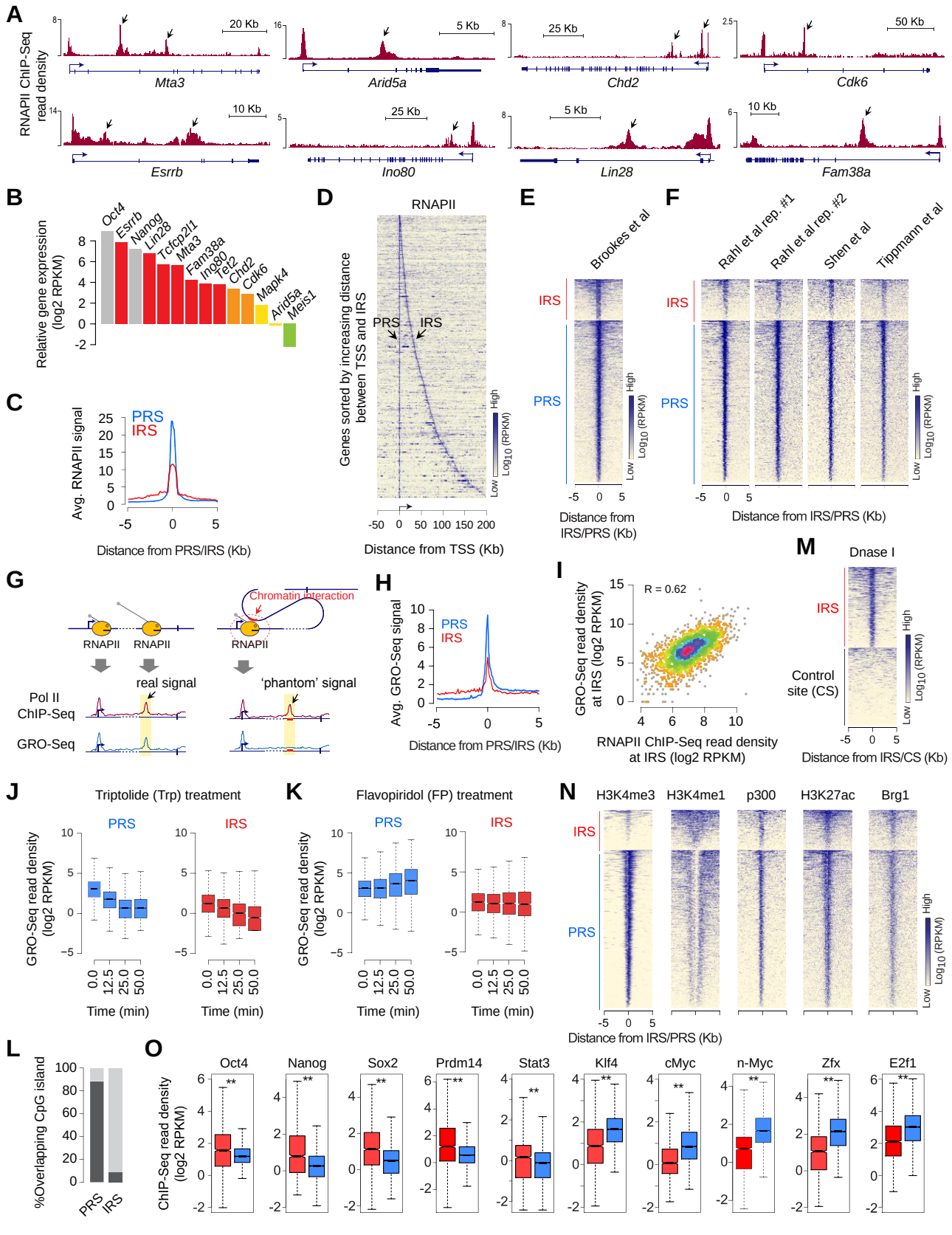
A
RNAPII ChIP-Seq read density
Mta3
Arid5a
Chd2
Cdk6
Esrrb
Ino80
Lin28
Fam38a
B
Relative gene expression (log2 RPKM)
Oct4
Esrrb
Nanog
Lin28
Tcfcp2l1
Mta3
Fam38a
Ino80
Tet2
Chd2
Cdk6
Mapk4
Arid5a
Meis1
C
Avg. RNAPII signal
PRS
IRS
Distance from PRS/IRS (Kb)
D
RNAPII
Genes sorted by increasing distance between TSS and IRS
Distance from TSS (Kb)
E
Brookes et al
Distance from IRS/PRS (Kb)
F
Rahl et al rep. #1
Rahl et al rep. #2
Shen et al
Tippmann et al
Distance from IRS/PRS (Kb)
G
RNAPII
Chromatin interaction
real signal
'phantom' signal
Pol II ChIP-Seq
GRO-Seq
H
Avg. GRO-Seq signal
Distance from PRS/IRS (Kb)
I
R = 0.62
GRO-Seq read density at IRS (log2 RPKM)
RNAPII ChIP-Seq read density at IRS (log2 RPKM)
M
Dnase I
Control site (CS)
Distance from IRS/CS (Kb)
J
Triptolide (Trp) treatment
GRO-Seq read density (log2 RPKM)
Time (min)
K
Flavopiridol (FP) treatment
N
H3K4me3
H3K4me1
p300
H3K27ac
Brg1
Distance from IRS/PRS (Kb)
L
%Overlapping CpG island
O
ChIP-Seq read density (log2 RPKM)
Oct4
Nanog
Sox2
Prdm14
Stat3
Klf4
cMyc
n-Myc
Zfx
E2f1

Figure S1 (related to Figures 1 and 2). Intragenic sites of RNAPII enrichment mark transcriptionally active intragenic enhancers

(A) ChIP-Seq profiles of RNAPII occupancy at select genes *Mta3*, Arid5a, *Chd2*, *Cdk6*, *Esrrb*, *Ino80*, *Lin28* and *Fam38a* (*Peizo1*) in mouse ESCs(Brookes et al., 2012).

