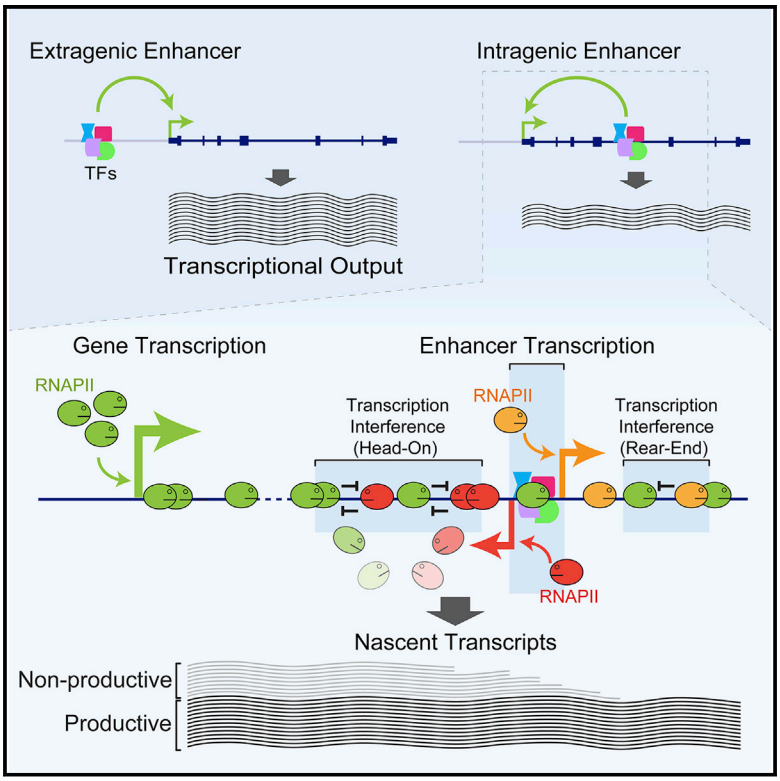
Article

Molecular Cell

# Intragenic Enhancers Attenuate Host Gene Expression

## Graphical Abstract




## Authors

Senthilkumar Cinghu, Pengyi Yang, Justin P. Kosak, ..., Andrew J. Oldfield, Karen Adelman, Raja Jothi

## Correspondence

jothi@nih.gov

## In Brief

Cinghu et al. report an unanticipated role for intragenic enhancers in attenuating host gene expression. They show that transcription at intragenic enhancers interferes with and attenuates host gene transcription during productive elongation. Genetic experiments reveal a physiological role for intragenic enhancer-mediated attenuation in cell fate choice during embryonic stem cell differentiation.

## Highlights

- Intragenic enhancers, besides activating genes, also attenuate host gene expression
- Transcription at intragenic enhancers interferes with host gene transcription
- The act of enhancer transcription alone, but not the eRNA, explains the attenuation
- Intragenic enhancer-mediated attenuation determines cell fate choice

Cinghu et al., 2017, Molecular Cell *68*, 104–117
October 5, 2017 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.molcel.2017.09.010

CellPress

Molecular Cell
Article

# Intragenic Enhancers Attenuate Host Gene Expression

Senthilkumar Cinghu,[1,4] Pengyi Yang,[1,2,4] Justin P. Kosak,[1] Amanda E. Conway,[1] Dhirendra Kumar,[1] Andrew J. Oldfield,[1] Karen Adelman,[1,3] and Raja Jothi[1,5,*]

[1]Epigenetics & Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC 27709, USA

[2]Present address: Charles Perkins Centre and School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006, Australia

[3]Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

[4]These authors contributed equally

[5]Lead Contact

*Correspondence: jothi@nih.gov

http://dx.doi.org/10.1016/j.molcel.2017.09.010

## SUMMARY

**Eukaryotic gene transcription is regulated at many steps, including RNA polymerase II (Pol II) recruitment, transcription initiation, promoter-proximal Pol II pause release, and transcription termination; however, mechanisms regulating transcription during productive elongation remain poorly understood. Enhancers, which activate gene transcription, themselves undergo Pol II-mediated transcription, but our understanding of enhancer transcription and enhancer RNAs (eRNAs) remains incomplete. Here we show that transcription at intragenic enhancers interferes with and attenuates host gene transcription during productive elongation. While the extent of attenuation correlates positively with nascent eRNA expression, the act of intragenic enhancer transcription alone, but not eRNAs, explains the attenuation. Through CRISPR/Cas9-mediated deletions, we demonstrate a physiological role for intragenic enhancer-mediated transcription attenuation in cell fate determination. We propose that intragenic enhancers not only enhance transcription of one or more genes from a distance but also fine-tune transcription of their host gene through transcription interference, facilitating differential utilization of the same regulatory element for disparate functions.**

