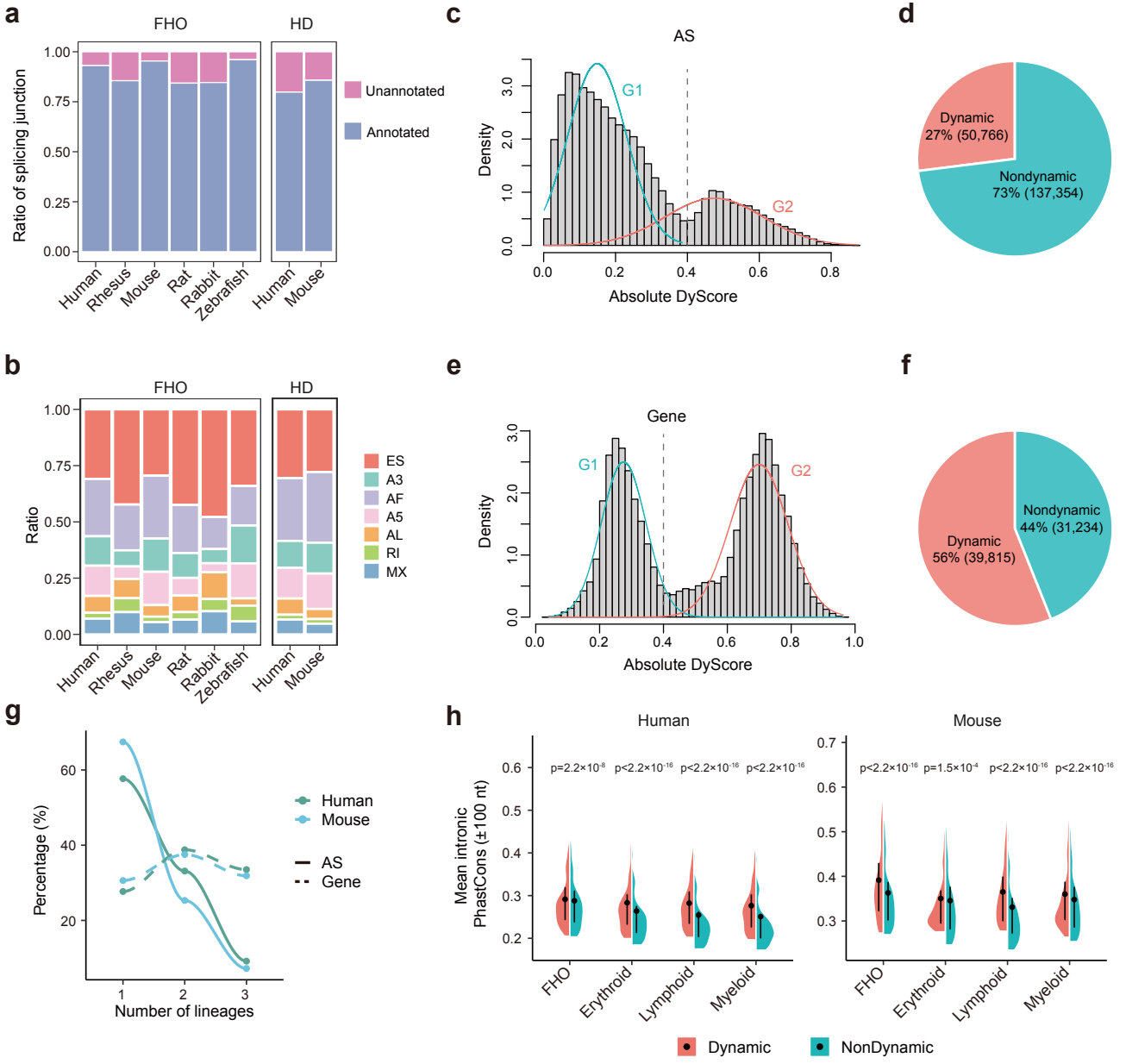

a
FHO
HD
Ratio of splicing junction
1.00
0.75
0.50
0.25
0.00
Unannotated
Annotated
Human
Rhesus
Mouse
Rat
Rabbit
Zebrafish
Human
Mouse
b
FHO
HD
Ratio
ES
A3
AF
A5
AL
RI
MX
c
AS
G1
G2
Density
Absolute DyScore
d
Dynamic
27% (50,766)
Nondynamic
73% (137,354)
e
Gene
G1
G2
Density
Absolute DyScore
f
Nondynamic
44% (31,234)
Dynamic
56% (39,815)
g
Percentage (%)
Number of lineages
Human
Mouse
AS
Gene
h
Human
Mouse
Mean intronic
PhastCons (±100 nt)
p=2.2×10-8
p<2.2×10-16
p<2.2×10-16
p<2.2×10-16
p<2.2×10-16
p=1.5×10-4
p<2.2×10-16
p<2.2×10-16
FHO
Erythroid
Lymphoid
Myeloid
Dynamic
NonDynamic

Supplementary Fig. 1 | Comparative analysis of dynamic features and classification of alternative splicing events and genes.

a, Proportions of annotated versus novel splicing junctions across all datasets. FHO: Fetal hematopoietic organ; HD: Hematopoietic differentiation dataset.

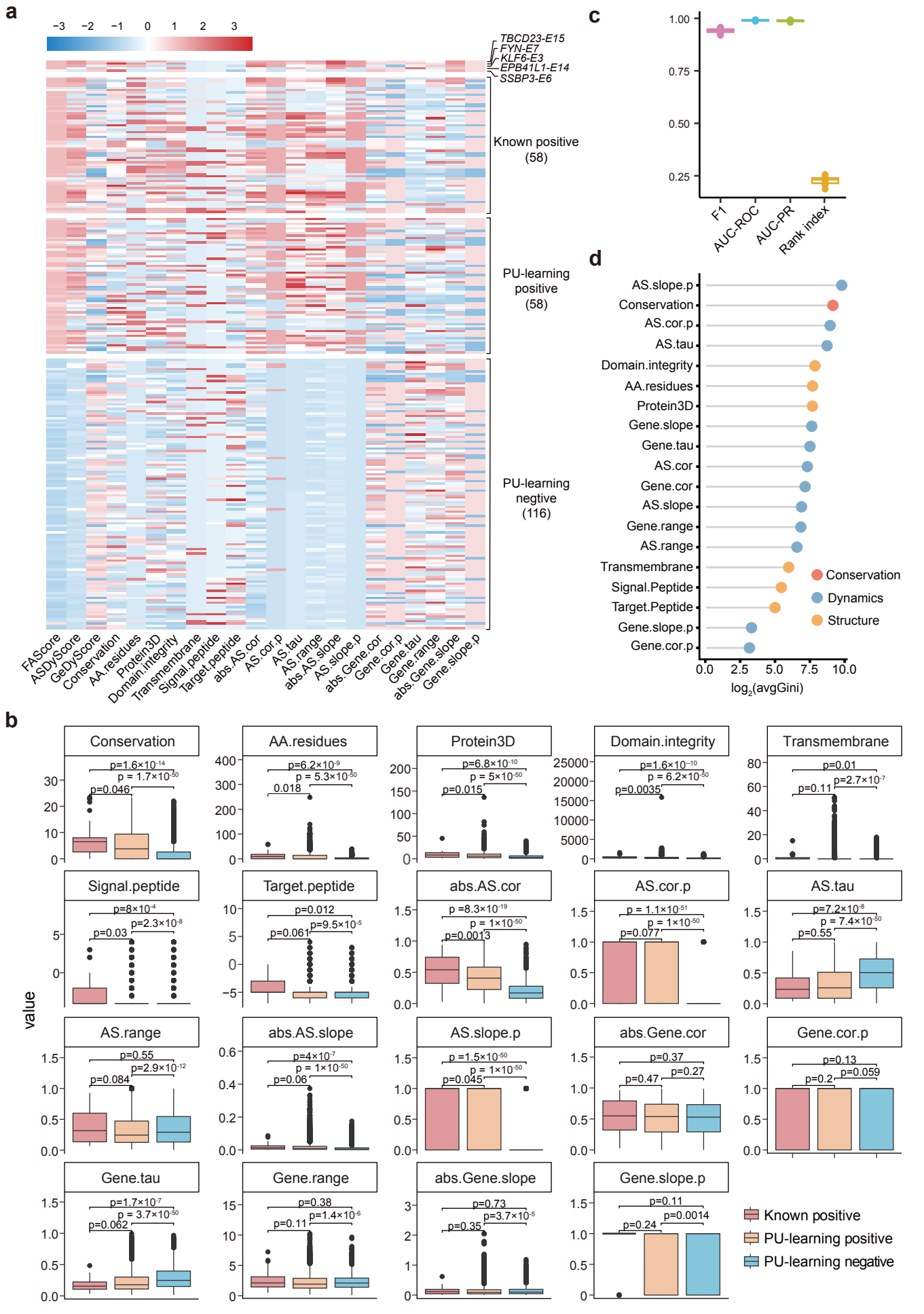
b, Distribution of alternative splicing (AS) event types inferred by SUPPA2 in all datasets, including ES (exon skipping), AF (alternative first exon), A3 (alternative 3' splice sites), A5 (alternative 5' splice sites), AL (alternative last exon), RI (retained intron), and MX (mutually exclusive exon).

c, e, Distribution of absolute dynamic score (DyScore) shown as overlaid histogram and KDE for AS events (**c**) and genes (**e**). G1 and G2 denote the two fitted Gaussian components used for classifying dynamic and non-dynamic elements.

d, f, Pie charts summarizing the proportion and number of dynamically versus non-dynamically regulated AS events (**d**) and genes (**f**). 27% (50,766) of AS events and 44% (31,234) of genes were classified as dynamic.

g, Lineage specificity comparison of dynamically regulated AS events and genes for both human and mice. The x-axis represents the number of lineages in which dynamically regulated AS events and genes are present, while the y-axis indicates the percentage of dynamically regulated AS or genes.

h, Distribution of mean primate sequence conservation (PhastCons) scores of intronic sequences flanking exons (100 bp of upstream and downstream) for humans and mice across FHO, erythroid, lymphoid and myeloid. Statistical significance was determined by paired two-sided Student's t-test.


