Science Advances
AAAS

# Supplementary Materials for

## Single-cell architecture and functional requirement of alternative splicing during hematopoietic stem cell formation

Fang Wang, Puwen Tan, Pengcheng Zhang, Yue Ren, Jie Zhou, Yunqiao Li, Siyuan Hou, Shuaili Li, Linlin Zhang, Yanni Ma, Chaojie Wang, Wanbo Tang, Xiaoshuang Wang, Yue Huo, Yongfei Hu, Tianyu Cui, Chao Niu, Dong Wang*, Bing Liu*, Yu Lan*, Jia Yu*

*Corresponding author. Email: wangdong79@smu.edu.cn (D.W.); bingliu17@yahoo.com (B.L.); rainyblue_1999@126.com (Y.Lan); j-yu@ibms.pumc.edu.cn (J.Y.)

Published 7 January 2022, *Sci. Adv.* **8**, eabg5369 (2022)
DOI: 10.1126/sciadv.abg5369

**The PDF file includes:**



**Other Supplementary Material for this manuscript includes the following:**

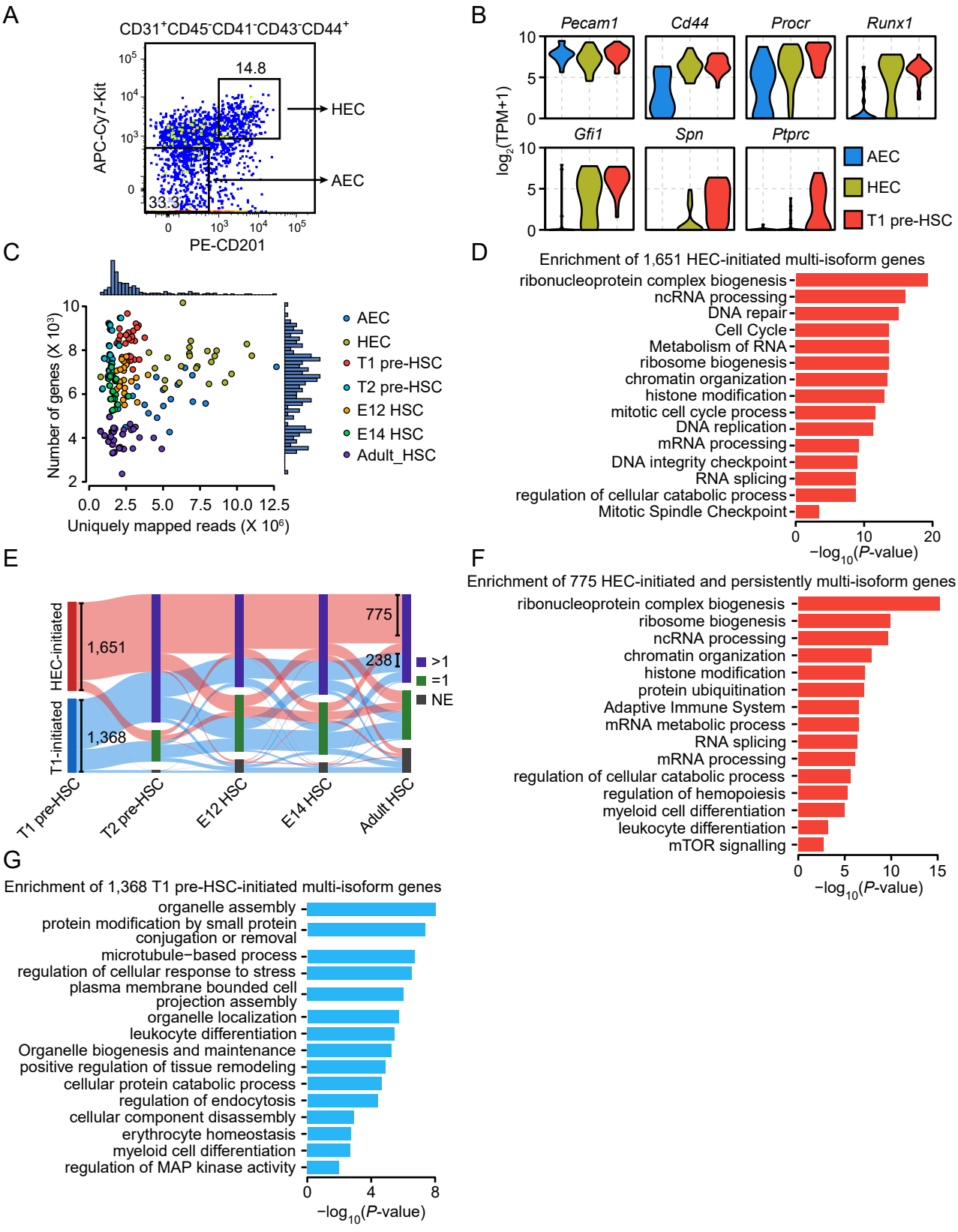

A
CD31+CD45-CD41-CD43-CD44+
APC-Cy7-Kit
PE-CD201
14.8
33.3
HEC
AEC
B
Pecam1
Cd44
Procr
Runx1
Gfi1
Spn
Ptprc
log2(TPM+1)
AEC
HEC
T1 pre-HSC
C
Number of genes (X 10³)
Uniquely mapped reads (X 10⁶)
AEC
HEC
T1 pre-HSC
T2 pre-HSC
E12 HSC
E14 HSC
Adult_HSC
D
Enrichment of 1,651 HEC-initiated multi-isoform genes
ribonucleoprotein complex biogenesis
ncRNA processing
DNA repair
Cell Cycle
Metabolism of RNA
ribosome biogenesis
chromatin organization
histone modification
mitotic cell cycle process
DNA replication
mRNA processing
DNA integrity checkpoint
RNA splicing
regulation of cellular catabolic process
Mitotic Spindle Checkpoint
−log10(P-value)
E
HEC-initiated
T1-initiated
1,651
1,368
775
238
>1
=1
NE
T1 pre-HSC
T2 pre-HSC
E12 HSC
E14 HSC
Adult HSC
F
Enrichment of 775 HEC-initiated and persistently multi-isoform genes
ribonucleoprotein complex biogenesis
ribosome biogenesis
ncRNA processing
chromatin organization
histone modification
protein ubiquitination
Adaptive Immune System
mRNA metabolic process