(B) Expression (RPKM), as measured using RNA-Seq in mouse ESCs (Brookes et al., 2012), of IRS-containing genes (in color) and master ESC transcription factors *Oct4* and *Nanog* (gray). Genes are sorted by expression (red/gray, top quartile; orange, second quartile; yellow, third quartile; green, bottom quartile).

(C) Read density plot showing average RNAPII occupancy (RPKM) at PRSs and IRSs.

(D) Heatmap representation of relative levels of RNAPII ChIP-Seq read density at IRS-containing genes. Each row represents an IRS-containing gene, with the genes aligned at their TSSs (zero on the x-axis). Genes are sorted (top to bottom; y-axis) by increasing distance between TSS and IRS. About 90% of the IRS-containing genes, with IRSs within 200 Kb of their TSSs, are shown.

(E,F) Heatmap representation of relative levels of RNAPII ChIP-Seq read density around IRSs and PRSs in mouse ESCs(Brookes et al., 2012; Rahl et al., 2010; Shen et al., 2012; Tippmann et al., 2012). Sites (rows) within the IRS/PRS class are ordered (top-to-bottom) based on total read count within each row.

(G) Schematic showing how a site inferred to have a RNAPII ChIP-Seq peak could be an artifact (phantom peak) due to chromatin interaction it has with one or more other sites with RNAPII occupancy. GRO-Seq is insensitive to chromatin interactions (since it measures nascent RNA).

(H) Read density plot showing average nascent RNA enrichment (RPKM), as measured using GRO-Seq in mouse ESCs (Jonkers et al., 2014), at PRSs and IRSs.

(I) Scatter plot showing the correlation between the levels of RNAPII occupancy(Brookes et al., 2012) (x-axis) and nascent RNA(Jonkers et al., 2014) (GRO-Seq; y-axis) at IRSs in mouse ESCs.

(J,K) Nascent RNA dynamics at IRSs in mouse ESCs treated with Triptolide (J) or Flavopiridol (K). Box plots showing changes in GRO-Seq read density (RPKM) upon Triptolide or Flavopiridol treatment at PRSs and IRSs over time.

(L) Overlap with CpG islands (annotations downloaded from the UCSC genome browser).

(M) Heatmap representation of relative levels of DNase I hypersensitivity, as measured using DNase-Seq (ENCODE, GSE37074), at IRSs and matched non-IRS intragenic control sites in mouse ESCs. Sites (rows) within each class are ordered (top-to-bottom) based on total read count within each row.

(N) Heatmap representation of relative levels of H3K4me3 (ENCODE, GSE31039), H3K4me1 (Creyghton et al., 2010), H3K27ac (ENCODE, GSE31039), p300 (Creyghton et al., 2010),

and Brg1 (Ho et al., 2009) occupancy, as measured using ChIP-Seq, at PRSs and IRSs in mouse ESCs. Sites (rows) within the IRS/PRS class are ordered (top-to-bottom) based on total read count within each row.

(O) Box plot representation showing relative levels (RPKM) of binding of various transcription factors at IRSs and PRSs in mouse ESCs (Chen et al., 2008; Ho et al., 2011; Ma et al., 2011; Marson et al., 2008). **$P < 2.2 \times 10^{-16}$ (Wilcoxon-Mann-Whitney U test; two-sided).

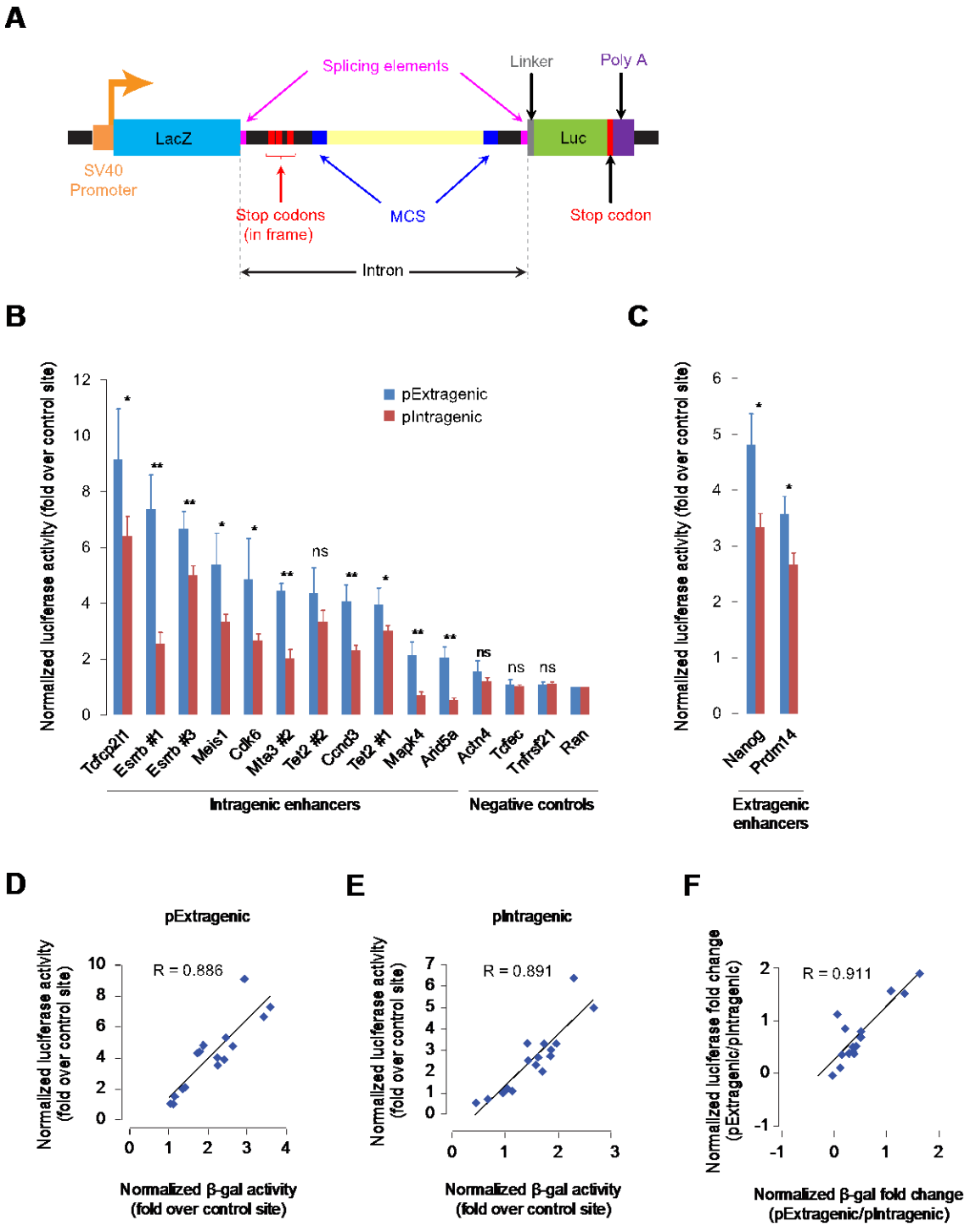

Figure S2 (related to Figure 3). Enhancers in intragenic position attenuate reporter activity