a
−3 −2 −1 0 1 2 3
TBCD23-E15
FYN-E7
KLF6-E3
EPB41L1-E14
SSBP3-E6
Known positive (58)
PU-learning positive (58)
PU-learning negtive (116)
FAScore
ASDyScore
GeDyScore
Conservation
AA.residues
Protein3D
Domain.integrity
Transmembrane
Signal.peptide
Target.peptide
abs.AS.cor
AS.cor.p
AS.tau
AS.range
abs.AS.slope
AS.slope.p
abs.Gene.cor
Gene.cor.p
Gene.tau
Gene.range
abs.Gene.slope
Gene.slope.p
c
1.00
0.75
0.50
0.25
F1
AUC-ROC
AUC-PR
Rank index
d
AS.slope.p
Conservation
AS.cor.p
AS.tau
Domain.integrity
AA.residues
Protein3D
Gene.slope
Gene.tau
AS.cor
Gene.cor
AS.slope
Gene.range
AS.range
Transmembrane
Signal.Peptide
Target.Peptide
Gene.slope.p
Gene.cor.p
Conservation
Dynamics
Structure
0.0 2.5 5.0 7.5 10.0
log2(avgGini)
b
value
Conservation
p=1.6×10-14
p = 1.7×10-50
p=0.046
AA.residues
p=6.2×10-9
p = 5.3×10-30
0.018
Protein3D
p=6.8×10-10
p = 5×10-50
p=0.015
Domain.integrity
p=1.6×10-10
p = 6.2×10-30
p=0.0035
Transmembrane
p=0.01
p=2.7×10-7
p=0.11
Signal.peptide
p=8×10-4
p=2.3×10-8
p=0.03
Target.peptide
p=0.012
p=9.5×10-5
p=0.061
abs.AS.cor
p =8.3×10-19
p = 1×10-50
p=0.0013
AS.cor.p
p = 1.1×10-51
p = 1×10-50
p=0.077
AS.tau
p=7.2×10-8
p = 7.4×10-30
p=0.55
AS.range
p=0.55
p=2.9×10-12
p=0.084
abs.AS.slope
p=4×10-7
p = 1×10-50
p=0.06
AS.slope.p
p =1.5×10-50
p = 1×10-50
p=0.045
abs.Gene.cor
p=0.37
p=0.47
p=0.27
Gene.cor.p
p=0.13
p=0.2
p=0.059
Gene.tau
p=1.7×10-7
p = 3.7×10-50
p=0.062
Gene.range
p=0.38
p=1.4×10-6
p=0.11
abs.Gene.slope
p=0.73
p=3.7×10-5
p=0.35
Gene.slope.p
p=0.11
p=0.24
p=0.0014
Known positive
PU-learning positive
PU-learning negative

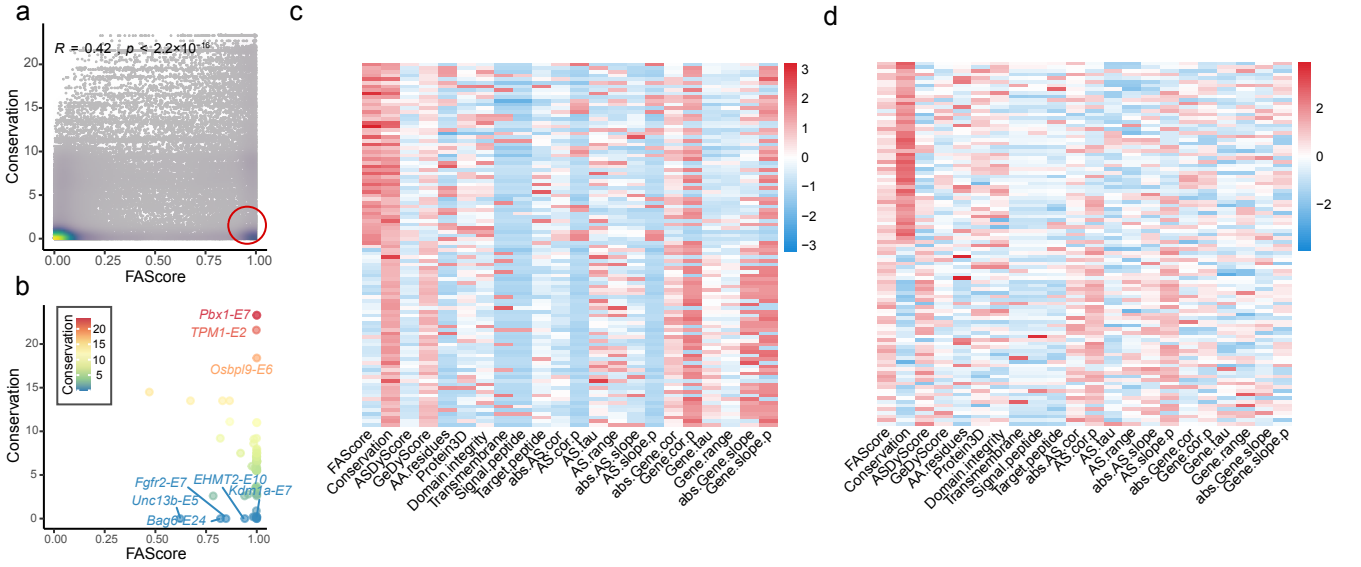
Supplementary Fig. 2 | The performance evaluation of the predicted model.

a, Heat map displays normalized feature scores for candidate AS events, known positive (58) by the literature, reliable positive (58 were randomly selected) and reliable negative (116 were randomly selected) AS events identified via PU-learning. Each row represents an AS event, and each column corresponds to a predictive feature, with the final FAScore. Values are z-score normalized.

b, Comparison of feature distributions across known positive AS sets, reliable positive AS and reliable negative AS sets identified by PU-learning. Statistical significance was determined by the two-sided Wilcoxon rank-sum test. abs.AS.cor/abs.Gene.cor and AS.cor.p/Gene.cor.p: absolute Spearman's rank coefficient and Spearman's test p-value of AS or Genes; AS.Tau/Gene.Tau: the specificity index of AS or Genes; AS.Range/ Gene.Range: the range of PSI or gene expression; abs.AS.Slope/abs.Gene.Slope and AS.Fit.p/Gene.Fit.p: absolute linear regression coefficient and p-value of the regression variable; AA.residues: functionally important residues score; Protein3D: the structural homologs and integrity score; Domain.integrity: domain integrity score; Transmenmbrane: trans-membrane helices score; Signal.peptide: signal peptide score; Target.peptide: target peptide score.

c, Performance evaluation of each model using 10-fold cross-validation. Metrics include F1 score, ROC-AUC (area under the receiver operating characteristic curve), PR-AUC (area under the precision-recall curve) and rank index, which assesses the ability of FAScore to identify functional AS.

d, Importance ranking of features used in the final model by average Gini index.


a
R = -0.42 ; p < 2.2×10⁻¹⁶
Conservation
FAScore
b
Conservation
Pbx1-E7
TPM1-E2
Osbpl9-E6
Fgfr2-E7
EHMT2-E10
Kdm1a-E7
Unc13b-E5
Bag6-E24
FAScore
c
FAScore
Conservation
ASDyScore
GeDyScore
AA.residues
Protein3D
Domain.integrity
Transmembrane
Signal.peptide
Target.peptide
abs.AS.cor
AS.cor.p
AS.tau
AS.range
abs.AS.slope
AS.slope.p
abs.Gene.cor
Gene.cor.p
Gene.tau
Gene.range
abs.Gene.slope
Gene.slope.p
d

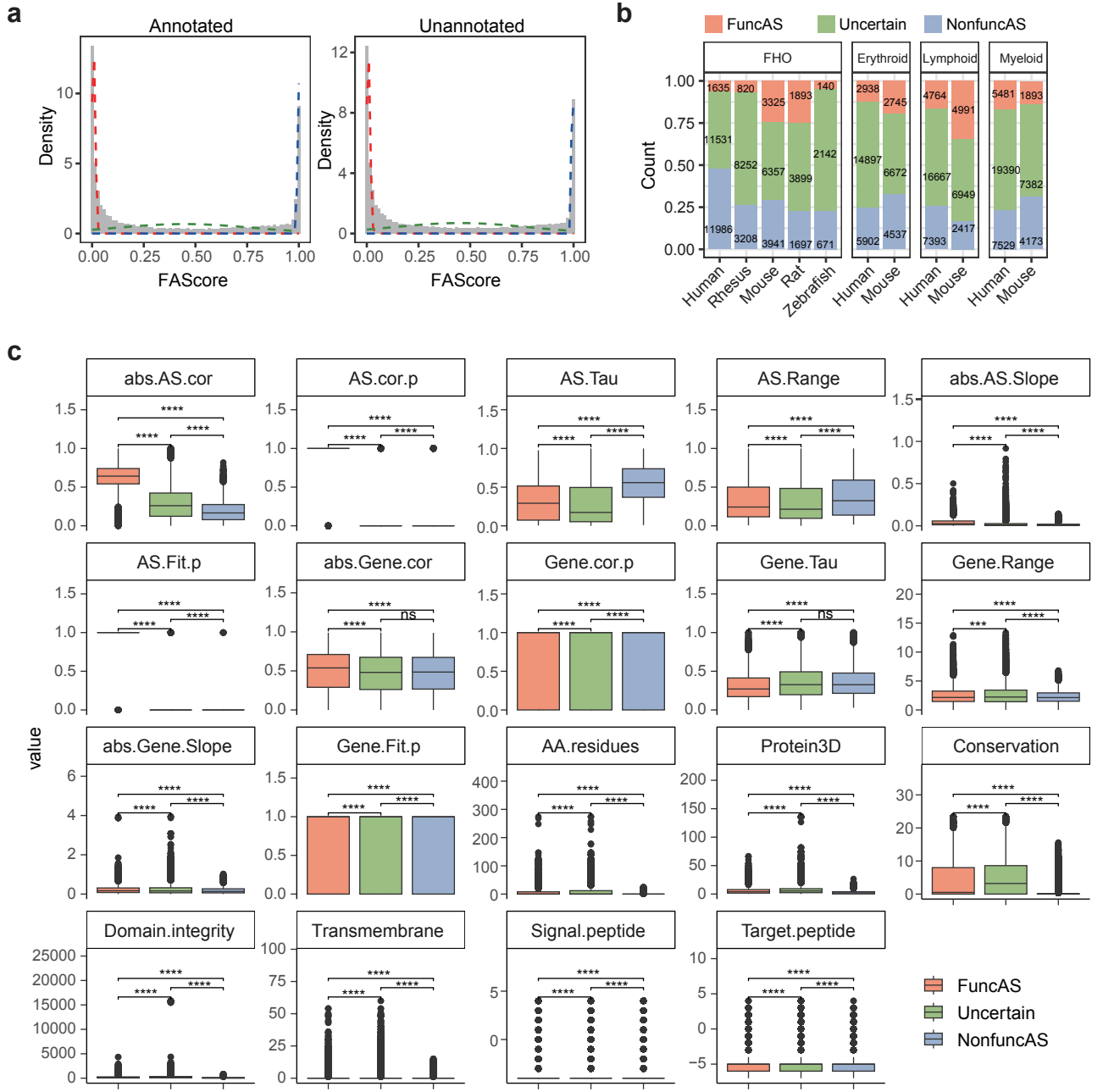
Supplementary Fig. 3 | Correlation between FAScore and evolutionary conservation across alternative splicing events.