RNA splicing
mRNA processing
regulation of cellular catabolic process
regulation of hemopoiesis
myeloid cell differentiation
leukocyte differentiation
mTOR signalling
−log10(P-value)
G
Enrichment of 1,368 T1 pre-HSC-initiated multi-isoform genes
organelle assembly
protein modification by small protein conjugation or removal
microtubule−based process
regulation of cellular response to stress
plasma membrane bounded cell projection assembly
organelle localization
leukocyte differentiation
Organelle biogenesis and maintenance
positive regulation of tissue remodeling
cellular protein catabolic process
regulation of endocytosis
cellular component disassembly
erythrocyte homeostasis
myeloid cell differentiation
regulation of MAP kinase activity
−log10(P-value)

**Fig. S1. Transcript Diversity during Mouse HSC Development.**

**(A)** Representative flow cytometry analysis for cell sorting of AECs (CD31+CD45-CD41-CD43-CD201-Kit-CD44+) and HECs (CD31+CD45-CD41-CD43-CD201+Kit+CD44+) from E10 AGM region. **(B)** Violin plot showing the expression levels of indicated genes in AEC, HEC and T1 pre-HSC stages. **(C)** Density diagram showing the uniquely mapped reads and the expressed gene number in each single-cell. **(D)** Bar plot showing the enriched terms of HEC-initiated multi-isoform genes. **(E)** Sankey diagram showing the dynamic changes of HEC-initiated mult-isoform genes (HEC-initiated) and T1 pre-HSC-initiated multi-isoform genes (T1-initiated) from T1 pre-HSCs to Adult HSCs. The numbers on the graph indicate the number of expressed genes. >1, multi-isoform gene; =1, single-isoform gene; NE, No expression. **(F)** Bar plot showing the enriched terms of HEC-initiated and persistently multi-isoform genes. **(G)** Bar plot showing the enriched terms of T1 pre-HSC-initiated multi-isoform genes.

