>p = 3.41E-14</td>
</tr>
<tr>
<td>H3K14ac</td>
<td>p = 1.61E-08</td>
<td>p = 7.18E-15</td>
<td>H3K23ac</td>
<td>p = 8.61E-41</td>
<td>p = 2.14E-14</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>p = 2.39E-38</td>
<td>p = 2.71E-13</td>
<td>H3K36me3</td>
<td>p = 9.02E-34</td>
<td>p = 1.77E-17</td>
</tr>
<tr>
<td>H2BK5ac</td>
<td>p = 2.36E-38</td>
<td>p = 2.09E-15</td>
<td>H2BK120ac</td>
<td>p = 7.14E-41</td>
<td>p = 8.42E-17</td>
</tr>
<tr>
<td>H3K4ac</td>
<td>p = 0.125</td>
<td>p = 0.0554</td>
<td>H3K4me1</td>
<td>p = 9.02E-34</td>
<td>p = 1.24E-11</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>p = 1.09E-24</td>
<td>p = 0.0203</td>
<td>H3K4me3</td>
<td>p = 0.737</td>
<td>p = 1.24E-11</td>
</tr>
<tr>
<td>H3K79me1</td>
<td>p = 1.09E-24</td>
<td>p = 0.0554</td>
<td>H3K9ac</td>
<td>p = 4.07E-04</td>
<td>p = 3.67E-12</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>p = 0.145</td>
<td>p = 3.55E-21</td>
<td>H4K8ac</td>
<td>p = 7.19E-20</td>
<td>p = 2.71E-20</td>
</tr>
<tr>
<td></td>
<td>p = 1.34E-34</td>
<td>p = 4.56E-21</td>
<td></td>
<td>p = 7.19E-20</td>
<td>p = 2.71E-20</td>
</tr>
</tbody>
</table>
Figure S5 A subset of AS events is significantly associated with some HMs upon hESC differentiation. (A) The differential ChIP-seq signal of a HM ($\Delta$HM) was defined as the difference of the summit heights of narrow peaks between two cell types, which were normalized by the distances (in kbs) from the peak summit to the 3’ splice site. (B) The statistic significances of changes for all 16 HMs in each cell lineage and pooling them together (pooled), represented as the -log10 p-values based on Mann-Whitney-Wilcoxon test between the HM profiles of inclusion-gain and inclusion–loss AS exons. The side bars represent the significances whether the changes of HMs are consistently differentially enriched in inclusion-gain and inclusion–loss AS exons across cell lineages, showing the link strength between AS and HMs as the -log10 p-value based on Fisher’s exact test. The yellow vertical line indicates the significance cutoff of 0.05. (C) Three correlation test methods were used to preselect the HMs that may associate with AS; ‘√’ indicates the HMs passing the test ($p \leq 0.05$) and will be considered for further quantitative analysis. PC, Pearson correlation; MLR, multiple linear regression; LLR, logistic regression. (D) A representative k-means clustering result of SE ‘inclusion-gain’ exons based on selected epigenetic features, showing one of the six clusters exhibits negative correlation between the differential H4K8ac signal and the inclusion level changes of 89 SE exons. (E) The statistic of the AS exons with at least one HM changes upon differentiation ($|\Delta$HM| > 0) and the HM-associated exons. The % in parentheses shows the percentage against the total AS exon number. (F) and (G) The overlaps between H3K36me3-associated AS genes and those from other literatures showing the potential mechanism by which H3K36me3 takes the role in AS regulation via proper chromatin-adapter systems. The indicated references are given in the main text.

Related to Figure 3.
Figure S5

A

Peak in cell type 1 (e.g. H1) Peak in cell type 2 (e.g. ME)

$\Delta HM = h_1/d_1 - h_2/d_2$

B

Selected

C

<table>
<thead>
<tr>
<th>MXE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC</td>
<td>MLR</td>
</tr>
<tr>
<td>H2AK5ac</td>
<td>0.0085</td>
</tr>
<tr>
<td>H2BK120ac</td>
<td>0.0281</td>
</tr>
<tr>
<td>H3K18ac</td>
<td>0.04346</td>
</tr>
<tr>
<td>H3K23ac</td>
<td>0.00860380</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>0.03921</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>0.01640025</td>
</tr>
<tr>
<td>H3K36me3</td>
<td>0.03104</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>0.0301</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>0.0480</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>0.0102</td>
</tr>
<tr>
<td>H3K9ac</td>
<td>0.0097</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>0.0010</td>
</tr>
<tr>
<td>H4K8ac</td>
<td>0.028800172</td>
</tr>
</tbody>
</table>