## INTRODUCTION

Eukaryotic gene transcription is regulated at many steps (Adelman and Lis, 2012; Jonkers and Lis, 2015; Porrua and Libri, 2015; Proudfoot, 2016). Emerging evidence points to much of the transcription regulation occurring well after RNA polymerase II (Pol II) recruitment, through controlled pause and release of promoter-proximal Pol II during early elongation (Adelman and Lis, 2012; Jonkers and Lis, 2015; Levine, 2011). After its regulated release into productive elongation, Pol II is generally assumed to processively progress through the gene, terminate, and eventually reinitiate transcription (Adelman and Lis, 2012). However, even after promoter-proximal pause release, Pol II must still contend with further roadblocks as it transcribes through the length of the gene body (Li et al., 2007; Teves et al., 2014). Certain DNA sequences are more difficult to transcribe than others. Intrinsic pause sites, even on naked DNA, make Pol II susceptible to transient pausing, which is exacerbated when Pol II encounters obstacles such as nucleosomes and R-loops, leading to backtracking of Pol II and transcriptional arrest (Bintu et al., 2012; Kireeva et al., 2005; Li et al., 2007; Teves et al., 2014). Arrested Pol II is typically reactivated by the prototypic transcription elongation factor TFIIS (Cheung and Cramer, 2011; Kireeva et al., 2005) or targeted for proteolytic degradation (Proudfoot, 2016).

Pol II pausing has also been observed to occur at intron-exon junctions and near 3′ cleavage/polyadenylation sites, coinciding with splicing factor recruitment (Kwak et al., 2013; Mayer et al., 2015; Nojima et al., 2015). This transient pause event is thought to provide sufficient time for spliceosome assembly and splicing, which would be consistent with exons having the strongest negative effect on elongation rates, delaying transcription through a gene by ~20–30 s per exon (Jonkers et al., 2014; Jonkers and Lis, 2015; Martin et al., 2013). Consequently, elongation rates differ greatly among genes (Danko et al., 2013; Jonkers et al., 2014). Although Pol II elongation rates correlate with gene expression (Danko et al., 2013), our understanding of mechanisms regulating transcription during productive elongation remains incomplete.

Precise spatiotemporal patterns of gene expression during development are regulated through integrated action of many *cis*-regulatory elements, which include promoters, enhancers, silencers, and insulators. Among this constellation of elements, enhancers, often located at greater distances from their target promoters, and their associated transcription factors (TFs) play a leading role in the activation of transcription (Spitz and Furlong, 2012). A hallmark of enhancers that has been repeatedly demonstrated is that they are relatively insensitive to distance or position relative to their target genes and can activate transcription independent of their orientation (Calo and Wysocka, 2013; Kim and Shiekhattar, 2015; Li et al., 2016; Shlyueva et al., 2014). Several studies have established that many, if not all, functionally active enhancers themselves undergo Pol II-mediated transcription and produce short enhancer-derived RNAs (eRNAs) (De Santa et al., 2010; Heinz et al., 2015; Kim et al., 2010; Kim and Shiekhattar, 2015; Li et al., 2016; Natoli and Andrau, 2012; Shlyueva et al., 2014). Distinct from mRNAs, eRNAs are

generally unstable, less abundant, and rapidly degraded by exosomes (Pefanis et al., 2015). Changes in eRNA expression highly correlate with changes in nearby gene expression; numerous studies have shown that knockdown of eRNAs accompanies a decrease in the expression of corresponding target genes, with several lines of evidence supporting a functional role for eRNAs in perhaps all stages of gene activation (Kim and Shiekhattar, 2015; Li et al., 2016; Schaukowitch et al., 2014). But whether the act of enhancer transcription per se regulates gene expression remains unclear.