(A) Double-reporter construct encoding β-galactosidase (β-gal) and luciferase proteins within a single reading frame but on separate exons, with an intervening intron sequence containing stop codons (Nasim et al., 2002). MCS: multiple cloning site.

(B, C) Normalized luciferase activity, in mouse ESCs, from pIntragenic and pExtragenic reporter constructs (Figure 2C) cloned with intragenic enhancers or negative controls (B) or extragenic enhancers (C). Data are normalized to control sequence from Ran locus, which is set to 1. Data represents mean of *n* = 5 to 15 biological replicates. Error bars represent SEM. *P < 0.05; ** P < 0.01 (Student's t-test; two-sided); ns, not significant.

(D, E) Scatter plot showing the correlation between normalized luciferase activity (y-axis) and normalized β-gal activity (x-axis) observed using either pExtragenic (D) or pIntragenic (E) reporter constructs.

(F) Scatter plot showing the correlation between normalized luciferase fold change in pExtragenic vs pIntragenic reporter constructs (y-axis) and normalized β-gal fold change in pExtragenic vs pIntragenic reporter constructs (x-axis).

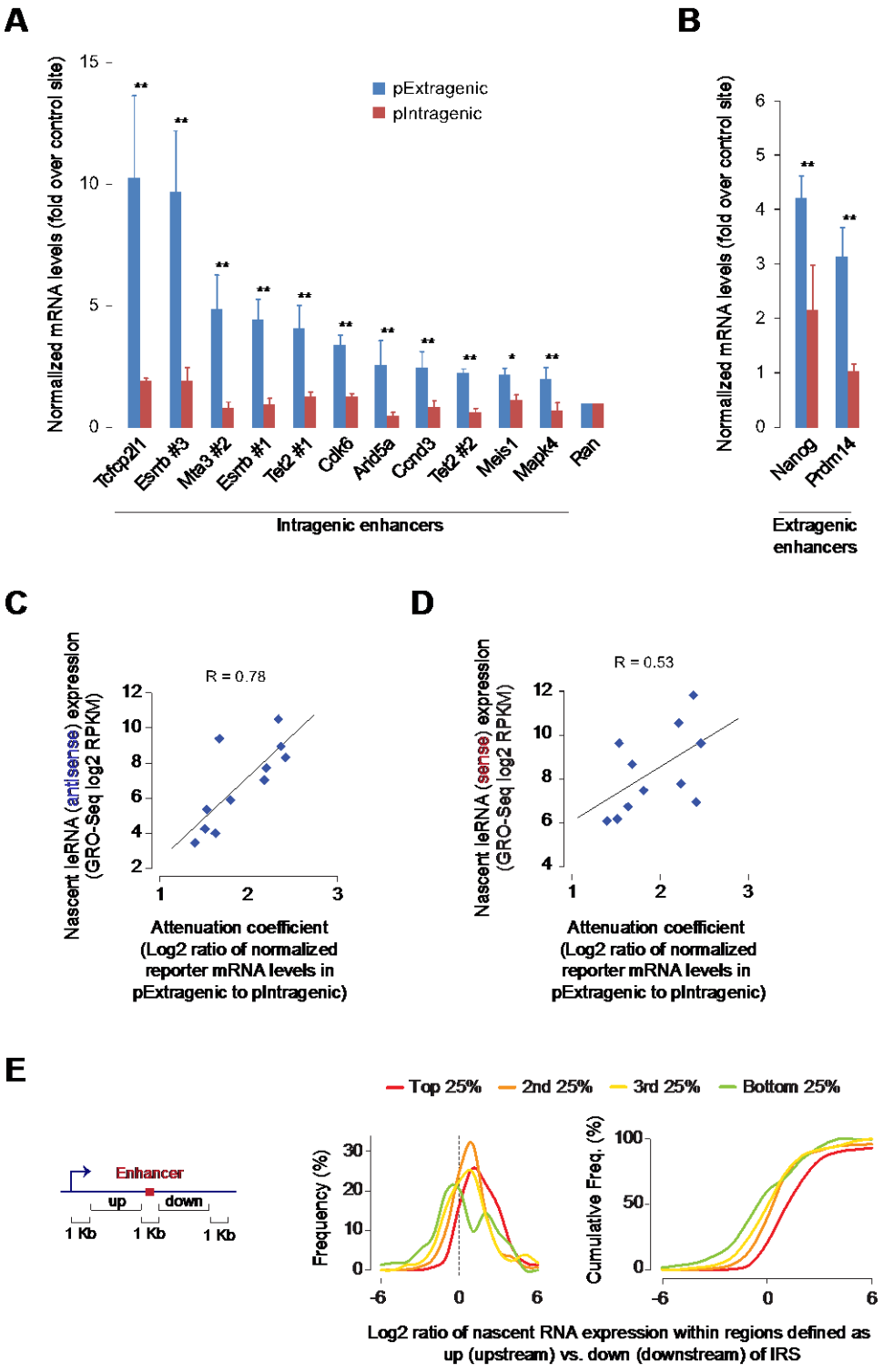

Figure S3 (related to Figure 3 and 4). Enhancers in intragenic position attenuate host gene transcription

(A, B) Normalized mRNA levels of the reporter gene, in mouse ESCs, from pIntragenic and pExtragenic reporter constructs cloned with intragenic enhancers (A) or extragenic enhancers (B). Data are normalized to control sequence from Ran locus, which is set to 1. Data represents mean of $n = 5$ biological replicates. Error bars represent SEM. *P < 0.05; ** P < 0.01 (Student's t-test; two-sided).