E

<table>
<thead>
<tr>
<th></th>
<th>Total AS exons</th>
<th>At least 1 HM changes</th>
<th>HM-associated AS exons</th>
</tr>
</thead>
<tbody>
<tr>
<td>MXE</td>
<td>3513</td>
<td>3194 (90.9%)</td>
<td>1891 (53.8%)</td>
</tr>
<tr>
<td>Pooled</td>
<td>7191</td>
<td>6861 (95.4%)</td>
<td>3797 (52.8%)</td>
</tr>
</tbody>
</table>

F

SRSF1-regulated AS genes

Olga Anczuków, et al., 2015

H3K36me3-associated AS genes (positive correlation)

Shatakshi Pandit, et al., 2013

G

PTB-regulated AS genes

(Yuanchao Xue, et al., 2009)

H3K36me3-associated AS genes (negative correlation)
Figure S6 K-means clustering based on selected epigenetic features of eight HMs for MXE and SE AS exons. Each panel represents one subset of exons identified by k-means clustering, showing the boxplots of differential signals (ΔHMs) and corresponding differential inclusion levels (ΔPSIs). The number of exons present in each cluster is indicated in the boxplot, with blue refers to the inclusion-loss exons and red the inclusion-gain exons.

Related to Figure 3.
Figure S7 HM-associated AS genes are more lineage-specific. (A) The length of HM-associated AS exons are much longer than that of HM-unassociated AS exons. p-values, student’s t-test. (B) Venn graph shows the overlap of HM-associated AS genes across all cell lineages, indicating more lineage specificity compared with all AS genes shown in Figure S3D, E. (C) Venn graph shows the overlap of HM-associated AS genes across cell lineages excluding the IMR90. (D) Pairwise overlaps of HM-associated AS genes across cell lineages. The percentages (%) in parentheses show the ratio between the number of pairwise intersection and union, which are used to define the shading darkness. (E) The most common HM-associated AS genes shared by all lineages. The last two underlined genes are the additional ones not shared by IMR90.

Related to Figure 4.
**Table E**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>UniProtKB/Swiss-Prot function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BARD1</strong></td>
<td>BRCA1 Associated RING Domain 1</td>
<td>This gene encodes a protein which interacts with the N-terminal region of BRCA1. Plays a central role in the control of the cell cycle in response to DNA damage.</td>
</tr>
<tr>
<td><strong>NASP</strong></td>
<td>Nuclear Autoantigenic Sperm Protein (Histone-Binding)</td>
<td>H1 histone binding protein that is involved in transporting histones into the nucleus of dividing cells. Required for DNA replication, normal cell cycle progression and cell proliferation.</td>
</tr>
<tr>
<td><strong>MARK2</strong></td>
<td>MAP/Microtubule Affinity-Regulating Kinase 2</td>
<td>Subunit of the splicing factor SF3A required for A complex assembly formed by the stable binding of U2 snRNP to the branchpoint sequence (BPS) in pre-mRNA.</td>
</tr>
<tr>
<td><strong>SF3A3</strong></td>
<td>Splicing Factor 3a, Subunit 3, 60kDa</td>
<td>Serine/threonine-protein kinase involved in cell polarity and microtubule dynamics regulation.</td>
</tr>
<tr>
<td><strong>TRAM1</strong></td>
<td>Translocation Associated Membrane Protein 1</td>
<td>E3 ubiquitin-protein ligase which mediates ubiquitination and subsequent proteasomal degradation of target proteins. Involved in chromosome segregation and cell cycle regulation.</td>
</tr>
<tr>
<td><strong>TRIM36</strong></td>
<td>Tripartite Motif Containing 36</td>
<td>This gene encodes an enzyme operative in the beta-oxidation system of the peroxisomes.</td>
</tr>
<tr>
<td><strong>ACAA1</strong></td>
<td>Acetyl-CoA Acyltransferase 1</td>
<td>A protein coding gene. GO annotations related to this gene include poly(A) RNA binding.</td>
</tr>
<tr>
<td><strong>PNISR</strong></td>
<td>PNN-Interacting Serine/Arginine-Rich Protein</td>
<td>Subunit of the splicing factor SF3A required for A complex assembly formed by the stable binding of U2 snRNP to the branchpoint sequence (BPS) in pre-mRNA.</td>
</tr>
</tbody>
</table>
Figure S8 HM-unassociated AS genes are enriched in G1 cell-cycle phase and pathways for self-renewal. (A) Gene ontology (GO) enrichment analysis shows HM-associated AS genes are more enriched in cell-cycle progression than HM-unassociated AS genes, shown as the enriched gene numbers of each subgroup. (B) The enrichments in cell cycle of HM-associated AS genes are consistent across cell lineages, with the MSC as an exception. (C) HM-unassociated AS genes involved in cell-cycle progression prefer to function in G1 phase and cell-cycle arrest, shown as the -log10 p-values after FDR (≤ %5) adjustment. (D) The enriched canonical pathways show HM-unassociated AS genes are related more with G1 phase and Wnt/β-catenin signaling, which are important for self-renewal. The vertical lines (yellow) indicate the significance cutoff of 0.05.