a, Density plot showing the correlation between FAScore and evolutionary conservation across species in ES events, with darker regions indicating higher density of events. The red circle highlights a region containing splicing events with high FAScore but low conservation. A nonparametric Spearman's rank-order correlation test was conducted, and the correlation coefficient (*R*) and its associated *p*-value are shown.

b, Scatter plot showing the correlation between FAScore and evolutionary conservation score across species. Each point represents a known functional splicing event.

c, Heatmap displays normalized feature scores for high conservation with high FAScore (50 were randomly selected) or low FAScore (50 were randomly selected) AS events identified via PU-learning. Each row represents an AS event, and each column corresponds to a predictive feature, with the final FAScore.

d, Heatmap displays normalized feature scores for high FAScore with high conservation (50 were randomly selected) or low conservation (50 were randomly selected) AS events identified via PU-learning. Each row represents an AS event, and each column corresponds to a predictive feature, with the final FAScore.

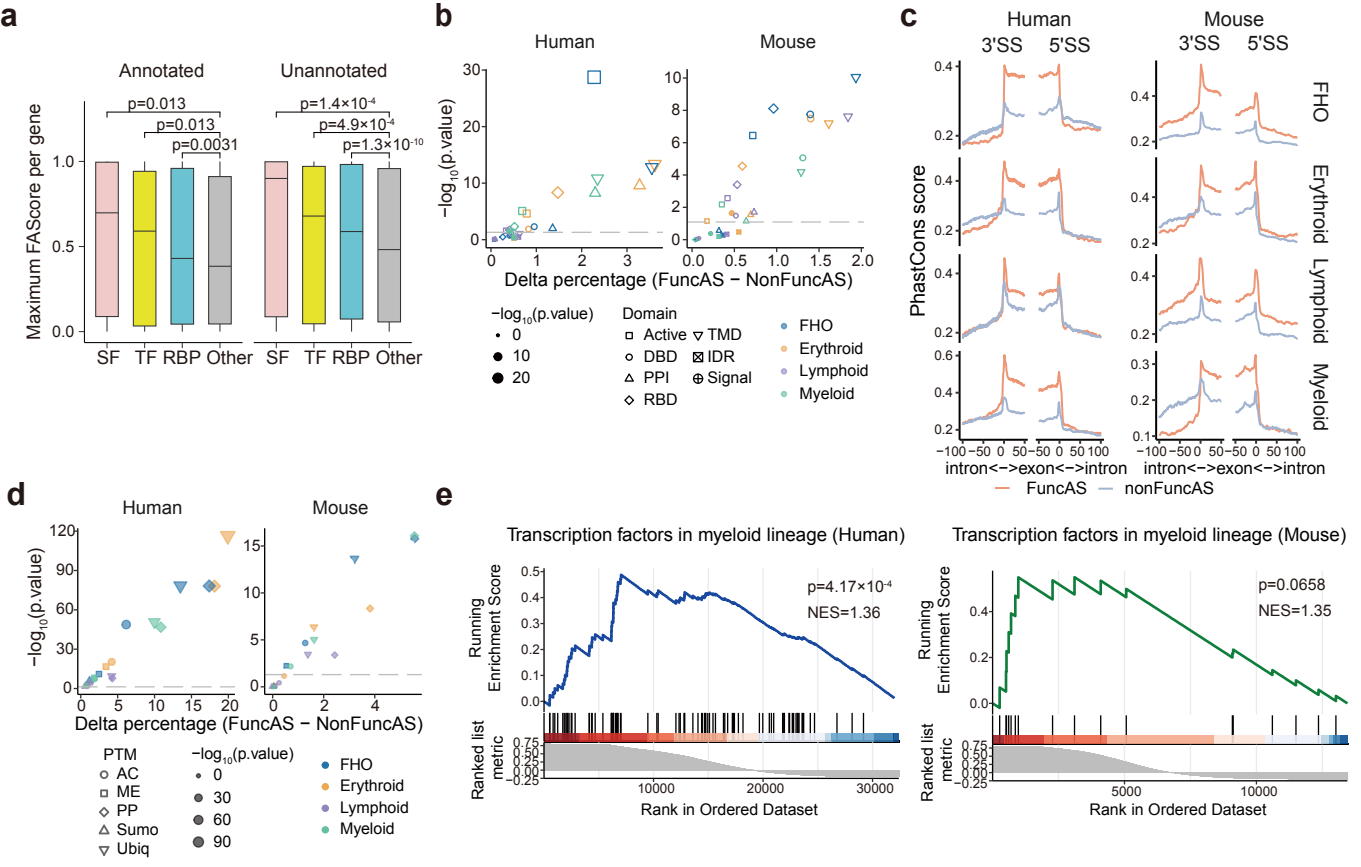

a
Annotated
Unannotated
Density
FAScore
b
FuncAS
Uncertain
NonfuncAS
FHO
Erythroid
Lymphoid
Myeloid
Count
Human
Rhesus
Mouse
Rat
Zebrafish
c
abs.AS.cor
AS.cor.p
AS.Tau
AS.Range
abs.AS.Slope
AS.Fit.p
abs.Gene.cor
Gene.cor.p
Gene.Tau
Gene.Range
abs.Gene.Slope
Gene.Fit.p
AA.residues
Protein3D
Conservation
Domain.integrity
Transmembrane
Signal.peptide
Target.peptide
value
ns

Supplementary Fig. 4 | Functional prediction and characterization of alternative splicing events using FAScore.

a, Density distributions of FAScore for annotated exon skipping (ES) events (Left) and novel splicing events (Right). Curves are stratified into functional (red), uncertain (green), and non-functional (blue) categories based on a Gaussian mixture model.

b, Stacked bar plot showing the functional classification of AS events across all datasets. The y-axis indicates the proportion of events. Colors correspond to functional categories as in (**a**). The number in plot indicates the number of AS events per category.

c, Comparison of feature distributions across functional (red), uncertain (green), and non-functional (blue) AS events using boxplots. Statistical significance was determined by the two-sided Wilcoxon rank-sum test.


a
Annotated
Unannotated
p=0.013
p=0.013
p=0.0031
p=1.4×10⁻⁴
p=4.9×10⁻⁴
p=1.3×10⁻¹⁰
Maximum FAScore per gene
SF TF RBP Other SF TF RBP Other
b
Human
Mouse
−log10(p.value)
Delta percentage (FuncAS − NonFuncAS)
−log10(p.value)
Domain
Active TMD
DBD IDR
PPI Signal
RBD
FHO
Erythroid
Lymphoid
Myeloid
c
Human
Mouse
3'SS 5'SS
PhastCons score
FHO
Erythroid
Lymphoid
Myeloid
intron<−>exon<−>intron
FuncAS nonFuncAS
d
Human
Mouse
−log10(p.value)
Delta percentage (FuncAS − NonFuncAS)
PTM
AC
ME
PP
Sumo
Ubiq
−log10(p.value)
FHO
Erythroid
Lymphoid
Myeloid
e
Transcription factors in myeloid lineage (Human)
p=4.17×10⁻⁴
NES=1.36
Running Enrichment Score
Ranked list metric
Rank in Ordered Dataset
Transcription factors in myeloid lineage (Mouse)
p=0.0658
NES=1.35
Running Enrichment Score
Ranked list metric
Rank in Ordered Dataset

Supplementary Fig. 5 | The validation of predicted functional AS.