Although about half of all annotated enhancers are intragenic (Andersson et al., 2014; Kowalczyk et al., 2012; Shen et al., 2012; Whyte et al., 2013), most, if not all, studies of enhancer function were conducted by cloning the enhancer of interest, even if it is intragenic, either upstream or downstream of transient/transgenic reporters. Consequently, the effects of intragenic enhancer transcription on host gene expression have not been addressed. Here we report that functionally active intragenic enhancers present yet another obstacle for Pol II transcribing the host gene. Our studies show that intragenic enhancers, besides activating genes, also attenuate host gene expression. While the extent of attenuation correlates positively with nascent eRNA expression, the act of intragenic enhancer transcription alone, but not eRNAs, explains the attenuation. Through CRISPR/Cas9-mediated deletions, we demonstrate a functional role for intragenic enhancer-mediated transcription attenuation in cell fate determination. Our findings suggest that the intragenic enhancer-mediated transcription attenuation could represent a general mechanism to attenuate and fine-tune host gene transcription during productive elongation.

## RESULTS

### Intragenic Sites of Pol II Enrichment

To gain insight into the regulation of transcription during productive elongation, we examined genome-wide Pol II chromatin immunoprecipitation sequencing (ChIP-seq) data in mouse embryonic stem cells (ESCs) (Brookes et al., 2012). We noted intragenic sites of Pol II enrichment tens of thousands of base pairs downstream of transcription start sites (TSSs) (Figures 1A and S1A), within genes with a wide range of expression (Figure S1B), and we hypothesized that they might reflect sites regulating transcription during productive elongation. Toward testing this theory, first, we systematically identified intragenic sites of Pol II enrichment by calculating Pol II Pausing Index (PI) (Adelman and Lis, 2012), defined as the ratio of Pol II density at an intragenic site of interest to the median Pol II density within the gene body (Figure 1B). Using a stringent threshold (PI $\geq$ 10), we identified 1,928 intragenic RNA Pol II sites (IRSs), located at least 1 kb away from TSSs and transcription end sites (TESs) of all known and predicted genes, which by definition excludes Pol II pause sites immediately downstream of promoters and near 3′ cleavage/polyadenylation sites (Table S1). For comparison purposes, we also identified 7,530 promoter-proximal RNA Pol II sites (PRSs), located within 500 bp downstream of TSSs (Table S1; Figure S1C).

IRSs were distributed across gene bodies (Figure 1C), with a median distance of ~40 kb from TSSs (Figure S1D). After confirming that the Pol II enrichment at IRSs is consistently observed across other Pol II datasets in ESCs (Figures S1E and S1F) (Rahl et al., 2010; Shen et al., 2012; Tippmann et al., 2012), we wondered whether IRSs might represent sites of phantom Pol II ChIP enrichment explainable by cross-linking artifacts resulting from their interaction with one or more Pol II-enriched promoters (Figure S1G). Our analyses of data from global run-on sequencing (GRO-seq) (Jonkers et al., 2014), which measures nascent RNA from transcriptionally engaged Pol II, revealed an enrichment for GRO-seq reads at IRSs (Figure S1H) and a strong positive correlation between Pol II and GRO-seq signals (Figure S1I), indicating that Pol II enrichment at IRSs is unlikely to be a ChIP artifact.

We considered two potential explanations for Pol II enrichment at IRSs: Pol II pausing during productive elongation and/or transcription initiation. To discriminate between the two possibilities, we analyzed nascent RNA from ESCs treated with Triptolide (Trp), which blocks transcription initiation altogether, or Flavopiridol (FP), which allows transcription initiation but blocks P-TEFb-dependent promoter-proximal Pol II pause release (Jonkers et al., 2014). Trp effectively blocked transcription initiation, resulting in the clearing of Pol II from promoters and gene bodies over time (Figures 1D, 1E, and S1J). FP, in contrast, allowed Pol II accumulation at its recruitment sites but blocked the release of paused Pol II, thus resulting in the clearing of Pol II from gene bodies except for at IRSs (Figures 1D, 1E, and S1K). These data suggested that Pol II enrichment at IRSs is due to intragenic transcription initiation rather than intrinsic Pol II pausing during productive synthesis. Indeed, our analysis of data from Start-seq (Williams et al., 2015), which measures transcription initiation-associated RNA, confirmed bidirectional transcription initiation from IRSs (Figures 1E and 1F).