Related to Figure 4.
Figure S8

A

Cell cycle

Enriched gene numbers

Cell division
Cell development
Cell cycle process
Cell cycle phase transition
Cell cycle phase
Cell cycle G2/M phase transition
Cell cycle checkpoint
Cell cycle

HM-unassociated AS genes
HM-associated AS genes

B

<table>
<thead>
<tr>
<th>Lineages</th>
<th>p-value</th>
<th>FDR B&amp;H adjustment</th>
<th>FDR B&amp;Y adjustment</th>
<th>Bonferroni adjustment</th>
<th>Genes from input (# enriched/total input)</th>
<th>Genes in annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>2.808E-5</td>
<td>9.018E-3</td>
<td>7.833E-2</td>
<td>9.325E-2</td>
<td>40/224</td>
<td>1780</td>
</tr>
<tr>
<td>TBL</td>
<td>2.622E-6</td>
<td>2.701E-3</td>
<td>2.415E-2</td>
<td>1.127E-2</td>
<td>56/323</td>
<td>1780</td>
</tr>
<tr>
<td>NPC</td>
<td>7.343E-6</td>
<td>7.527E-3</td>
<td>6.836E-2</td>
<td>3.624E-2</td>
<td>81/540</td>
<td>1780</td>
</tr>
<tr>
<td>MSC</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>56/387</td>
<td>1780</td>
</tr>
<tr>
<td>IMR90</td>
<td>7.564E-12</td>
<td>1.782E-8</td>
<td>1.682E-7</td>
<td>5.346E-8</td>
<td>199/1340</td>
<td>1780</td>
</tr>
</tbody>
</table>

C

Cell cycle progression of tumor cell lines
Duplication of centrosome
Arrest in interphase
Arrest in cell cycle progression of fibroblasts
Modification of chromatin
Cell cycle progression of epithelial cells
Cell cycle progression of fibroblasts
Arrest in G1 phase
DNA damage checkpoint
Cycling of centrosome
Replication of centriole
G1/S phase transition
Checkpoint control
Interphase
G1 phase
Arrest in cell cycle progression
Mitosis
Cytokinesis
M phase
Cell cycle progression

D

Polyamine regulation in colon cancer
Cell cycle: G1/S checkpoint regulation
DNA methylation and transcriptional repression signaling
Cyclins and cell cycle regulation
Ovarian cancer signaling
HIPPO signaling
Wnt/β-catenin signaling
Epithelial adherens junction signaling
Protein kinase A signaling
Protein ubiquitination signaling
Mitotic roles of polo-like kinase
Figure S9 HM-associated cell-cycle genes are enriched in DNA damage response pathway. (A) IPA pathway view of the enriched AS cell-cycle genes in the pathway called ‘ATM/ATR-mediated DNA damage response’. The darkness of filled shapes is corresponding to the number of lineages that share the given AS gene. Some paths without any enriched genes are omitted. See Figure S10 for the profiles of the full list of these thirteen AS genes. (B) The overlap of enriched AS cell-cycle genes across lineages. (C) The enrichment in DNA damage response of HM-associated AS cell-cycle genes is consistent across cell lineages. The enriched genes for either pooling all lineages together (pooled) or considering each lineage separately were provided. An exception was observed that no AS gene from MSC enriched in this pathway.