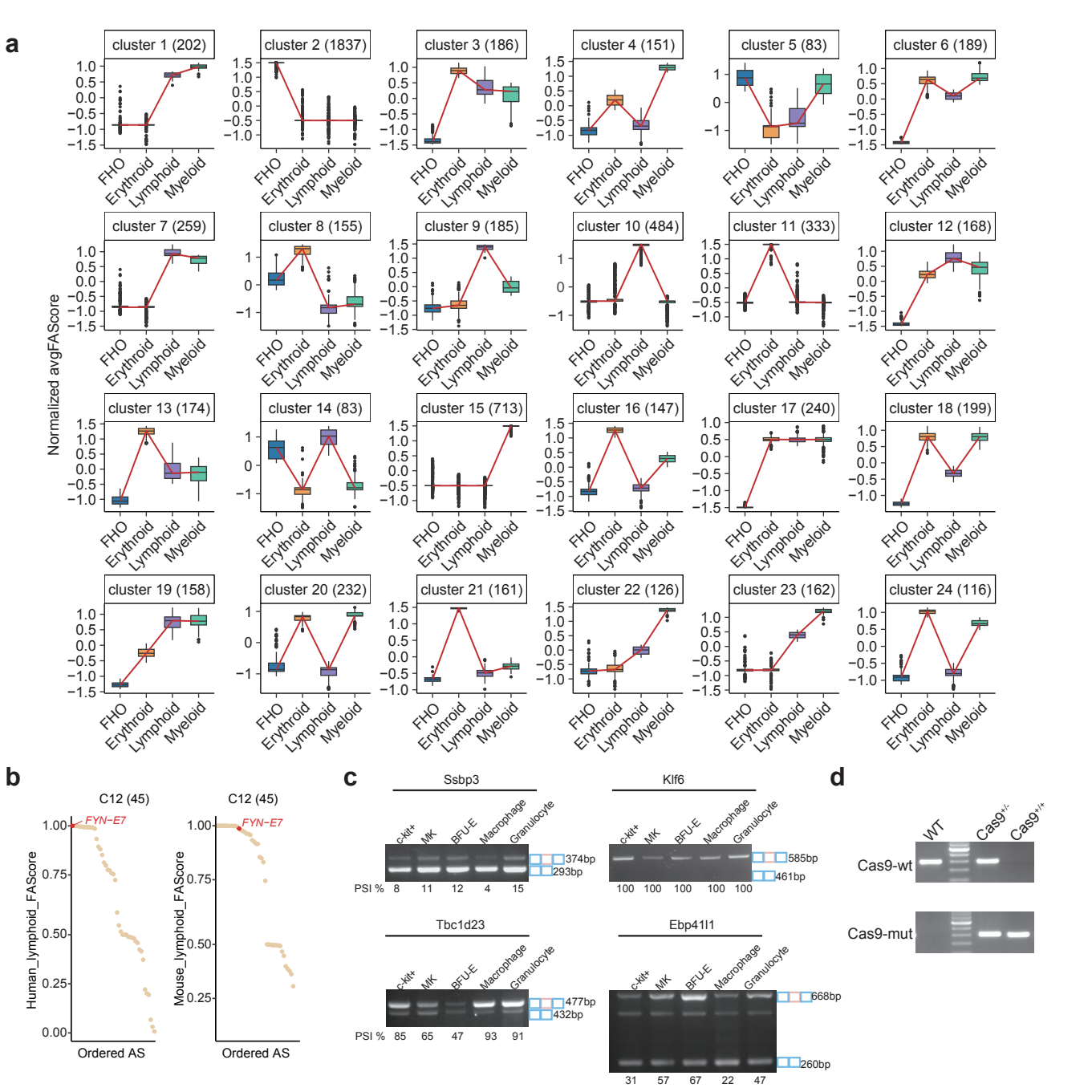
a, Boxplot illustrating the distribution of maximum FAScore values for AS events across each gene, color-coded by gene type: splicing factor (SF), RNA binding protein (RBP), transcription factor (TF), and other. The p-value was calculated using an unpaired two-sided Student's t-test.

b, Volcano plots illustrate significant protein domain differences between functional and non-functional unannotated AS events in humans (Left) and mice (Right) The x-axis shows percentage difference, while the y-axis displays $-\log_{10}$(*p*-value) from the two-sided proportion tests. Colors denote distinct datasets, and point shapes indicate different protein domain types, including Active (Active site), DBD (DNA binding domain), PPI (Protein-protein interaction domain), RBD (RNA binding domain), TMDs (Transmembrane domains), IDR (Intrinsically disordered regions), and Signal (Signal peptide). The grey dashed line indicates statistical significance (*p*-value = 0.05).

c, Comparison of conservation scores (PhastCons) surrounding splice sites (3'SS and 5'SS) between functional and non-functional AS of unannotated type in different lineages. The x-axis shows position relative to splice sites (0), covering intronic (-100 bp) and exonic (+50 bp) regions. The y-axis shows the average PhastCons scores for each site of functional and non-functional AS.

d, Volcano plot illustrating important post-translational modification (PTM) sites contained in the exonic region of unannotated AS in humans (Left) and mice (Right). The x-axis shows the percentage difference of PTM sites between functional and non-functional AS, while the y-axis presents $-\log_{10}$(*p*-value) from the two-sided proportion test. Colors represent distinct datasets, with point shapes indicating types of PTM sites, including AC (Acetylation), ME (Methylation), PP (Phosphorylation), SUMO (SUMOylation), and Ubiq (Ubiquitination). The gray dashed line represents the significance threshold for enrichment (*p*-value = 0.05).

e, Gene Set Enrichment Analysis (GSEA) indicates significant enrichment of myeloid-specific transcription factor sets in genes exhibiting functional AS in humans (Left) and mice (Right).


a
cluster 1 (202)
cluster 2 (1837)
cluster 3 (186)
cluster 4 (151)
cluster 5 (83)
cluster 6 (189)
cluster 7 (259)
cluster 8 (155)
cluster 9 (185)
cluster 10 (484)
cluster 11 (333)
cluster 12 (168)
cluster 13 (174)
cluster 14 (83)
cluster 15 (713)
cluster 16 (147)
cluster 17 (240)
cluster 18 (199)
cluster 19 (158)
cluster 20 (232)
cluster 21 (161)
cluster 22 (126)
cluster 23 (162)
cluster 24 (116)
FHO
Erythroid
Lymphoid
Myeloid
Normalized avgFAScore
b
C12 (45)
FYN−E7
Human_lymphoid_FAScore
Mouse_lymphoid_FAScore
Ordered AS
c
Ssbp3
Klf6
Tbc1d23
Ebp41l1
c-kit+
MK
BFU-E
Macrophage
Granulocyte
374bp
293bp
585bp
461bp
477bp
432bp
668bp
260bp
PSI %
8
11
12
4
15
100
100
100
100
100
85
65
47
93
91
31
57
67
22
47
d
WT
Cas9+/-
Cas9+/+
Cas9-wt
Cas9-mut

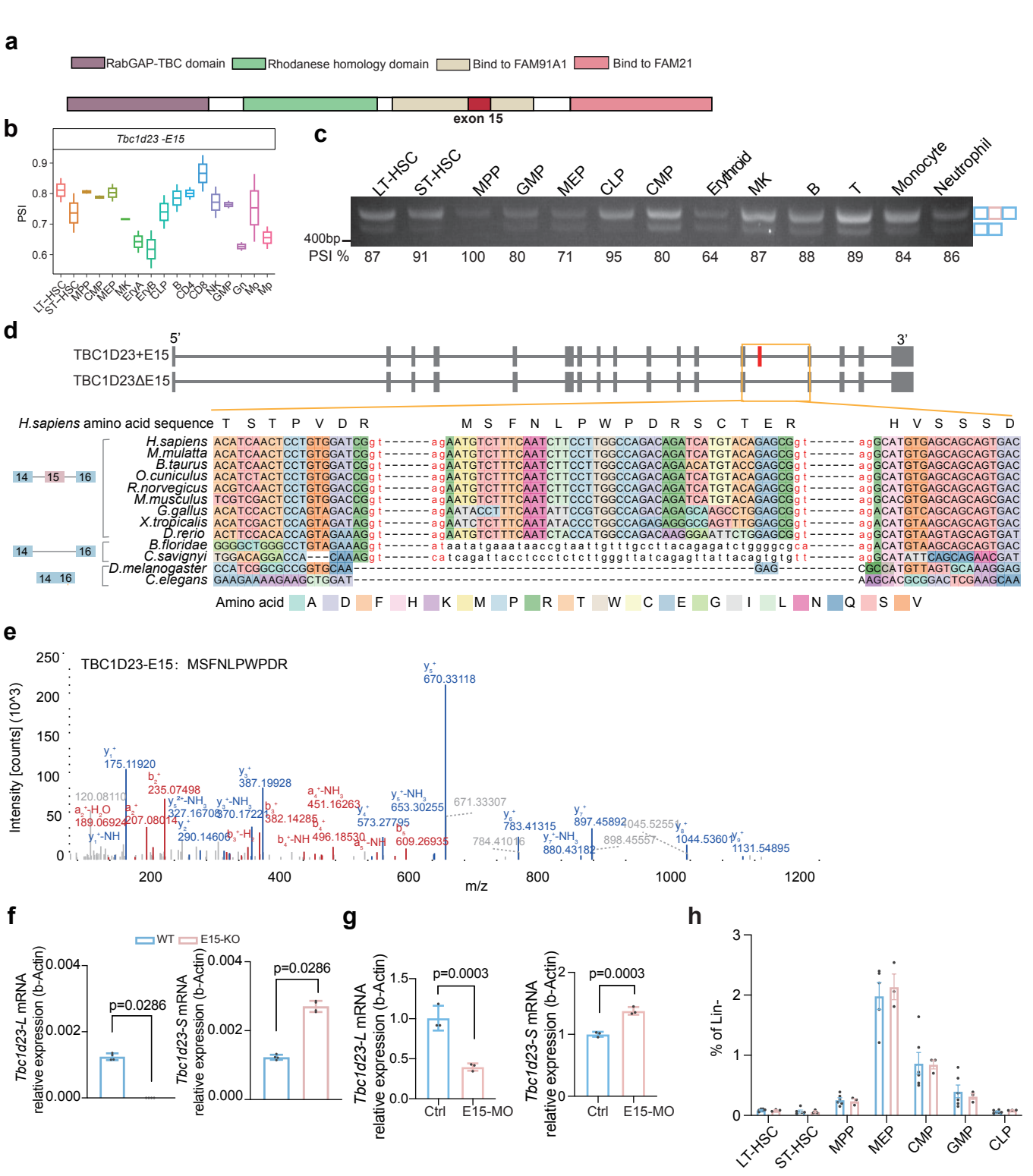
Supplementary Fig. 6 | Identification of patterns of evolutionarily conserved AS

a, The distribution of the normalized average FAScore of ancient AS events (age class 5) for four hematopoietic lineages in 24 clusters from fuzzy c-means clustering. Plot numbers indicate the number of AS events assigned to each of these clusters. Boxes indicate the interquartile range (IQR, 25^{th} and 75^{th} percentiles), with the center line representing the median. Whiskers extend to the most extreme values within 1.5 x IQR, and points beyond are shown as outliers. A red line connects the median values of the lineages, delineating the overall function trend across the four lineages for that cluster.