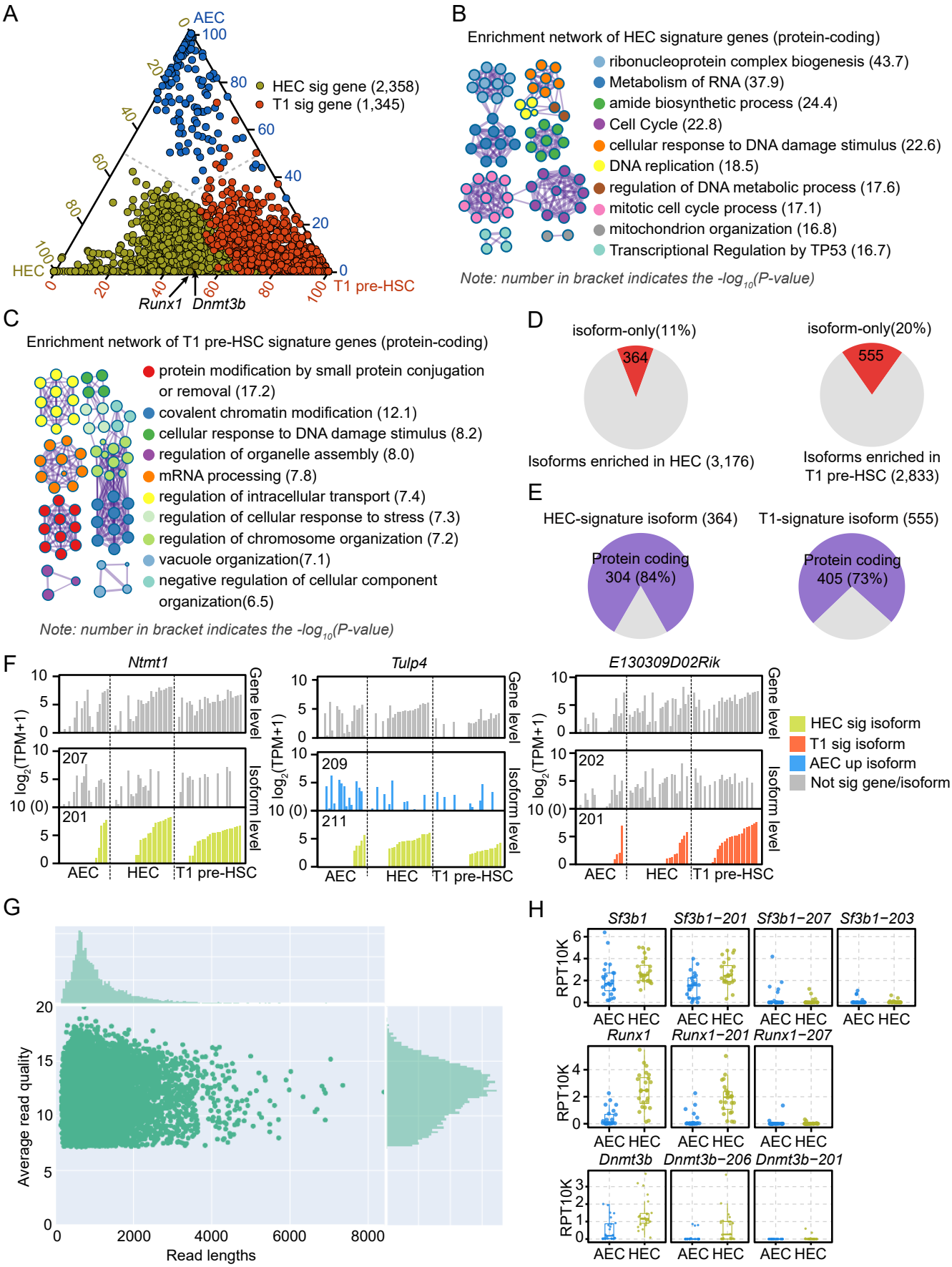

A
AEC
HEC
T1 pre-HSC
HEC sig gene (2,358)
T1 sig gene (1,345)
Runx1
Dnmt3b
B
Enrichment network of HEC signature genes (protein-coding)
ribonucleoprotein complex biogenesis (43.7)
Metabolism of RNA (37.9)
amide biosynthetic process (24.4)
Cell Cycle (22.8)
cellular response to DNA damage stimulus (22.6)
DNA replication (18.5)
regulation of DNA metabolic process (17.6)
mitotic cell cycle process (17.1)
mitochondrion organization (16.8)
Transcriptional Regulation by TP53 (16.7)
Note: number in bracket indicates the -log10(P-value)
C
Enrichment network of T1 pre-HSC signature genes (protein-coding)
protein modification by small protein conjugation or removal (17.2)
covalent chromatin modification (12.1)
cellular response to DNA damage stimulus (8.2)
regulation of organelle assembly (8.0)
mRNA processing (7.8)
regulation of intracellular transport (7.4)
regulation of cellular response to stress (7.3)
regulation of chromosome organization (7.2)
vacuole organization(7.1)
negative regulation of cellular component organization(6.5)
Note: number in bracket indicates the -log10(P-value)
D
isoform-only(11%)
364
Isoforms enriched in HEC (3,176)
isoform-only(20%)
555
Isoforms enriched in T1 pre-HSC (2,833)
E
HEC-signature isoform (364)
Protein coding 304 (84%)
T1-signature isoform (555)
Protein coding 405 (73%)
F
Ntmt1
Tulp4
E130309D02Rik
Gene level
Isoform level
log2(TPM+1)
207
201
209
211
202
AEC
HEC
T1 pre-HSC
HEC sig isoform
T1 sig isoform
AEC up isoform
Not sig gene/isoform
G
Average read quality
Read lengths
H
Sf3b1
Sf3b1-201
Sf3b1-207
Sf3b1-203
Runx1
Runx1-201
Runx1-207
Dnmt3b
Dnmt3b-206
Dnmt3b-201
RPT10K
AEC HEC

**Fig. S2. HEC and T1 pre-HSC Signature Genes and Isoforms.**

(**A**) Ternary phase diagram showing the relative enrichment of HEC (HEC sig) and T1 pre-HSC (T1 sig) signature genes. Upregulated genes in AEC compared to HEC were plot in blue. Indicated genes in Fig. 2F were labeled. (**B-C**) Enrichment network representing the top 10 enriched terms of HEC **(B)** and T1 pre-HSC **(C)** signature genes with protein coding abilities. Enriched terms with highly similarity were clustered and rendered as a network, while each node represents an enriched term and is colored by its cluster. Node size indicates the number of enriched genes and the line thickness indicates the similarity score shared by two enriched terms. The term with smallest *P*-value from each cluster was labeled. **(D)** Proportion of HEC (left) and T1 pre-HSC (right) signature isoforms in all enriched isoforms in HEC or T1 pre-HSC. (**E**) Proportion of HEC (left) and T1 pre-HSC (right) signature isoforms with protein coding abilities. (**F**) Bar plot showing the expression levels of indicated HEC and T1 pre-HSC signature isoforms in Fig. 2A. Number on the upper left represents the isoform ID from Ensembl. **(G)** Density diagram showing the distribution of average read quality and read lengths from Nanopore-seq data. (**H**) Boxplot showing the expression level of *Sf3b1*, *Runx1* and *Dnmt3b* genes, as well as isoforms in AEC and HEC quantified using Nanopore-seq data.