Related to Figure 4.
**Figure S9**

A. ATM/ATR-mediated DNA damage response

Ionizing Radiation → DNA damage → S phase Checkpoint control → Complex A (BRCA1, BARD1)

Nucleus

B. Venn diagram showing enriched gene sets in different lineages.

C. Table summarizing enrichment analysis results.

<table>
<thead>
<tr>
<th>Lineages</th>
<th>p-value</th>
<th># enriched genes from Input</th>
<th># genes in annotation</th>
<th>Enriched gene names</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled</td>
<td>4.89E-10</td>
<td>13/319</td>
<td>78</td>
<td>BARD1, BLM, BRCA1, FANCA, MLH1, MSH2, MSH6, PBRM1, RAD50, RFC1, RFC2, RFC4, RPA1</td>
</tr>
<tr>
<td>ME</td>
<td>5.90E-04</td>
<td>4/40</td>
<td>78</td>
<td>BARD1, MSH2, RAD50, RFC1</td>
</tr>
<tr>
<td>TBL</td>
<td>2.15E-02</td>
<td>3/56</td>
<td>78</td>
<td>BARD1, MSH2, RFC1</td>
</tr>
<tr>
<td>MSC</td>
<td>NA</td>
<td>1/56</td>
<td>78</td>
<td>RFC4</td>
</tr>
<tr>
<td>NPC</td>
<td>1.82E-05</td>
<td>6/81</td>
<td>78</td>
<td>BARD1, BLM, BRCA1, MLH1, MSH2, PBRM1</td>
</tr>
<tr>
<td>IMR90</td>
<td>1.65E-05</td>
<td>7/199</td>
<td>78</td>
<td>BARD1, FANCA, MLH1, MSH6, RFC2, RFC4, RPA1</td>
</tr>
</tbody>
</table>
Figure S10 Thirteen HM-associated AS cell-cycle genes are enriched in ATM/ATR-mediated DNA damage response pathway. Table shows the detailed profiles of these genes. The ‘AS profiles’ column show the graphic views of PSIs for AS exons.

Related to Figure 4.
<table>
<thead>
<tr>
<th>Gene</th>
<th>AS type</th>
<th># of Exons</th>
<th>AS exons and locations</th>
<th>Shared lineages</th>
<th>AS profiles</th>
<th>HM-associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>BARD1</td>
<td>MXE</td>
<td>11</td>
<td>Exon 4: Chr2(-):215645283-215646233 Exon 3: Chr2(-):215657020-215657169</td>
<td>ME,TBL,NPC, IMR90</td>
<td>PSI</td>
<td>H3K36me3 negative</td>
</tr>
<tr>
<td>MSH2</td>
<td>SE</td>
<td>16</td>
<td>Exon 11: Chr2(+)::47698103-47698201</td>
<td>ME,TBL,NPC</td>
<td>PSI</td>
<td>H3K36me3 positive</td>
</tr>
<tr>
<td>RFC1</td>
<td>MXE</td>
<td>25</td>
<td>Exon 8: Chr4(-)::39322906-39322994 Exon 7: Chr4(+)::39324959-39325037</td>
<td>ME</td>
<td>PSI</td>
<td>H3K36me3 negative</td>
</tr>
<tr>
<td>MLH1</td>
<td>MXE</td>
<td>19</td>
<td>Exon 16: Chr3(+):37089009-37089174 Exon 17: Chr3(+):37090007-37090100</td>
<td>NPC, IMR90</td>
<td>PSI</td>
<td>H3K36me3 positive</td>
</tr>
<tr>
<td>RFC4</td>
<td>MXE</td>
<td>11</td>
<td>Exon 10: Chr3(+):186507930-186508044 Exon 9: Chr3(+):186508114-186508195</td>
<td>MSC, IMR90</td>
<td>PSI</td>
<td>H3K36me3 negative</td>
</tr>
<tr>
<td>RAD50</td>
<td>MXE</td>
<td>25</td>
<td>Exon 12: Chr5(+):131930560-131930736 Exon 13: Chr5(+):131931264-131931502</td>
<td>ME</td>
<td>PSI</td>
<td>H3K36me3 positive</td>
</tr>
<tr>
<td>BLM</td>
<td>SE</td>
<td>22</td>
<td>Exon 6: Chr15(+):91303376-91303509</td>
<td>NPC</td>
<td>PSI</td>
<td>H3K36me3 positive</td>
</tr>
<tr>
<td>BRCA1</td>
<td>SE</td>
<td>24</td>
<td>Exon 9: Chr17(+):41247862-41247939</td>
<td>NPC</td>
<td>PSI</td>
<td>H3K36me3 positive</td>
</tr>
<tr>
<td>PBRM1</td>
<td>SE</td>
<td>30</td>
<td>Exon 27: Chr3(-):52588739-52588895</td>
<td>NPC</td>
<td>PSI</td>
<td>H3K36me3 negative</td>
</tr>
<tr>
<td>FANCA</td>
<td>MXE</td>
<td>42</td>
<td>Exon 38: Chr16(-):89807211-89807274 Exon 37: Chr16(-):89809207-89809346</td>
<td>IMR90</td>
<td>PSI</td>
<td>H3K36me3 positive</td>
</tr>
<tr>
<td>MSH6</td>
<td>MXE</td>
<td>10</td>
<td>Exon 2: Chr2(+):48018065-48018262 Exon 3: Chr2(+):48023032-48023202</td>
<td>IMR90</td>
<td>PSI</td>
<td>H3K36me3 negative</td>
</tr>
<tr>
<td>RFC2</td>
<td>MXE</td>
<td>11</td>
<td>Exon 4: Chr7(-):73663341-73663448 Exon 3: Chr7(-):73664068-73664110</td>
<td>IMR90</td>
<td>PSI</td>
<td>H3K36me3 negative</td>
</tr>
<tr>
<td>RPA1</td>
<td>MXE</td>
<td>17</td>
<td>Exon 14: Chr17(+):1791968-1792145 Exon 15: Chr17(+):1795126-1795234</td>
<td>IMR90</td>
<td>PSI</td>
<td>H3K36me3 positive</td>
</tr>
</tbody>
</table>
Figure S11 H3K36me3-associated isoform switch of PBX1 is involved in hESC differentiation. (A) The sequence information of transcript and protein isoforms of PBX1. (B) The expression levels of NANOG and OCT4 genes are significantly positive correlated with the expression of PBX1a, shown as the relative expression levels. (C) The expression levels of NANOG and OCT4 genes are significantly positive correlated with the PSI of exon 7 for PBX1. (D) No significant correlations between expression levels of PTB/MRG15 and the PSI for exon 7 of PBX1 are observed. Shapes with black outlines in Figure Sb-d indicate the ESCs or iPSCs. (E) Histone modifications are linked to ESC fate decision via epigenetic association with the alternative splicing of critical pathways and transcriptional factors in a cell-cycle-dependent manner.