b, Rank-ordered distribution of functional AS scores of humans and mice for ES events in cluster 12. AS events are ranked by descending FAScore (0.00 to 1.00). Each point represents an AS event. The number of ES events in each cluster is indicated in parentheses. The candidate AS event is highlighted in red.

c, Quantification of PSI values for *Klf6* exon 3, *Ssbp3* exon 6, *Tbc1d23* exon 15, and *Epb41l1* exon 14 was performed by RT-PCR analysis in the following cell populations isolated from colony-forming unit (CFU) colonies: c-kit$^+$ cells, megakaryocytes (MK), erythroid burst-forming units (BFU-E), macrophages, and granulocytes. Representative results from three independent experiments are shown.

d, Genotyping of Cas9$^+$ mice. Representative results from three independent experiments are shown.


a
RabGAP-TBC domain
Rhodanese homology domain
Bind to FAM91A1
Bind to FAM21
exon 15
b
Tbc1d23 -E15
PSI
LT-HSC
ST-HSC
MPP
CMP
MEP
MK
EryA
EryB
CLP
B
CD4
CD8
NK
GMP
Gn
Mo
Mp
c
LT-HSC
ST-HSC
MPP
GMP
MEP
CLP
CMP
Erythroid
MK
B
T
Monocyte
Neutrophil
400bp
PSI %
87
91
100
80
71
95
80
64
87
88
89
84
86
d
5'
3'
TBC1D23+E15
TBC1D23ΔE15
H.sapiens amino acid sequence T S T P V D R M S F N L P W P D R S C T E R H V S S S D
H.sapiens
M.mulatta
B.taurus
O.cuniculus
R.norvegicus
M.musculus
G.gallus
X.tropicalis
D.rerio
B.floridae
C.savignyi
D.melanogaster
C.elegans
14
15
16
Amino acid
A
D
F
H
K
M
P
R
T
W
C
E
G
I
L
N
Q
S
V
e
TBC1D23-E15: MSFNLPWPDR
Intensity [counts] (10^3)
m/z
f
WT
E15-KO
Tbc1d23-L mRNA relative expression (b-Actin)
Tbc1d23-S mRNA relative expression (b-Actin)
p=0.0286
p=0.0286
g
Tbc1d23-L mRNA relative expression (b-Actin)
Tbc1d23-S mRNA relative expression (b-Actin)
p=0.0003
p=0.0003
Ctrl
E15-MO
Ctrl
E15-MO
h
% of Lin-
LT-HSC
ST-HSC
MPP
MEP
CMP
GMP
CLP

Supplementary Fig. 7 | Exon 15 of TBC1D23 regulates erythropoiesis.

a, Schematic of TBC1D23 protein domains, highlighting the position of exon 15 within the FAM91A1-binding region.

b, Box plot of PSI values for *Tbc1d23-E15* skipping events derived from RNA-seq data of hematopoietic differentiation in mouse. LT-HSC: Long-time hematopoietic stem cell, ST-HSC: Short-time hematopoietic stem cell, MPP: Multipotent progenitor, CLP: Common lymphoid progenitor, CMP: Common myeloid progenitor, GMP: Granulocyte monocyte progenitor, MEP: Megakaryocyte erythrocyte progenitor, Ery: Erythroblasts, MK: Megakaryocytes, Mo: Monocyte, Mp: Macrophage, B: B cell, CD4: CD4+ T cell, CD8: CD8+ T cell, NK: Natural killer cell.

c, RT-PCR validation of spliced transcripts in mouse hematopoietic cells. Calculate the PSI value of each lane was calculated by dividing the gray value of the longer transcript by the sum of the gray value of the longer and shorter transcripts. Representative results from three independent experiments are shown.

d, Gene structure of *TBC1D23* based on transcript data from the ENSEMBL database, with exon 15 indicated in red (Top). The alignment of the base sequence around exon 15 and its flanking exon regions of *TBC1D23* across 13 species, ranging from invertebrates to vertebrates. The 5' and 3' splice sites are marked in red, and codon bases are color-coded in the background. The dashed line indicates the intronic regions. Species are ranked by evolutionary distance (Bottom).

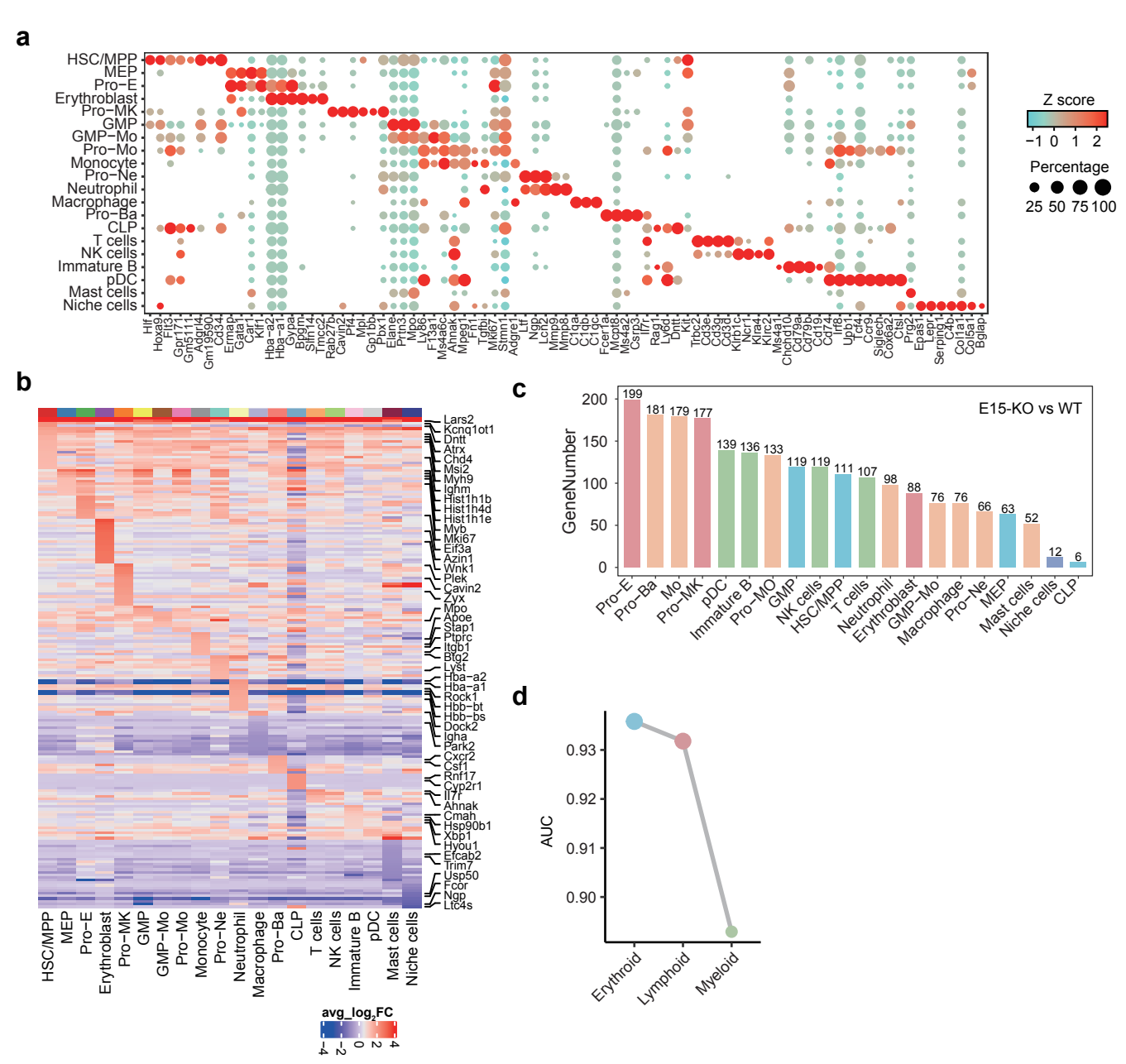
e, MS/MS analysis of peptides from TBC1D23-L immunoprecipitation..

f, RT-qPCR analysis of *Tbc1d23-L* (Left), and *Tbc1d23-S* (Right) mRNA levels in WT and E15-KO (n=3 mice/group).

g, RT-qPCR analysis of *Tbc1d23-L* (Left), and *Tbc1d23-S* (Right) mRNA levels in Ctrl-MO and E15-MO zebrafish embryos at 48 hpf (n=3 zebrafish/group).

h, Statistical analysis of the percentage of LT-HSC, ST-HSC, MPP, MEP, CMP, and GMP cells in Lin- cells from BM of WT (n = 6 mice) and E15-KO (n = 3 mice) by FACS.