(C, D) Scatter plot, similar to the ones in Figure 4A, D but using the GRO-Seq data from Williams et al. (Williams et al., 2015), showing the correlation between attenuation coefficient, calculated based on data shown in Figure 3D, and levels of nascent antisense (C) or sense (D) ieRNA expression from native, genomic locus.

(E) *Left*: Schematic showing intragenic regions defined as upstream (up) and downstream (down) of intragenic enhancer. One kilobase (Kb) regions immediately downstream of TSS, immediately upstream of TES, and centered around the enhancer were excluded. *Middle*: Intragenic enhancers were binned into four groups based on their host gene expression. Frequency distribution of the ratios of GRO-Seq (Jonkers et al., 2014) read density within the 'up' region to that within the corresponding 'down' region for each group is shown. Only GRO-Seq reads from the sense strand was used. *Right*: Cumulative frequency distribution of the same data shown in the middle panel is shown.

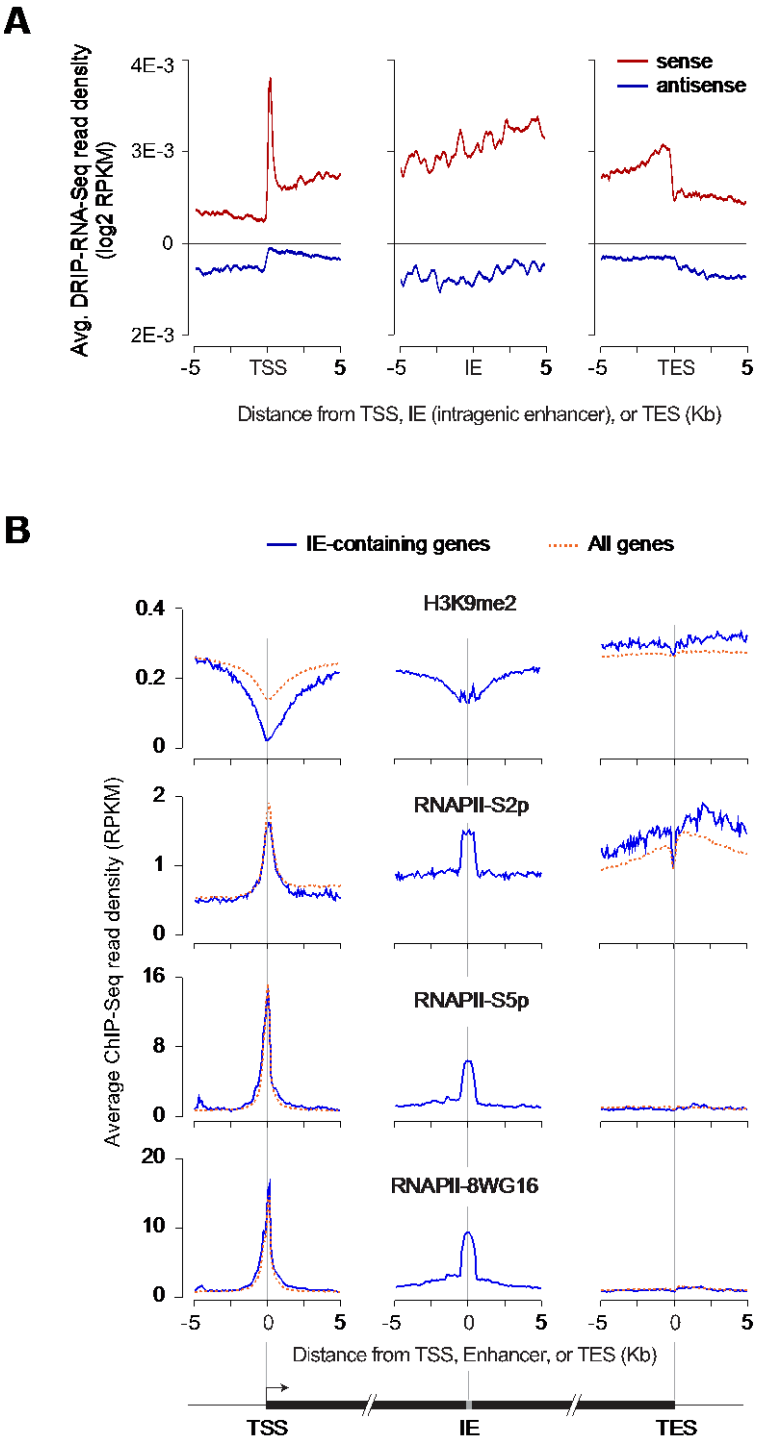

Figure S4 (related to Figure 5). Intragenic enhancers are not enriched for R-loops or R-loop-mediated RNAi-dependent H3K9me2

(A) Read density plot showing relative enrichment of R-loops, as measured using DRIP-RNA-Seq in mouse ESCs (Chen et al., 2015), near TSS, intragenic enhancer (IE), and TES of intragenic enhancer-containing genes in mouse ESCs.

(B) Read density plot showing relative levels (RPKM) of H3K9me2, RNAPII, and phosphorylated forms of RNAPII (RNAPII-S5P and RNAPII-S5P primarily associated with transcription initiation and elongation, respectively), as measured using ChIP-Seq in mouse ESCs (Brookes et al., 2012; Kurimoto et al., 2015; Rahl et al., 2010), near TSS, intragenic enhancer (IE), and TES of intragenic enhancer-containing genes in mouse ESCs.