Related to Figures 5 and 6.
### Figure S11

#### A

<table>
<thead>
<tr>
<th>Name</th>
<th>GeneID</th>
<th>Ensembl ID</th>
<th>CCDS ID</th>
<th>UniProt</th>
<th>RefSeq ID</th>
<th>mRNA length</th>
<th>Protein length</th>
<th>Protein sequence variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBX1a / PBX1-001</td>
<td>5087</td>
<td>ENSG00000185630</td>
<td>ENST00000420696</td>
<td>ENSP00000405890</td>
<td>NM_002585</td>
<td>6918nt</td>
<td>430aa</td>
<td>Canonical sequence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NP_002576</td>
<td></td>
<td></td>
<td>Full length: 1-430 aa</td>
</tr>
<tr>
<td>PBX1b / PBX1-002</td>
<td>5087</td>
<td>ENSG00000185630</td>
<td>ENST00000367897</td>
<td>ENSP00000356872</td>
<td>NM_001204961</td>
<td>6805nt</td>
<td>347aa</td>
<td>334-347: SSSSFNMSNSGDLF→GYPSCYQPDRRIQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NP_001191890</td>
<td></td>
<td></td>
<td>Full length: 1-347aa</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>401-420: ANGGWQDATTPSSVTSPTEG→HLPHRPRQAHYHFLPTWH</td>
</tr>
<tr>
<td>PBX1c / PBX1-020</td>
<td>5087</td>
<td>ENSG00000185630</td>
<td>ENST00000627490</td>
<td>ENSP00000405892</td>
<td>NM_001204963</td>
<td>4162nt</td>
<td>420aa</td>
<td>401-420: ANGGWQDATTPSSVTSPTEG→HLPHRPRQAHYHFLPTWH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NP_001191892</td>
<td></td>
<td></td>
<td>Full length: 1-420aa</td>
</tr>
</tbody>
</table>

nt: nucleotide; aa: amino acid

#### B

- **Expression of NANO**
  - **NANO**
    - $R^2 = 0.5312$
    - $p = 8.87E-07$
  - **OCT4**
    - $R^2 = 0.38$
    - $p = 5.29E-06$

#### C

- **Relative Expression of NANO**
  - Relative expression of NANO vs. relative expression of PBX1a

- **Relative Expression of OCT4**
  - Relative expression of OCT4 vs. relative expression of PBX1a

#### D

- **FPKM of NANO**
  - FPKM vs. PSI

- **FPKM of OCT4**
  - FPKM vs. PSI

#### E

- **Alternative splicing** → **Cell-cycle progression** → **ESC fate decision**
- **Histone modifications**
- **ATM/ATR-mediated DNA damage response pathway** → **Pluripotency state dissolution (PSD)**
- **ESC self-renewal (core pluripotency regulatory network)** → **ESC differentiation**