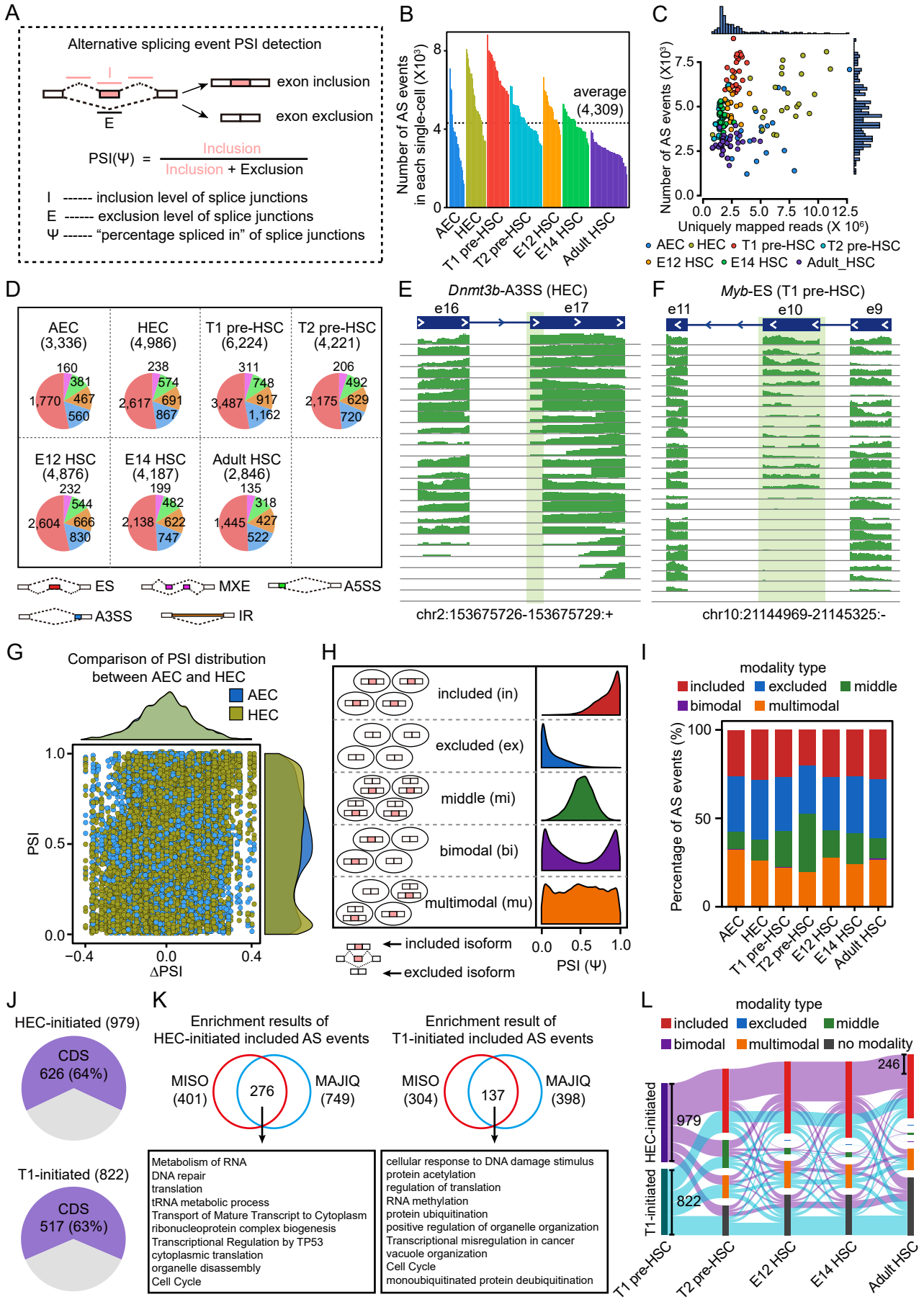

A
Alternative splicing event PSI detection
I
E
exon inclusion
exon exclusion
PSI(Ψ) = Inclusion / Inclusion + Exclusion
I ------ inclusion level of splice junctions
E ------ exclusion level of splice junctions
Ψ ------ "percentage spliced in" of splice junctions
B
Number of AS events in each single-cell (X10³)
average (4,309)
AEC
HEC
T1 pre-HSC
T2 pre-HSC
E12 HSC
E14 HSC
Adult HSC
C
Number of AS events (X10³)
Uniquely mapped reads (X 10⁶)
AEC HEC T1 pre-HSC T2 pre-HSC
E12 HSC E14 HSC Adult_HSC
D
AEC (3,336)
HEC (4,986)
T1 pre-HSC (6,224)
T2 pre-HSC (4,221)
E12 HSC (4,876)
E14 HSC (4,187)
Adult HSC (2,846)
ES
MXE
A5SS
A3SS
IR
E
Dnmt3b-A3SS (HEC)
e16
e17
chr2:153675726-153675729:+
F
Myb-ES (T1 pre-HSC)
e11
e10
e9
chr10:21144969-21145325:-
G
Comparison of PSI distribution between AEC and HEC
AEC
HEC
PSI
ΔPSI
H
included (in)
excluded (ex)
middle (mi)
bimodal (bi)
multimodal (mu)
included isoform
excluded isoform
PSI (Ψ)
I
modality type
included excluded middle
bimodal multimodal
Percentage of AS events (%)
J
HEC-initiated (979)
CDS 626 (64%)
T1-initiated (822)
CDS 517 (63%)
K
Enrichment results of HEC-initiated included AS events
MISO (401)
276
MAJIQ (749)
Metabolism of RNA
DNA repair
translation
tRNA metabolic process
Transport of Mature Transcript to Cytoplasm
ribonucleoprotein complex biogenesis
Transcriptional Regulation by TP53
cytoplasmic translation
organelle disassembly
Cell Cycle
Enrichment result of T1-initiated included AS events
MISO (304)
137
MAJIQ (398)
cellular response to DNA damage stimulus
protein acetylation
regulation of translation
RNA methylation
protein ubiquitination
positive regulation of organelle organization
Transcriptional misregulation in cancer
vacuole organization
Cell Cycle
monoubiquitinated protein deubiquitination
L
modality type
included excluded middle
bimodal multimodal no modality
HEC-initiated
T1-initiated
979
822
246
T1 pre-HSC
T2 pre-HSC
E12 HSC
E14 HSC
Adult HSC