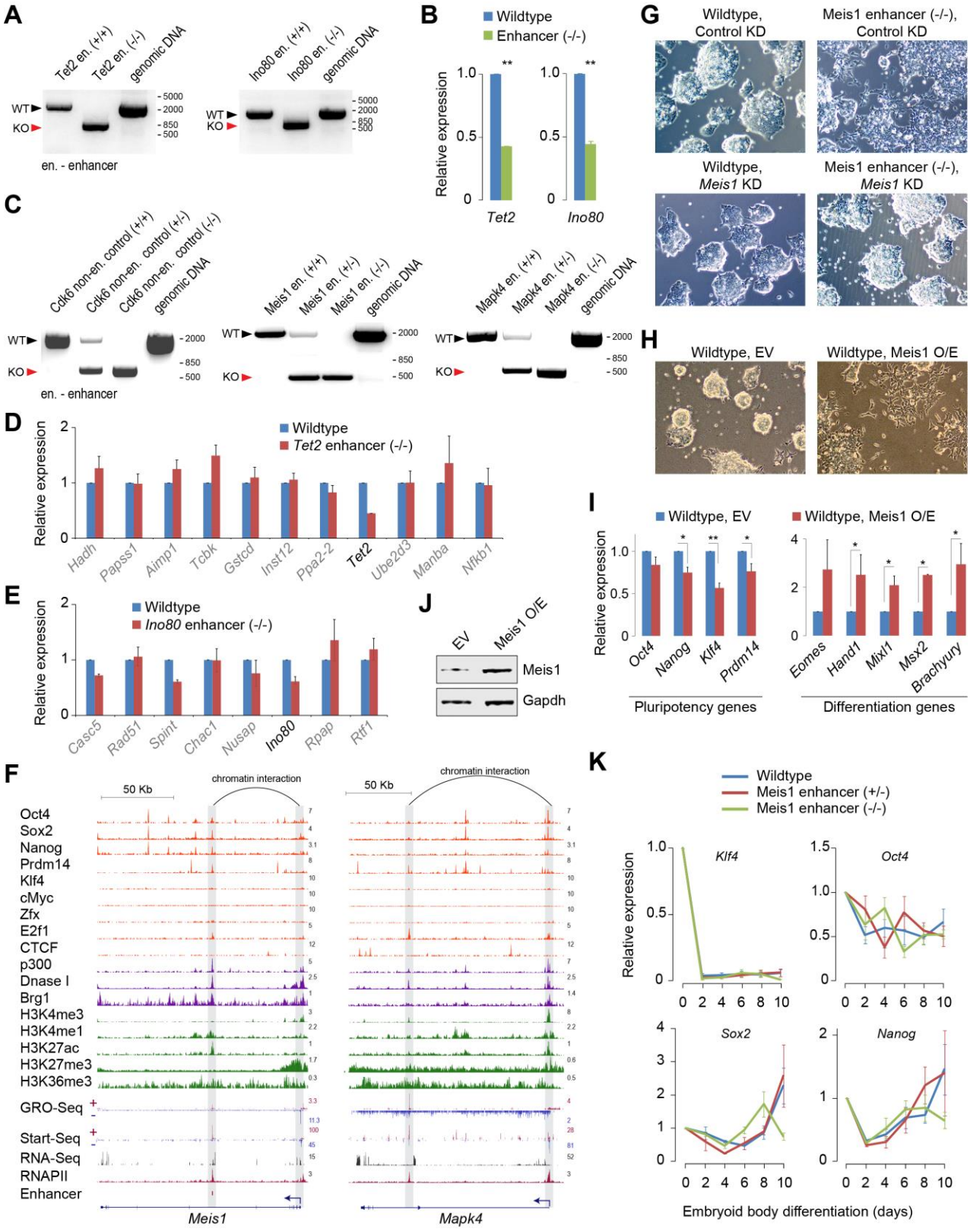
A
Tet2 en. (+/+)
Tet2 en. (-/-)
genomic DNA
Ino80 en. (+/+)
Ino80 en. (-/-)
genomic DNA
WT
KO
5000
2000
850
500
en. - enhancer
B
Wildtype
Enhancer (-/-)
Relative expression
1.0
0.5
0
**
Tet2
Ino80
C
Cdk6 non-en. control (+/+)
Cdk6 non-en. control (+/-)
Cdk6 non-en. control (-/-)
genomic DNA
Meis1 en. (+/+)
Meis1 en. (+/-)
Meis1 en. (-/-)
genomic DNA
Mapk4 en. (+/+)
Mapk4 en. (+/-)
Mapk4 en. (-/-)
genomic DNA
WT
KO
2000
850
500
en. - enhancer
D
Wildtype
Tet2 enhancer (-/-)
Relative expression
2
1
0
Hadh
Papss1
Aimp1
Tcbk
Gstcd
Inst12
Ppa2-2
Tet2
Ube2d3
Manba
Nfkb1
E
Wildtype
Ino80 enhancer (-/-)
Relative expression
2
1
0
Casc5
Rad51
Spint
Chac1
Nusap
Ino80
Rpap
Rtf1
F
50 Kb
chromatin interaction
Oct4
Sox2
Nanog
Prdm14
Klf4
cMyc
Zfx
E2f1
CTCF
p300
Dnase I
Brg1
H3K4me3
H3K4me1
H3K27ac
H3K27me3
H3K36me3
GRO-Seq
Start-Seq
RNA-Seq
RNAPII
Enhancer
Meis1
Mapk4
G
Wildtype, Control KD
Meis1 enhancer (-/-), Control KD
Wildtype, Meis1 KD
Meis1 enhancer (-/-), Meis1 KD
H
Wildtype, EV
Wildtype, Meis1 O/E
I
Wildtype, EV
Wildtype, Meis1 O/E
Relative expression
1.0
0.5
0
4
2
0
*
**
Oct4
Nanog
Klf4
Prdm14
Eomes
Hand1
Mixl1
Msx2
Brachyury
Pluripotency genes
Differentiation genes
J
EV
Meis1 O/E
Meis1
Gapdh
K
Wildtype
Meis1 enhancer (+/-)
Meis1 enhancer (-/-)
Relative expression
Klf4
Oct4
Sox2
Nanog
0 2 4 6 8 10
Embryoid body differentiation (days)

Figure S5 (related to Figure 6). CRISPR/Cas9-mediated genetic deletion of candidate intragenic enhancers and a role for intragenic enhancer-mediated attenuation in cell-fate determination

(A) Genotyping of genomic DNA from enhancer knockout (-/-) clonal cell lines, using standard gel electrophoresis, showing loss of the wildtype (WT) allele in enhancer (-/-) mouse ESCs. Black and red arrows mark the expected genomic products for screening-primers designed to evaluate deletions. Bands excised from the gel were subsequently sequence-verified.