### Intragenic Sites of Pol II Enrichment Mark Transcriptionally Active Intragenic Enhancers

To determine whether IRSs represent unannotated gene promoters or enhancers, which themselves undergo Pol II-mediated transcription (De Santa et al., 2010; Kim et al., 2010; Kim and Shiekhattar, 2015; Li et al., 2016), we examined previously defined stereotypical chromatin signatures that distinguish enhancers from promoters (Calo and Wysocka, 2013; Kim and Shiekhattar, 2015; Li et al., 2016; Shlyueva et al., 2014). Unlike annotated and predicted promoters, ~70% of which overlap CpG islands, only ~7% of IRSs overlap CpG islands, which is more in line with the ~5% overlap reported for enhancers (Natoli and Andrau, 2012) (Figure S1L). Furthermore, examination of ~4.4M mouse-expressed sequence tags (ESTs) revealed enrichment for ends of ESTs at annotated TSSs, but not at IRSs (Figure 2A), suggesting that IRSs are less likely to represent unannotated gene promoters.

Further characterization revealed that IRSs exhibit an accessible chromatin architecture—marked by enhanced DNase I hypersensitivity—accompanied by high levels of histone modifications H3K4me1 and H3K27ac, associated with active enhancers, but not H3K4me3, associated with active promoters (Calo and Wysocka, 2013; Creyghton et al., 2010; Ho et al., 2009; Kim and Shiekhattar, 2015; Li et al., 2016; Shlyueva