**Fig. S3. Splicing Profiles during Endothelial-to-Hematopoietic Transition.**

(**A)** Schematic illustration of the simplified calculation of PSI of AS events. (**B)** Number of AS events detected with PSI in an individual cell of each cell population. (**C)** Density diagram showing the uniquely mapped reads and the AS event number in each single-cell. (**D)** Proportion of five types of AS events at each development stage. Schematic illustration of AS events types was plotted at bottom. The number on the graph represent the average number of AS events for each type of AS event, while the number in the brackets indicate the total number of AS events in each stage. ES, exon skipping; MXE, mutually exclusive exons; A5SS, alternative 5' splice sites; A3SS, alternative 3' splice sites; IR, intron retention. **(E)** Coverage track of alternative spliced exon 17 from *Dnmt3b* in HEC. Each row represents a single cell. The AS event type followed with gene name at top. The genomic location of alternative spliced exon labeled at the bottom. (**F)** Coverage track of alternative spliced exon 10 from *Myb* in T1 pre-HSC. Each row represents a single cell. The AS event type followed with gene name at top. The genomic location of alternative spliced exon labeled at the bottom. **(G)** Density diagram showing the PSI distribution between AEC and HEC using two different methods. X-axis represents the average difference (δPSI) of an AS events between AEC and HEC, while y-axis represents the PSI in each single cell from AEC and HEC stages. (**H)** Schematic illustration of five AS modality types. Each oval on the left represents a cell in a specific cell population. Distributions plot on the right represent the typical PSI distribution in five modalities. (**I)** Bar plot showing the proportion of AS event with different modalities in each cell population. (**J)** Proportion of HEC-initiated (top) and T1 pre-HSC-initiated (bottom) included AS events located in CDS regions. (**K)** Veen diagram showing the overlap between MISO and MAJIQ algorithms concerning the enriched terms of HEC- (left) or T1-initiated (right) included AS events. (**L)** Sankey diagram showing modality changes of HEC- (HEC-initiated) and T1 pre-HSC- (T1-initiated) initiated included AS events from T1 pre-HSCs to Adult HSCs. The numbers on the graph indicate the number of AS events.