(B) RT-qPCR analysis of mRNA levels of intragenic enhancer-containing genes *Tet2* and *Ino80* in corresponding wild-type and enhancer knockout (-/-) mouse ESCs. ** $P < 0.01$ (Student's t-test; two-sided)

(C) Genotyping of genomic DNA from enhancer heterozygous (+/-) and knockout (-/-) clonal cell lines, using standard gel electrophoresis, showing loss of WT allele(s) in enhancer (+/-) and enhancer (-/-) mouse ESCs. Black and red arrows mark the expected genomic products for screening-primers designed to evaluate deletions. Bands excised from the gel were subsequently sequence-verified.

(D, E) mRNA levels of genes within *Tet2* (D) or *Mapk4* (E) locus in wildtype or *Tet2* (*Ino80*, respectively) enhancer (-/-) ESCs. Data are normalized to *Actin*. Error bars represent SEM of three biological replicates.

(F) Genome browser shot of Meis1 and Mapk4 loci showing ChIP-Seq read density profiles of RNAPII, GRO-Seq, Start-Seq, RNA-Seq, various transcription regulators and chromatin remodelers, and histone modifications in mouse ESCs. Known chromatin interaction between Meis1/Mapk4 intragenic enhancer and its promoter in mouse ESCs (Schoenfelder et al., 2015) is shown at the top.

(G) Colony morphology of wildtype and Meis1 enhancer (-/-) mouse ESCs upon control or *Meis1* KD. Representative images, taken 96 h after siRNA transfection, are shown.

(H) Colony morphology of mouse ESCs overexpressing Meis1 or an empty vector (EV). Representative images, taken 96 h after plasmid transfection, are shown.

(I) mRNA levels of pluripotency-associated ESC identity genes (left) and differentiation/developmental genes (right) in ESCs overexpressing Meis1 or an empty vector (EV). Data are normalized to *Actin*. Error bars represent SEM of three biological replicates. *$P < 0.05$; **$P < 0.01$ (Student's t-test; two-sided).

(J) Western blot analysis of Meis1 (endogenous+exogenous) in mouse ESCs expressing EV or endogenous Meis1. Gapdh is used as a loading control.

(K) mRNA levels of ESC identity genes during embryoid body formation of wildtype, Meis1 enhancer heterozygous (+/-) and knockout (-/-) mouse ESCs. Data are normalized to *Actin*. Error bars represent SEM of three biological replicates.

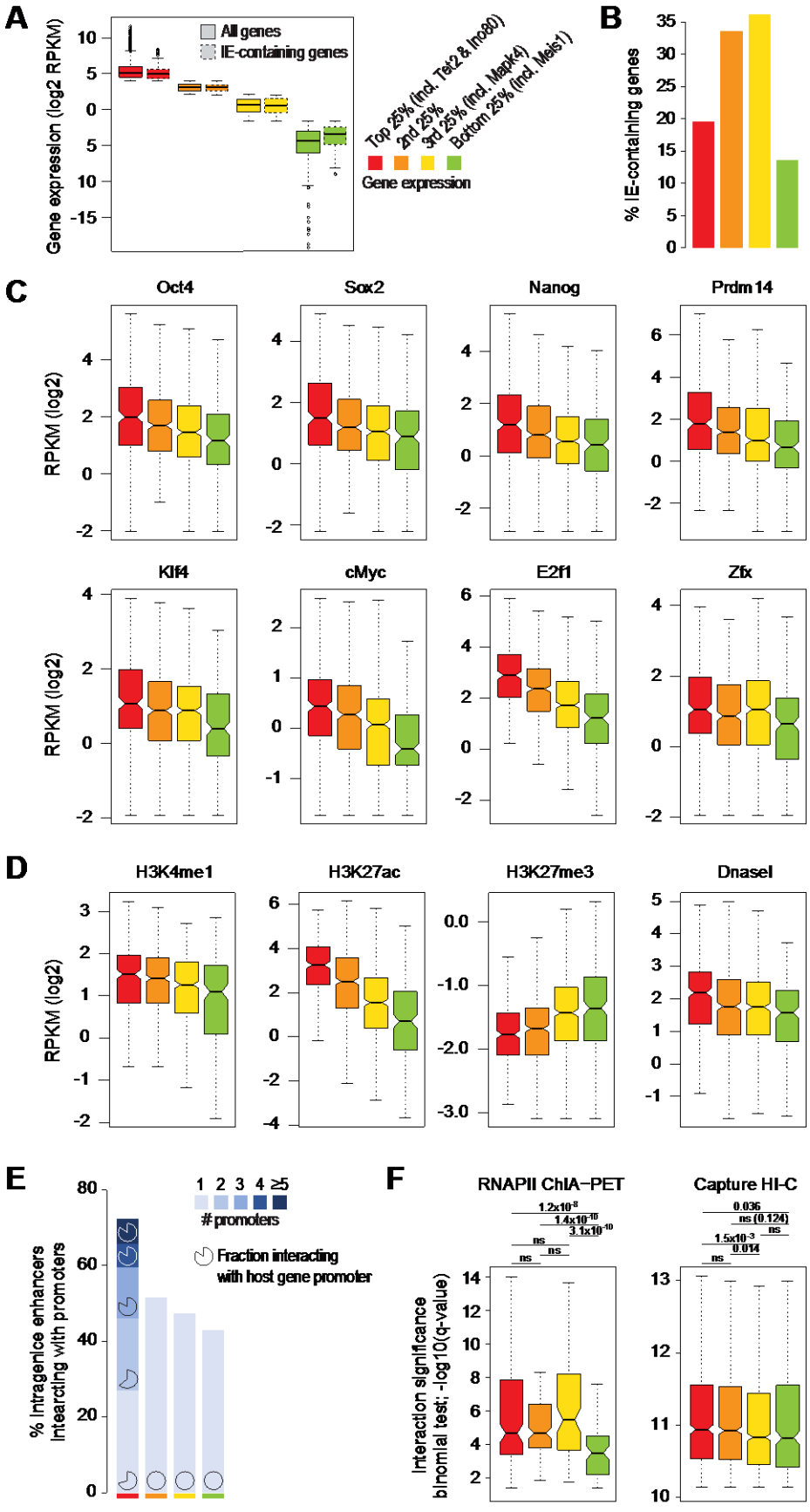
A
Gene expression (log2 RPKM)
All genes
IE-containing genes
Top 25% (incl. Tet2 & Ino80)
2nd 25%
3rd 25% (incl. Mapk4)
Bottom 25% (incl. Meis1)
Gene expression
B
% IE-containing genes
C
Oct4
Sox2
Nanog
Prdm14
Klf4
cMyc
E2f1
Zfx
RPKM (log2)
D
H3K4me1
H3K27ac
H3K27me3
DnaseI
RPKM (log2)
E
% intragenice enhancers interacting with promoters
1 2 3 4 ≥5
#promoters
Fraction interacting with host gene promoter
F
RNAPII ChIA−PET
Capture HI-C
Interaction significance binomial test; -log10(q-value)
1.2x10-8
1.4x10-10
3.1x10-10
ns
0.036
ns (0.124)
1.5x10-3
0.014