For Statistical analysis, the following test was used: Unpaired two-tailed Student's t-test (**f, g**) and Two-tailed Mann-Whitney U tests (**h**). Data are shown as mean ± s.d and are representative of three independent experiments.


a
HSC/MPP
MEP
Pro−E
Erythroblast
Pro−MK
GMP
GMP−Mo
Pro−Mo
Monocyte
Pro−Ne
Neutrophil
Macrophage
Pro−Ba
CLP
T cells
NK cells
Immature B
pDC
Mast cells
Niche cells
Z score
−1 0 1 2
Percentage
25 50 75 100
b
avg_log2FC
−4 −2 0 2 4
c
E15-KO vs WT
GeneNumber
0 50 100 150 200
199 181 179 177 139 136 133 119 119 111 107 98 88 76 76 66 63 52 12 6
Pro−E Pro−Ba Mo Pro−MK pDC Immature B Pro−MO GMP NK cells HSC/MPP T cells Neutrophil Erythroblast GMP−Mo Macrophage Pro−Ne MEP Mast cells Niche cells CLP
d
AUC
0.90 0.91 0.92 0.93
Erythroid Lymphoid Myeloid

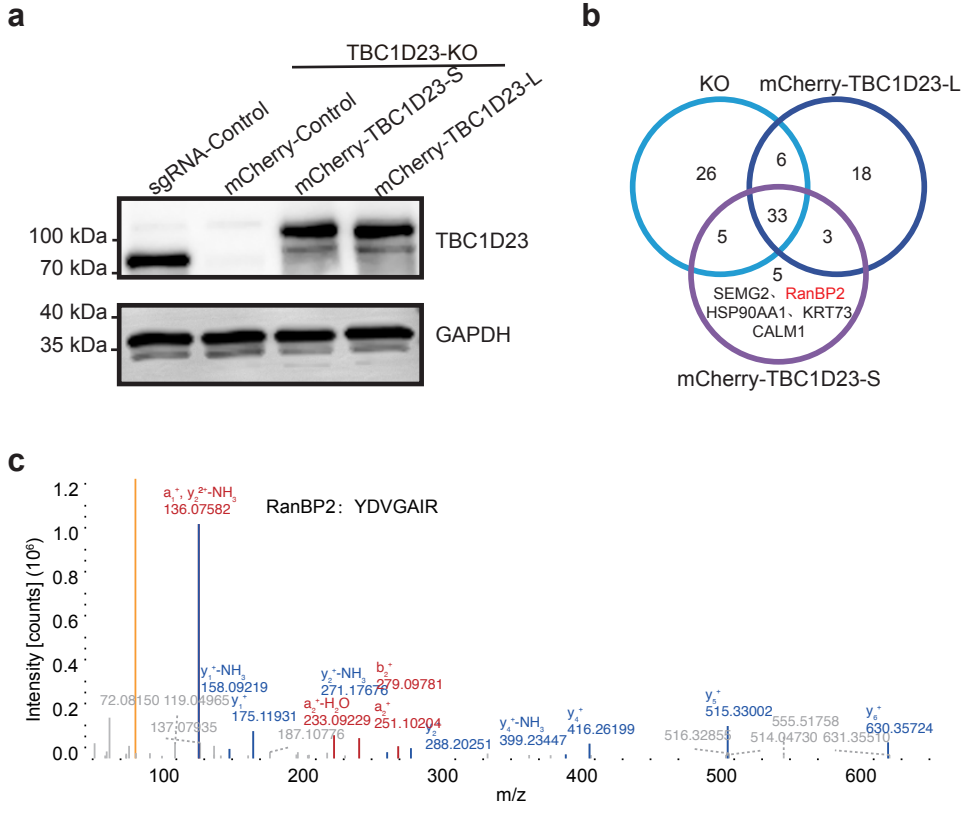
Supplementary Fig. 8| Single-cell analysis confirms that Exon 15 of TBC1D23 regulates erythroid differentiation.

a, Dot plot displaying expression level of marker genes for each cluster. The color scale indicates the expression levels and the size of dots is proportional to the fraction of cells. HSC/MPP: Hematopoietic stem cell/Multipotent progenitor, MEP: Megakaryocyte erythrocyte progenitor, Pro-E: Pro-erythroblast, Pro-MK: Pro-megakaryocytes, GMP: Granulocyte monocyte progenitor, GMP-Mo: GMP-monocyte, Pro-Mo: Pro-monocyte, Pro-Ne: Pro-neutrophil, Pro-Ba: Pro-basophil, CLP: Common lymphoid progenitor, pDC: plasmacytoid dendritic cell.

b, Heatmap illustrating the row-scaled expression of the top 20 differentially expressed genes (DEGs) per cell type, identified using the Bonferroni-corrected *p*-values (< 0.05) from the Wilcoxon rank sum test.

c, Count of DEGs in each cell type comparing E15-KO and WT samples.

d, Dot plot displaying the AUC value (identifying which lineage exhibits the most transcriptional changes between E15-KO and WT) of differential lineages.

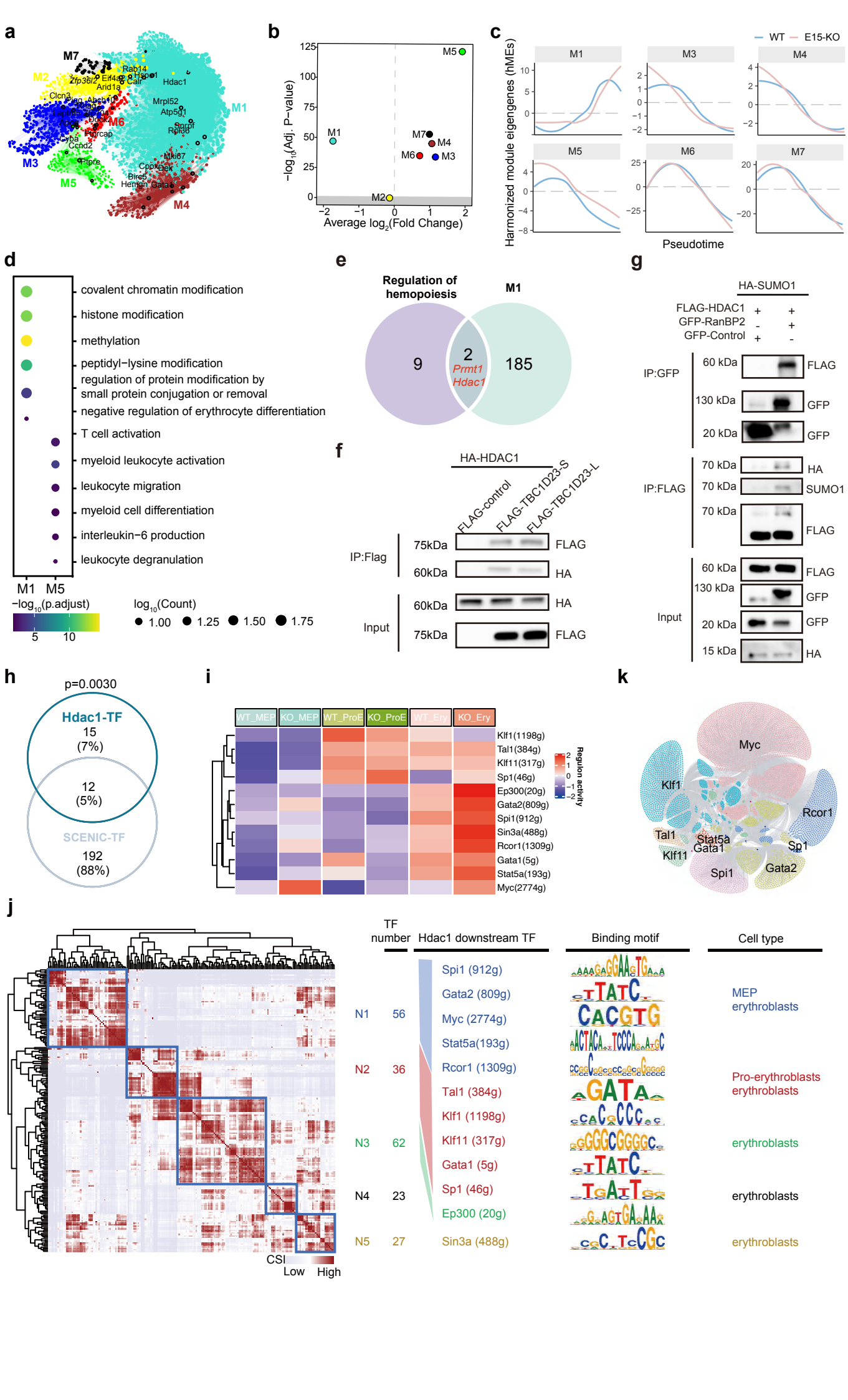

a
TBC1D23-KO
sgRNA-Control
mCherry-Control
mCherry-TBC1D23-S
mCherry-TBC1D23-L
100 kDa
70 kDa
TBC1D23
40 kDa
35 kDa
GAPDH
b
KO
mCherry-TBC1D23-L
26
6
18
33
5
3
5
SEMG2、RanBP2
HSP90AA1、KRT73
CALM1
mCherry-TBC1D23-S
c
RanBP2：YDVGAIR
Intensity [counts] (10⁶)
m/z
136.07582
158.09219
175.11931
271.17678
279.09781
233.09229
251.10204
288.20251
399.23447
416.26199
515.33002
630.35724

Supplementary Fig. 9 | TBC1D23-E15 encodes a sequence that is critical for interacting with the E3-sumo ligase RanBP2 complex.

a, Western blot analysis of TBC1D23 protein expression levels in HEK293T cells under four conditions: sgRNA-control cells, TBC1D23-KO cells reconstituted with either empty mCherry vector (mCherry), mCherry-tagged short isoform of TBC1D23 (mCherry-TBC1D23-S), or mCherry-tagged long isoform of TBC1D23 (mCherry-TBC1D23-L). Representative results from three independent experiments are shown.

b, Venn diagram of all proteins identified by MS / MS of TBC1D23 knockout (KO), TBC1D23-L, and TBC1D23-S. The five specific interaction proteins of TBC1D23-S were shown.

c, MS / MS analysis of the peptide found in the gel band from the sample of TBC1D23-S immunoprecipitation.