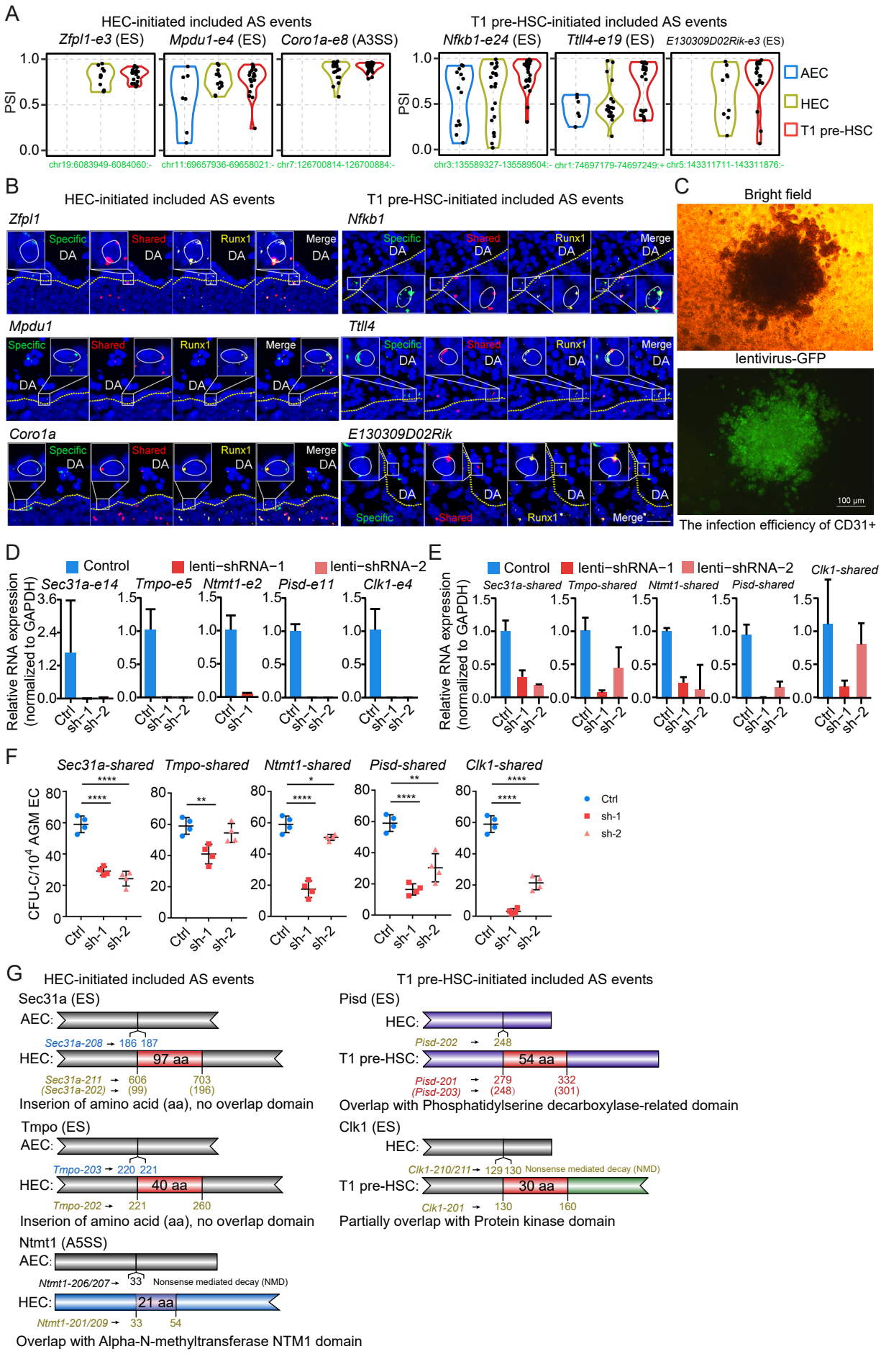

A
HEC-initiated included AS events
Zfpl1-e3 (ES)
Mpdu1-e4 (ES)
Coro1a-e8 (A3SS)
T1 pre-HSC-initiated included AS events
Nfkb1-e24 (ES)
Ttll4-e19 (ES)
E130309D02Rik-e3 (ES)
PSI
AEC
HEC
T1 pre-HSC
chr19:6083949-6084060:-
chr11:69657936-69658021:-
chr7:126700814-126700884:-
chr3:135589327-135589504:-
chr1:74697179-74697249:+
chr5:143311711-143311876:-
B
HEC-initiated included AS events
T1 pre-HSC-initiated included AS events
Zfpl1
Nfkb1
Mpdu1
Ttll4
Coro1a
E130309D02Rik
Specific
Shared
Runx1
Merge
DA
C
Bright field
lentivirus-GFP
100 μm
The infection efficiency of CD31+
D
Control
lenti−shRNA−1
lenti−shRNA−2
Relative RNA expression (normalized to GAPDH)
Sec31a-e14
Tmpo-e5
Ntmt1-e2
Pisd-e11
Clk1-e4
Ctrl
sh-1
sh-2
E
Sec31a-shared
Tmpo-shared
Ntmt1-shared
Pisd-shared
Clk1-shared
F
CFU-C/10^4 AGM EC
G
HEC-initiated included AS events
T1 pre-HSC-initiated included AS events
Sec31a (ES)
AEC:
Sec31a-208 → 186 187
HEC:
97 aa
Sec31a-211 → 606 703
(Sec31a-202) → (99) (196)
Inserion of amino acid (aa), no overlap domain
Pisd (ES)
HEC:
Pisd-202 → 248
T1 pre-HSC:
54 aa
Pisd-201 → 279 332
(Pisd-203) → (248) (301)
Overlap with Phosphatidylserine decarboxylase-related domain
Tmpo (ES)
AEC:
Tmpo-203 → 220 221
HEC:
40 aa
Tmpo-202 → 221 260
Inserion of amino acid (aa), no overlap domain
Clk1 (ES)
HEC:
Clk1-210/211 → 129 130 Nonsense mediated decay (NMD)
T1 pre-HSC:
30 aa
Clk1-201 → 130 160
Partially overlap with Protein kinase domain
Ntmt1 (A5SS)
AEC:
Ntmt1-206/207 → 33 Nonsense mediated decay (NMD)
HEC:
21 aa
Ntmt1-201/209 → 33 54
Overlap with Alpha-N-methyltransferase NTM1 domain

**Fig. S4. Expression and *In Vitro* Function Validation of HEC- and T1 pre-HSC-Initiated Included AS Events.**

(**A)** Violin plot representing the PSI changes of indicated HEC- (left) or T1 pre-HSC- (right) initiated included AS events. The genomic location of alternative spliced exons was labeled in green at bottom. Number of alternative spliced exon was labeled after gene name, followed with AS event type in brackets. ES, exon skipping; A3SS, alternative 3' splice sites. (**B)** Simultaneous detection of RNA and protein by using probes against the specific and shared exons, as well as antibody against Runx1 at E11 AGM region. DA, dorsal aorta. Scale bar, 25μm. **(C)** The infection efficiency of $CD31^+$ cells by purified lentivirus. (**D)** Q-PCR detection of each alternative spliced exon in $CD31^+$ cells infected with indicated lentivirus. (Data are represented as mean ± SEM.) **(E)** Q-PCR detection of shared exons of each gene in $CD31^+$ cells infected with indicated lentivirus. (Data are represented as mean ± SEM.) (**F)** Relative CFU-C number of lentivirus-mediated knockdown of shared exons of indicated genes in AGM $CD31^+$ cells. (Data are represented as mean ± SEM. Two-sided Wilcoxon rank-sum test, *$P$<0.05, **$P$<0.01, ***$P$<0.001, ****$P$<0.0001). **(G)** Schematic illustration showing the protein structure alterations of indicated AS events. Numbers under each bar representing the location of amino acid sequence. ES, exon skipping. A5SS, alternative 5' splicing site.