Figure S6 (related to Figure 6). Enhancer strength and attenuation

(A) The set of all genes were binned into four quartiles based on their expression in mouse ESCs, and the expression of intragenic enhancer (IE)-containing (host) genes within each quartile is shown.

(B) Percentage of IE-containing genes within each of the four quartiles; color code similar to that in (A).

(C, D) Intragenic enhancers were binned into four groups based on their host gene expression; color code similar to that in (A). Intragenic enhancers within *Tet2* and *Ino80* are in the first quartile (top 25%) and those within *Mapk4* and *Meis1* and are in the third and fourth quartiles, respectively. Box plots show relative levels (RPKM) of transcription factor occupancy (A) and enhancer-associated chromatin signatures (B) at intragenic enhancers in mouse ESCs (Chen et al., 2008; Creyghton et al., 2010; Ma et al., 2011; Marson et al., 2008) (ENCODE, GSE31039 and GSE37074).

(E) Bar plot shows the fraction of intragenic enhancers, from within each group, interacting with one or more gene promoters; enhancer-promoter interaction data from RNAPII ChIA-PET (Zhang et al., 2013) and Capture Hi-C (Schoenfelder et al., 2015) experiments in mouse ESCs was used. Shades of blue represent number of gene promoters enhancers interact with. Pie-charts are used to represent fraction of intragenic enhancers, within each class, that interact with host gene promoter.

(F) Intragenic enhancers were binned into four groups (red, orange, yellow, and green) based on their host gene expression; color code similar to that in (A). Box plots show relative enhancer-promoter interaction significance, as assessed using RNAPII ChIA-PET (Zhang et al., 2013) and Capture Hi-C (Schoenfelder et al., 2015) experiments, in mouse ESCs. P-values shown were computed using the Wilcoxon-Mann-Whitney U test (two-sided).