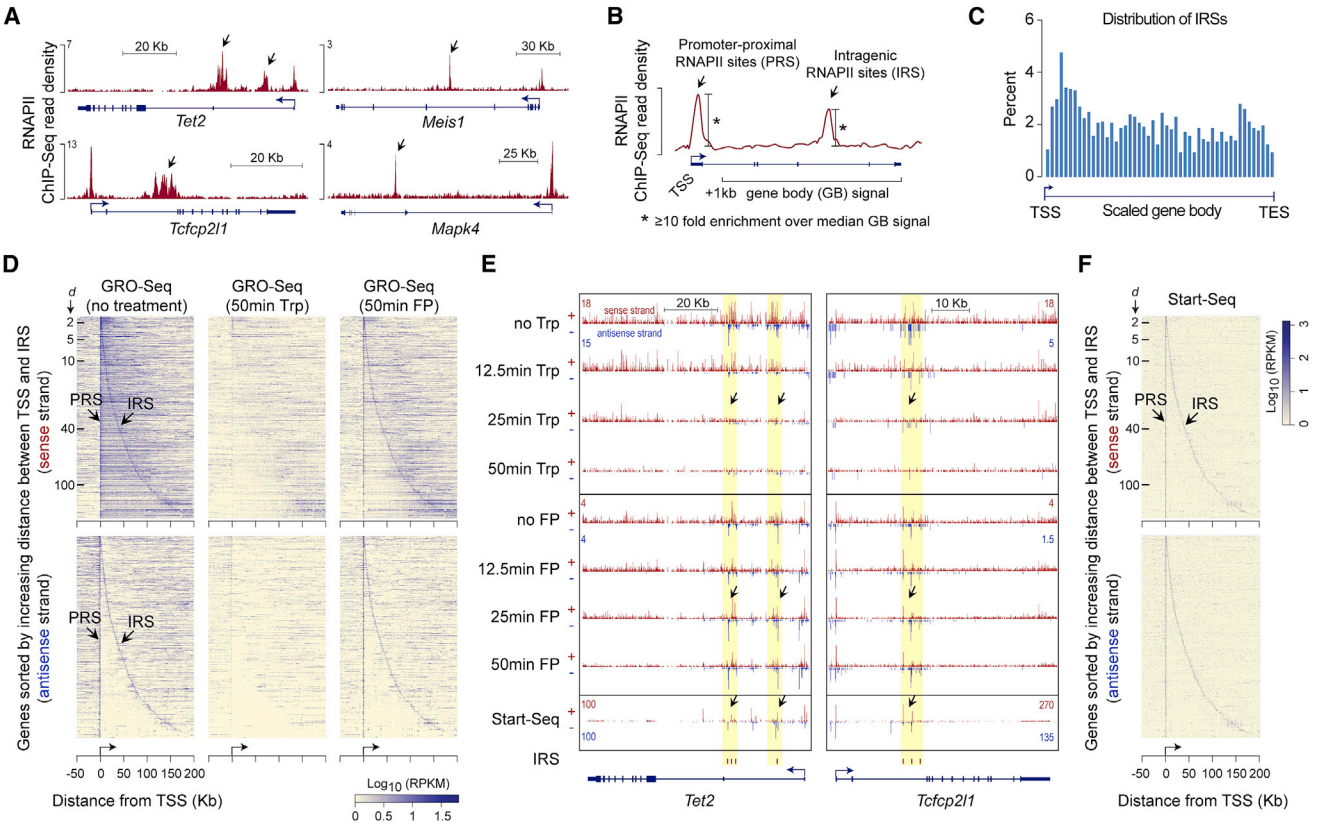



**Figure 1. Characterization of Intragenic Pol II Sites**

(A) ChIP-seq profiles of RNA Pol II (RNAPII) occupancy at select genes in mouse ESCs (Brookes et al., 2012).

(B) Schematic describing the calculation used to determine promoter-proximal RNA Pol II sites (PRSs) and intragenic RNA Pol II sites (IRSs).

(C) Histogram showing distribution of IRSs within genes.

(D) Heatmap representation of nascent RNA expression, as measured using GRO-seq (Jonkers et al., 2014) in untreated mouse ESCs, near IRS-containing genes. Each row represents an IRS-containing gene, with genes aligned at their TSSs. Genes are sorted (top to bottom) by increasing distance between TSS and IRS; the distance (*d*, kb) between TSS and IRS is indicated by tick marks. The top and bottom panels show data from sense and antisense strands, respectively. About 90% of the IRS-containing genes, with IRSs within 200 kb of their TSSs, are shown.

(E) Genome browser shots of genes *Tet2* and *Tcfcp2l1* showing GRO-seq read density in ESCs treated with Triptolide (Trp) or Flavopiridol (FP) for 12.5, 25, and 50 min (Jonkers et al., 2014) and transcription initiation-associated RNA enrichment in mouse ESCs, as measured using Start-seq (Williams et al., 2015). IRS loci are highlighted in yellow.

(F) Heatmap representation of Start-seq read density at IRS-containing genes. Data representation is similar to that in (D).

See also Figure S1.

et al., 2014) (Figures 2B, 2C, S1M, and S1N). Using published ChIP-seq data for various TFs in ESCs (Chen et al., 2008; Ho et al., 2011; Ma et al., 2011; Marson et al., 2008), we found enhanced binding of Oct4, Sox2, Nanog, Prdm14, and Stat3 (TFs generally known to bind enhancers) at IRSs, but not those that typically bind gene promoters (Klf4, Myc, Zfx, and E2f1) (Figures 2B, 2D, and S1O). Consistent with this observation, the frequency of known TF-binding motifs within IRSs revealed a significant enrichment for motifs targeted by Oct4, Sox2, Nanog, and Stat3, but not Klf4, Myc, TBP, or CTCF (Figure 2E). Additionally, published chromatin interaction maps in ESCs (Schoenfelder et al., 2015; Zhang et al., 2013) revealed 67% of IRSs interacting with one or more promoters (Figure 2F), with 48% of IRSs interacting with their host gene promoter. Altogether, these observations are consistent with previous criteria used to identify active enhancers. We used reporter experiments to confirm that many IRSs, indeed, can enhance gene expression (Figure 2G). While we cannot rule out that some IRSs could be unannotated (alternative) TSSs of coding or long non-coding RNAs, we conclude that IRSs generally represent transcriptionally active intragenic enhancers.

## Transcriptionally Active Intragenic Enhancers Attenuate Host Gene Expression

Pol II engaged in transcription is extremely stably associated with DNA. This intrinsic stability of Pol II, although obviously advantageous for efficient transcription of long genes, might actually be deleterious whenever Pol II encounters obstacles, such as another stably engaged Pol II. Indeed, Pol II transcribing one strand of a DNA fragment can interfere with a convergent Pol II transcribing the other strand of the same DNA fragment in vitro (Hobson et al., 2012; Prescott and Proudfoot, 2002;