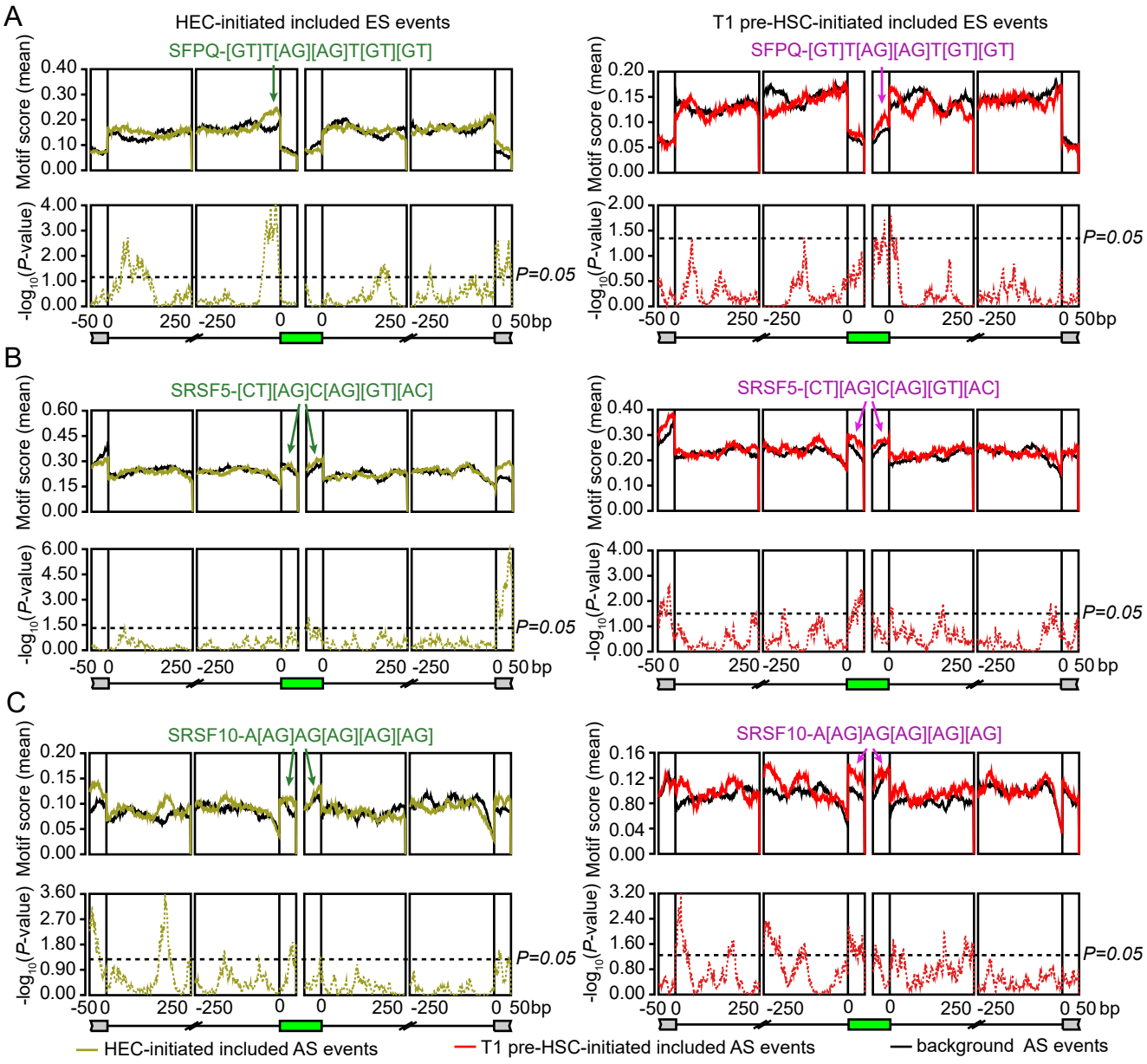
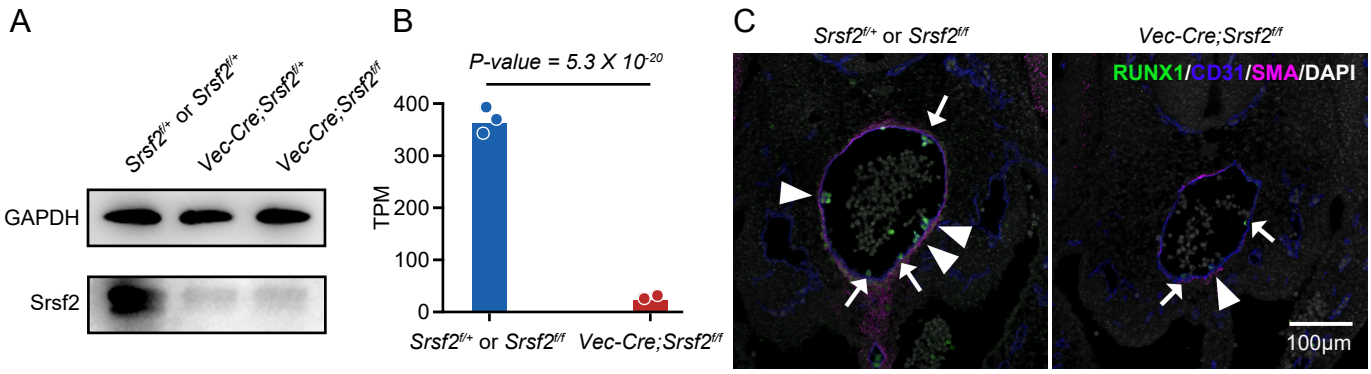