a
M7
M2
M1
M6
M3
M5
M4
Rab14
Eif4g3
Arid1a
Hspa1
Calr
Hdac1
Clcn3
Mrpl52
Atp5g1
Sprpf
Rpl10
Ptprcap
Cybb
Cend2
Mki67
Cpox
Dtx4
Birc6
Hemgn
Gata2
Ptpre
b
M5
M1
M7
M4
M6
M3
M2
−log10(Adj. P−value)
Average log2(Fold Change)
c
WT
E15-KO
M1
M3
M4
M5
M6
M7
Harmonized module eigengenes (hMEs)
Pseudotime
d
covalent chromatin modification
histone modification
methylation
peptidyl−lysine modification
regulation of protein modification by small protein conjugation or removal
negative regulation of erythrocyte differentiation
T cell activation
myeloid leukocyte activation
leukocyte migration
myeloid cell differentiation
interleukin−6 production
leukocyte degranulation
M1
M5
−log10(p.adjust)
log10(Count)
1.00
1.25
1.50
1.75
e
Regulation of hemopoiesis
M1
9
2
Prmt1
Hdac1
185
f
HA-HDAC1
FLAG-control
FLAG-TBC1D23-S
FLAG-TBC1D23-L
IP:Flag
75kDa
FLAG
60kDa
HA
Input
60kDa
HA
75kDa
FLAG
g
HA-SUMO1
FLAG-HDAC1
GFP-RanBP2
GFP-Control
IP:GFP
60 kDa
FLAG
130 kDa
GFP
20 kDa
GFP
IP:FLAG
70 kDa
HA
70 kDa
SUMO1
70 kDa
FLAG
Input
60 kDa
FLAG
130 kDa
GFP
20 kDa
GFP
15 kDa
HA
h
p=0.0030
Hdac1-TF
15
(7%)
12
(5%)
SCENIC-TF
192
(88%)
i
WT_MEP
KO_MEP
WT_ProE
KO_ProE
WT_Ery
KO_Ery
Klf1(1198g)
Tal1(384g)
Klf11(317g)
Sp1(46g)
Ep300(20g)
Gata2(809g)
Spi1(912g)
Sin3a(488g)
Rcor1(1309g)
Gata1(5g)
Stat5a(193g)
Myc(2774g)
Regulon activity
k
Myc
Klf1
Rcor1
Tal1
Stat5a
Gata1
Sp1
Klf11
Spi1
Gata2
j
CSI
Low
High
TF number
Hdac1 downstream TF
Binding motif
Cell type
N1
56
Spi1 (912g)
Gata2 (809g)
Myc (2774g)
Stat5a(193g)
MEP erythroblasts
N2
36
Rcor1 (1309g)
Tal1 (384g)
Klf1 (1198g)
Pro-erythroblasts erythroblasts
N3
62
Klf11 (317g)
Gata1 (5g)
Sp1 (46g)
erythroblasts
N4
23
Ep300 (20g)
erythroblasts
N5
27
Sin3a (488g)
erythroblasts

Supplementary Fig. 10 | The differential expression analysis for co-expression network modules between WT and TBC1D23-E15 KO.

a, UMAP plot of the erythroid co-expression network derived from scRNA-seq using Hierarchical Dynamic Weighted Gene Co-expression Network Analysis (hdWGCNA). Each node represents an individual gene, while edges indicate co-expression links between genes and module hub genes. Point size reflects eigengene-based connectivity (kME), and nodes are colored according to co-expression module assignments. The top five hub genes within each module are labeled.

b, Volcano plot depicting differential networks between knockout (KO) and wild-type (WT) samples. The x-axis shows effect size; the y-axis shows significance ($-\log_{10}$(*p*-value), two-sided Wilcoxon rank-sum test). The grey area marks non-significant modules.

c, Module eigengene (ME) dynamics along pseudotime for each module, with loess regression lines fitted separately for each condition.

d, GO pathways analysis of genes from Module 1 and Module 5. Bubble size indicates gene count per term; color indicates enrichment significance (one-sided hypergeometric test with Benjamini-Hochberg correction).

e, Venn diagram showing gene overlap between the protein modification pathway (enriched in Module 1) and the hematopoiesis regulation pathway (enriched in SUMOylation substrates from Fig. 4e).

f, HA-HDAC1 were co-transfected with Flag-TBC1D23-S or Flag-TBC1D23-L plasmid in HEK293T cells, and the cell lysates were subjected to IP/WB assays as indicated. Representative results from three independent experiments are shown.

g, Pull-down assay in HEK293T cells stably expressing HA-SUMO1 and co-transfected with Flag-HDAC1 and GFP-RanBP2/GFP-Control. Crosslinking IP was performed with anti-GFP (top) or anti-Flag (bottom) antibodies, followed by immunoblotting. Representative results from three independent experiments are shown.

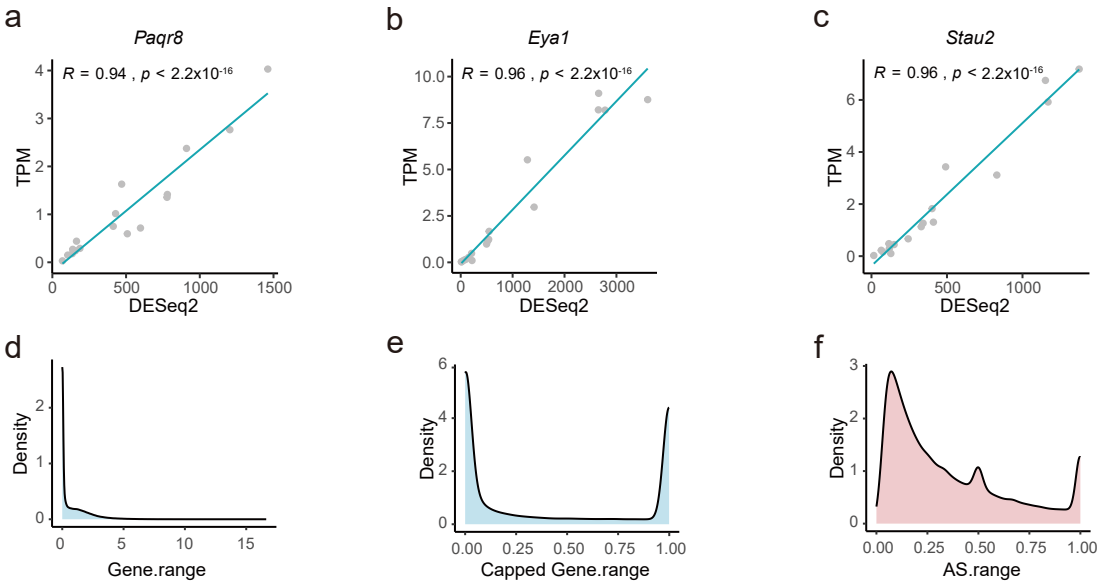
h, Venn diagram of overlapping transcription factors (TFs) between erythropoiesis-related motifs (SCENIC analysis) and previously reported Hdac1-related TFs in erythroid differentiation. Significance was assessed by one-sided hypergeometric test.

i, Heatmap of regulon activity scores (RAS) for 12 TFs (from panel **h**) across MEP, pro-

erythroblast, and erythroblast clusters in E15-KO vs. WT..

j, Regulon modules identified by the connection specificity index (CSI) matrix, showing representative TFs, Hdac1-related TFs, binding motifs, and associated cell types.

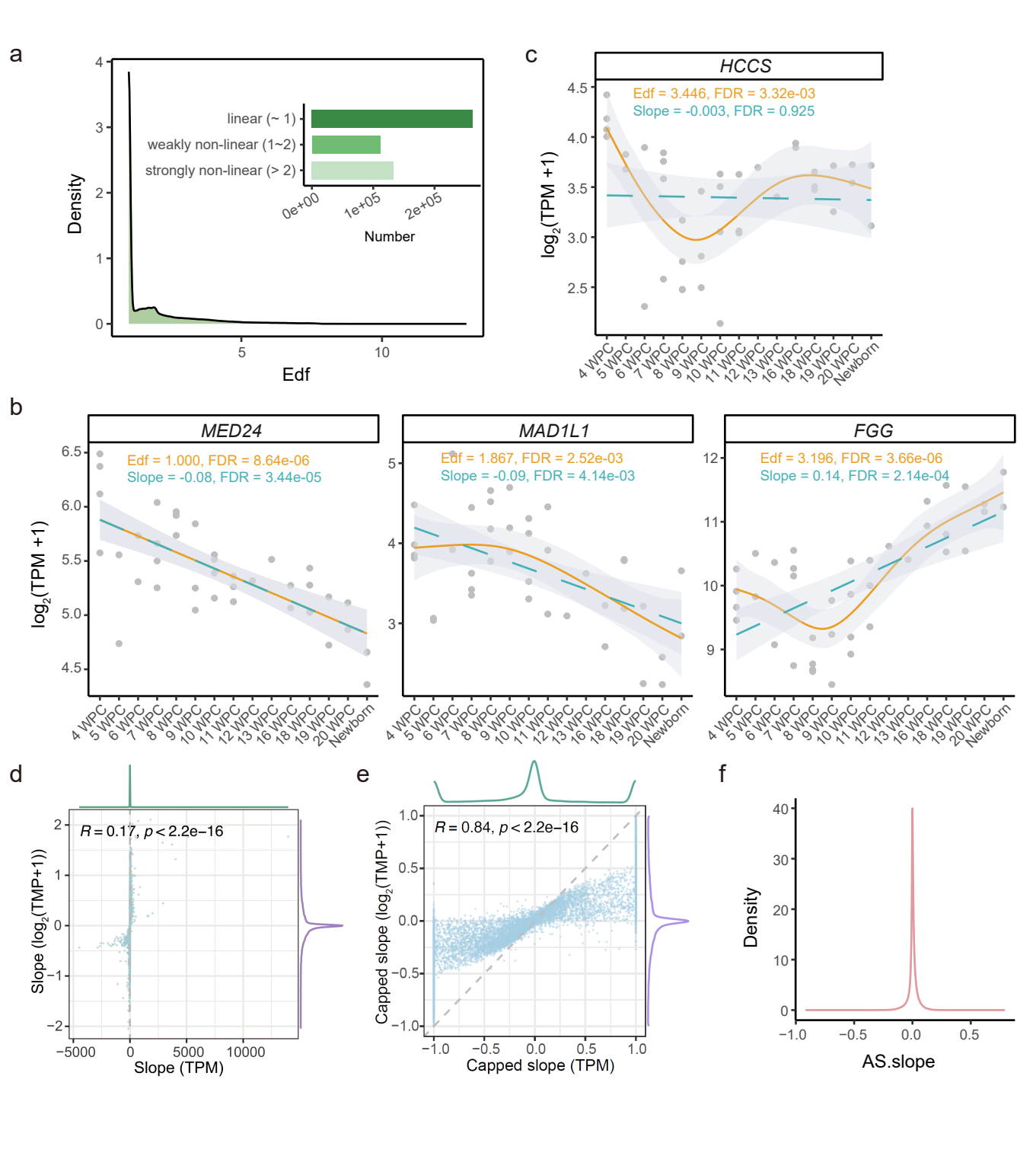
k, Network diagram of key TFs and their target genes. Nodes represent genes; edges denote regulatory influence. Colors distinguish target genes of different TFs.