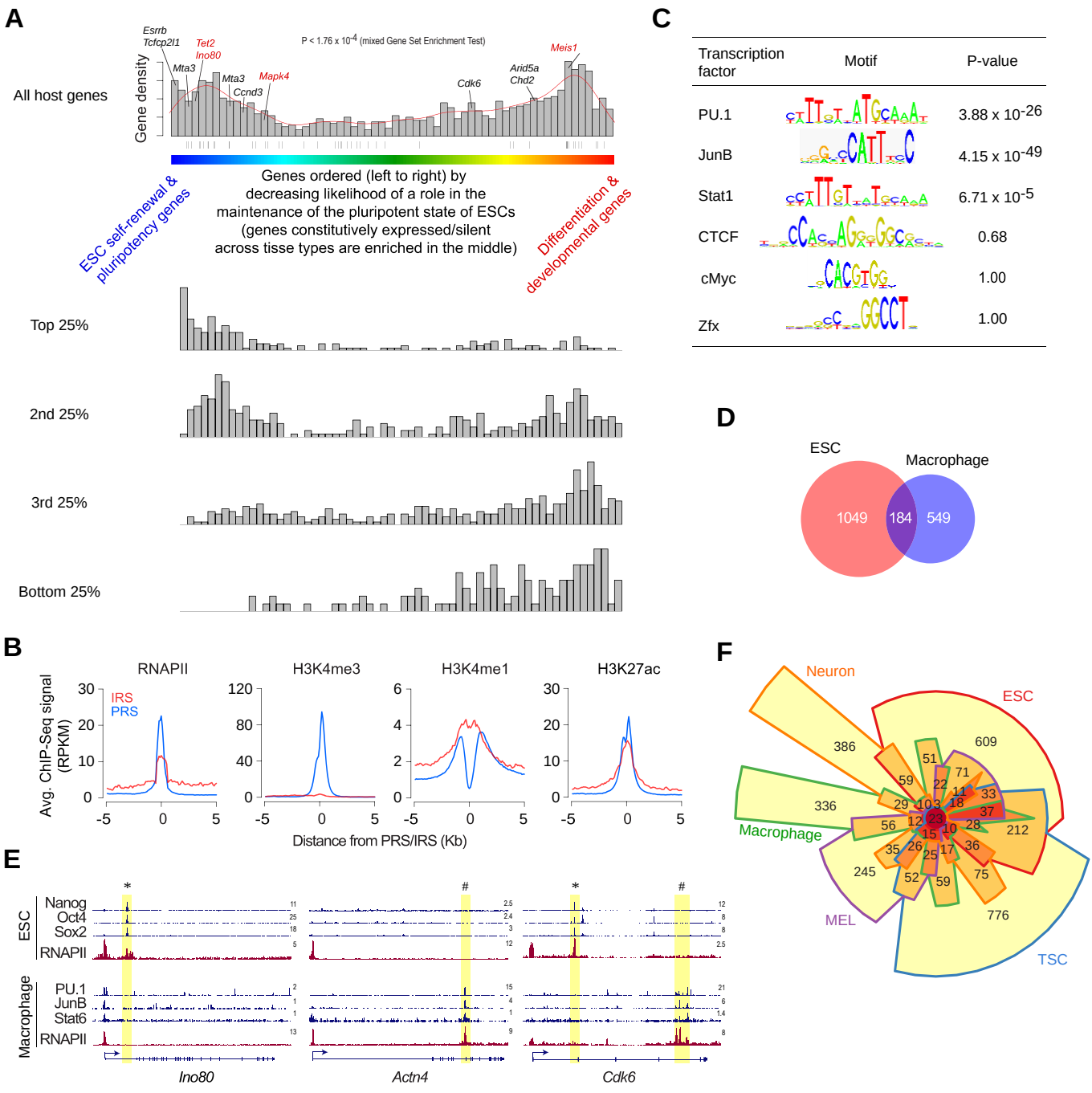
A
All host genes
Gene density
Esrrb
Tcfcp2l1
Tet2
Ino80
Mta3
Mta3
Ccnd3
Mapk4
Cdk6
Arid5a
Chd2
Meis1
P < 1.76 x 10-4 (mixed Gene Set Enrichment Test)
Genes ordered (left to right) by decreasing likelihood of a role in the maintenance of the pluripotent state of ESCs (genes constitutively expressed/silent across tisse types are enriched in the middle)
ESC self-renewal & pluripotency genes
Differentiation & developmental genes
Top 25%
2nd 25%
3rd 25%
Bottom 25%
C
Transcription factor
Motif
P-value
PU.1
3.88 x 10-26
JunB
4.15 x 10-49
Stat1
6.71 x 10-5
CTCF
0.68
cMyc
1.00
Zfx
1.00
D
ESC
Macrophage
1049
184
549
B
RNAPII
H3K4me3
H3K4me1
H3K27ac
Avg. ChIP-Seq signal (RPKM)
IRS
PRS
Distance from PRS/IRS (Kb)
E
ESC
Nanog
Oct4
Sox2
RNAPII
Macrophage
PU.1
JunB
Stat6
RNAPII
Ino80
Actn4
Cdk6
F
Neuron
ESC
Macrophage
MEL
TSC
386
609
336
212
245
776
G
ESC
Cell-cell adhesion
Establishment of cell polarity
Embryonic development
Phosphorylation
Cell morphogenesis
Chromatin modification
Embryonic organ development
Neg. regulation of DNA binding
Regulation of transcription
Stem cell differentiation
Macrophage
DNA metabolic process
Cellular response to stress
Response to DNA damage stimulus
Response to wounding
Cell activation
Leukocyte activation
Defense response
Inflammatory response
Innate immune response
Lymphocyte activation
Terminal neuron
Behavior
Synaptic transmission
Chloride transport
Ion transport
Transmission of nerve impulse
Chromosome organization
Regulation of transmission of nerve impulse
Cell-cell signaling
Regulation of neuron apoptosis
Nerve-nerve synaptic transmission
Trophoblast stem cell
Regulation of transcription
Cell morphogenesis
Transcription
Regulation of RNA metabolic process
Cellular component morphogenesis
Neg. regulation of cellular biosynthetic process
Pos. regulation of gene expression
Neg. regulation of biosynthetic process
Neg. regulation of macromolecule biosynthetic process
Pos. regulation of transcription
MEL
Regulation of T cell proliferation
Lipid biosynthetic process
Phosphorylation
Indole derivative biosynthetic process
Indolalkylamine biosynthetic process
Regulation of calcium ion transport
Monosaccharide metabolic process
Icosanoid metabolic process
Regulation of mononuclear cell proliferation
Regulation of lymphocyte proliferation
10-2
P-value
10-6

Figure S7 (related to Figure 6). Functional characterization of intragenic enhancer-containing genes

(A) Functional characterization of intragenic enhancer-containing genes. *Top*: Frequency distribution of all intragenic enhancer-containing (host) genes in relation to list of genes, rank-ordered (left to right; x-axis) based on the likelihood of their role in the maintenance of mouse ESC identity (Cinghu et al., 2014). Candidate intragenic enhancer-containing genes are highlighted. Genes harboring intragenic enhancers are enriched for ESC identity genes (distribution on the left) and developmental/differentiation genes (distribution on the right). *Bottom*: Host genes were binned into four groups based on their host gene expression (high to low). For each group, frequency distribution (same as that in the top panel) is shown. Host genes with expression in the top 25% are enriched for ESC self-renewal and pluripotency genes. In contrast, host genes with expression in the bottom 25% are enriched for differentiation and developmental genes.

(B) Read density plots showing relative levels (RPKM) of H3K4me3, H3K4me1 and H3K27ac at IRSs in mouse bone marrow-derived macrophages.

(C) Relative enrichment of TF sequence recognition motifs within macrophage IRSs

(D) Overlap between IRS-associated genes in mouse ESCs and macrophages.

(E) Genome browser shots of select genes showing mouse ESC-specific (*) and macrophage-specific (#) IRSs bound by cell-specific master transcription factors.

(F) Chow-Ruskey diagram showing five-way overlap of intragenic enhancer-associated genes in mouse ESCs (polygon with a red colored border), macrophages (green polygon), terminal neurons (orange polygon), trophoblast stem cells (TSCs) (blue polygon) and murine erythroleukemia (MEL) (purple polygon) cells. Regions with yellow fill color mark cell type-specific intragenic enhancer-associated genes. Regions marked by shades of orange and red mark intersection between cell types. Regions with light and dark shade of orange for fill color represents intersection between two and three cell types, respectively. Regions with red shade for fill color represents intersection between four cell types and the region with reddish brown shade for fill color (the circle in the middle) represents the overlap of all five cell types. Area of each intersection is proportional to number of genes within the intersection.

(G) Intragenic enhancer-associated genes have prominent roles in cell type-specific biology. Gene ontology (GO) categories enriched within intragenic enhancer-associated genes genes in mouse ESCs, macrophages, terminal neurons, trophoblast stem cells, and murine erythroleukemia (MEL) cells. P-value for each enriched GO category is displayed as a blue square, with color scale bar denoted below.