A
HEC-initiated included ES events
T1 pre-HSC-initiated included ES events
SFPQ-[GT]T[AG][AG]T[GT][GT]
Motif score (mean)
-log10(P-value)
P=0.05
B
SRSF5-[CT][AG]C[AG][GT][AC]
C
SRSF10-A[AG]AG[AG][AG][AG]
HEC-initiated included AS events
T1 pre-HSC-initiated included AS events
background AS events

**Fig. S5. RBP regulators for EHT.**

**(A-C)** Positional distribution of indicated RBP-binding motifs of HEC- (left) and T1 pre-HSC- (right) initiated included ES events. Motif enrichment scores (upper, solid line) and *P*-values (lower, dashed line) were plotted according to AS event positions. Arrows indicate peaks of enrichment for exons. ES: exon skipping.




D

*The number of small and large intra-aortic clusters in E10.0 Srsf2$^{f/+}$ or Srsf2$^{f/f}$ (control) and Vec-Cre;Srsf2$^{f/f}$ (KO) embryos.*

| | No.of small clusters | No.of large clusters | No.of all clusters | No.of IHC slices |
|---|---|---|---|---|
| control-1 | 59 | 12 | 71 | 14 |
| control-2 | 44 | 15 | 59 | 13 |
| control average (clusters/slices) | 3.8 | 1.0 | 4.8 | |
| KO-1 | 39 | 14 | 53 | 16 |
| KO-2 | 39 | 15 | 54 | 21 |
| KO average (clusters/slices) | 2.1 | 0.8 | 3.9 | |

**Fig. S6. Immunohistochemistry of *Srsf2* KO embryos.**

**(A)** Western blot showing the expression of Srsf2 protein in $CD44^{+}$ cells from E10.0 *Srsf2 f/+ or Srsf2 f/f* and *Vec-Cre;Srsf2 f/f* cells. **(B)** Bar plot showing the expression of *Srsf2* in *Srsf2 f/+ or Srsf2 f/f* and *Vec-Cre;Srsf2 f/f* cells from RNA sequencing data. **(C)** Immunohistochemistry results of the E10.0 *Srsf2 f/+ or Srsf2 f/f* and *Vec-Cre;Srsf2 f/f* embryos. Arrowheads indicate the large intra-aortic clusters. Arrows indicate the small clusters. **(D)** The number of small and large intra-aortic clusters in E10.0 $Srsf2^{f/+}$ or $Srsf2^{f/f}$ (control) and $Vec\text{-}Cre;Srsf2^{f/f}$ (KO) embryos.

## Supplementary Tables:

**Table S1. Types of HEC- and T1 pre-HSC-initiated multi-isoform genes in 7 HSC development stages.**

**Table S2. HEC and T1 pre-HSC signature genes and isoforms. (A)** HEC and T1 pre-HSC signature genes. **(B)**, HEC and T1 pre-HSC signature isoforms.

**Table S3. HEC- and T1 pre-HSC-initiated included AS events and their enrichment results. (A)** HEC- and T1 pre-HSC-initiated included AS events. **(B-C)** Enrichment results of HEC- **(B)** and T1 pre-HSC-initiated **(C)** included AS events located into the CDS region.

**Table S4. Sequences for the validation of HEC- and T1 pre-HSC-initiated included AS events. (A)** Sequences used for FISH probe design. **(B)** Primers used for shRNA lentivirus construction. **(C)** Primers used for qPCR detection of selected AS events.

**Table S5. Prediction of RBPs responsible for the inclusion of hemogenic-specific included AS events. (A)** Enriched RNA binding motifs of HEC- and T1-initiated included ES events. **(B)** Relative expression level of enriched RBPs in AEC, HEC and T1 pre-HSC. **(C)** RNA binding proteins used for motif enrichment analyses.

**Table S6. Loss of *Srsf2* disrupts gene programs required for HSC formation. (A)** Differential expressed genes between *Srsf2*$^{f/+}$ or *Srsf2*$^{f/f}$ (control) and *Vec-Cre;Srsf2*$^{f/f}$ (KO) cells. **(B)** Overlap genes between HEC/T1 pre-HSC signature genes and *Srsf2* KO differently expressed genes. **(C)** Differential AS events between control and *Srsf2* KO samples.