a
Paqr8
R = 0.94 , p < 2.2x10-16
TPM
DESeq2
b
Eya1
R = 0.96 , p < 2.2x10-16
TPM
DESeq2
c
Stau2
R = 0.96 , p < 2.2x10-16
TPM
DESeq2
d
Density
Gene.range
e
Density
Capped Gene.range
f
Density
AS.range

Supplementary Fig. 11 | The distribution of expression breadth of genes and alternative splicing events.

a-c, Scatter plots showing the correlation between TPM and DESeq2-normalized expression values for three example genes: (**a**) *Paqr8* (R = 0.94, $p < 2.2\times10^{-16}$), (**b**) *Eya1* (R = 0.96, $p < 2.2\times10^{-16}$), and (**c**) *Stau2* (R = 0.96, $p < 2.2\times10^{-16}$). The blue line in each panel represents the linear regression fit.

d-f, Density distributions of expression breadth of genes and AS, including Gene.range (**d**), Capped Gene.range (**e**) and AS.range (**f**).


a
Density
Edf
linear (~ 1)
weakly non-linear (1~2)
strongly non-linear (> 2)
0e+00
1e+05
2e+05
Number
c
HCCS
Edf = 3.446, FDR = 3.32e-03
Slope = -0.003, FDR = 0.925
log2(TPM +1)
b
MED24
Edf = 1.000, FDR = 8.64e-06
Slope = -0.08, FDR = 3.44e-05
MAD1L1
Edf = 1.867, FDR = 2.52e-03
Slope = -0.09, FDR = 4.14e-03
FGG
Edf = 3.196, FDR = 3.66e-06
Slope = 0.14, FDR = 2.14e-04
4 WPC
5 WPC
6 WPC
7 WPC
8 WPC
9 WPC
10 WPC
11 WPC
12 WPC
13 WPC
16 WPC
18 WPC
19 WPC
20 WPC
Newborn
d
R = 0.17, p < 2.2e−16
Slope (log2(TMP+1))
Slope (TPM)
e
R = 0.84, p < 2.2e−16
Capped slope (log2(TMP+1))
Capped slope (TPM)
f
Density
AS.slope

Supplementary Fig. 12 | Validation of linear modeling approach as dynamic features.

a, Distribution of effective degrees of freedom (edf) from generalized additive models (GAM). Categories: linear (edf ≈ 1.0), weakly nonlinear (1 < edf < 2), strongly nonlinear (edf > 2).

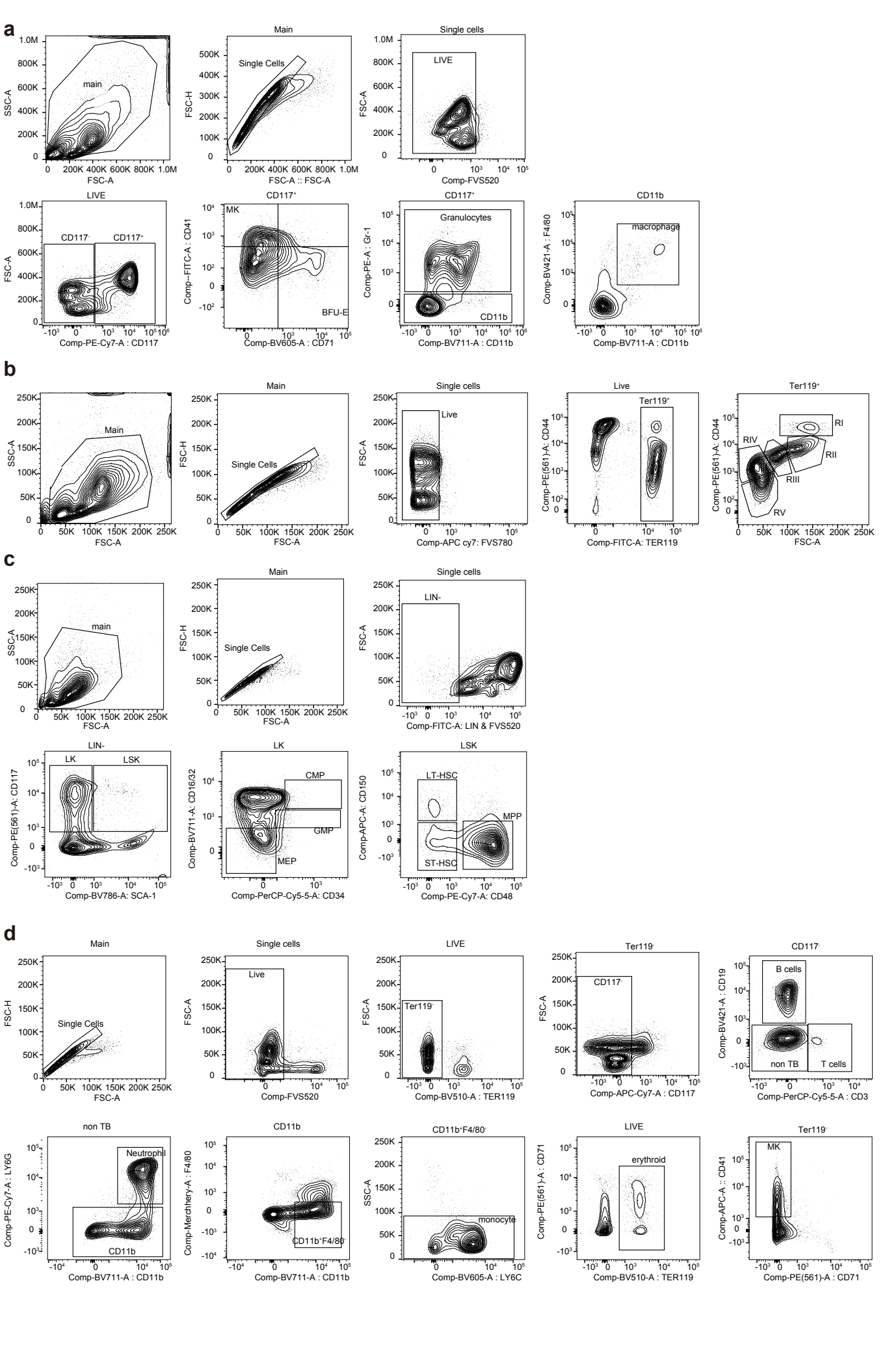
b, Dynamics of expression of representative genes (*MED24*, *MAD1L1*, *FGG*) across edf pattern categories. The yellow solid lines represent the GAM fits, while the blue dashed line indicates the linear regression fits. Gray shaded areas represent 95% confidence intervals.

c, Dynamics of expression of representative genes (*HCCS*) in high edf and near-zero slope. The yellow solid lines represent the GAM fits, while the blue dashed line indicates the linear regression fits. Gray shaded areas represent 95% confidence intervals.

d, Correlation between expression slopes of raw and log-transformed TPM expression per gene. Pearson's correlation coefficients (*R*) and p-value are labeled.

e, Correlation of capped slope between raw and log-transformed TPM expression per gene. Pearson's correlation coefficients (*R*) and p-value are labeled.

f, Density distribution of slopes of alternative splicing based on PSI.


a
main
SSC-A
FSC-A
Main
Single Cells
FSC-H
FSC-A :: FSC-A
Single cells
LIVE
FSC-A
Comp-FVS520
LIVE
CD117-
CD117+
Comp-PE-Cy7-A : CD117
CD117+
MK
BFU-E
Comp-FITC-A : CD41
Comp-BV605-A : CD71
CD117+
Granulocytes
CD11b
Comp-PE-A : Gr-1
Comp-BV711-A : CD11b
CD11b
macrophage
Comp-BV421-A : F4/80
Comp-BV711-A : CD11b
b
Main
SSC-A
FSC-A
Main
Single Cells
FSC-H
FSC-A
Single cells
Live
FSC-A
Comp-APC cy7: FVS780
Live
Ter119+
Comp-PE(561)-A: CD44
Comp-FITC-A: TER119
Ter119+
RI
RII
RIII
RIV
RV
Comp-PE(561)-A: CD44
FSC-A
c
main
SSC-A
FSC-A
Main
Single Cells
FSC-H
FSC-A
Single cells
LIN-
FSC-A
Comp-FITC-A: LIN & FVS520
LIN-
LK
LSK
Comp-PE(561)-A: CD117
Comp-BV786-A: SCA-1
LK
CMP
GMP
MEP
Comp-BV711-A: CD16/32
Comp-PerCP-Cy5-5-A: CD34
LSK
LT-HSC
MPP
ST-HSC
Comp-APC-A: CD150
Comp-PE-Cy7-A: CD48
d
Main
Single Cells
FSC-H
FSC-A
Single cells
Live
FSC-A
Comp-FVS520
LIVE
Ter119-
FSC-A
Comp-BV510-A : TER119
Ter119-
CD117-
FSC-A
Comp-APC-Cy7-A : CD117
CD117-
B cells
non TB
T cells
Comp-BV421-A : CD19
Comp-PerCP-Cy5-5-A : CD3
non TB
Neutrophil
CD11b
Comp-PE-Cy7-A : LY6G
Comp-BV711-A : CD11b
CD11b
Comp-Mrchhery-A : F4/80
CD11b+F4/80-
Comp-BV711-A : CD11b
CD11b+F4/80-
monocyte
SSC-A
Comp-BV605-A : LY6C
LIVE
erythroid
Comp-PE(561)-A : CD71
Comp-BV510-A : TER119
Ter119-
MK
Comp-APC-A :: CD41
Comp-PE(561)-A : CD71

Supplementary Fig. 13 | FACS gating strategy.

a, Gating strategy for the flow cytometric sorting of hematopoietic cell populations from colony-forming unit (CFU) assay.

b, Gating strategy for the flow cytometric analysis of erythroid cells in the bone marrow of the mouse.

c, Gating strategy for the flow cytometric analysis and sorting of hematopoietic stem and progenitor cells in the bone marrow of the mouse.

d, Gating strategy for the flow cytometric analysis and sorting of mature hematopoietic cell populations in the bone marrow of the mouse.