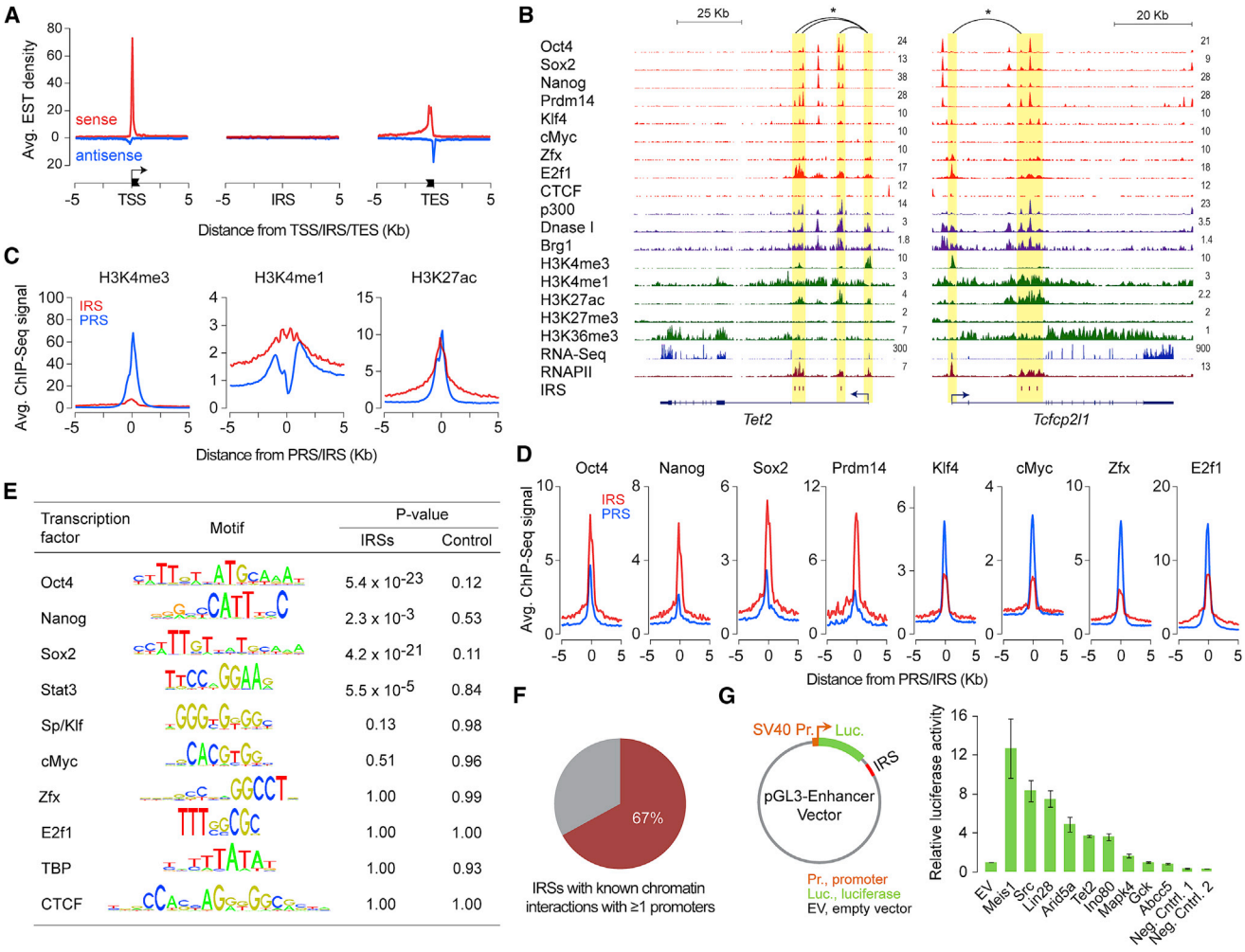



**Figure 2. Intragenic Sites of Pol II Enrichment Mark Transcriptionally Active Intragenic Enhancers**

(A) Density plot showing average expressed sequence tag (EST) enrichment at TSSs, IRSs, and transcription end sites (TESs).

(B) Genome browser shots of genes *Tet2* and *Tcfcp2l1* showing ChIP-seq read density profiles of Pol II, various transcription regulators and chromatin remodelers, and histone modifications in mouse ESCs. Also shown are read density profiles for DNase I hypersensitivity and gene expression (RNA sequencing [RNA-seq]). IRS loci are highlighted in yellow. *Known chromatin interactions.

(C) Relative levels (reads per kilobase of transcript per million mapped reads [RPKM]) of H3K4me3, H3K4me1, and H3K27ac at IRSs and PRSs in mouse ESCs.

(D) Relative binding levels (RPKM) of master ESC TFs Oct4, Sox2, Nanog, and Prdm14 and other TFs (Klf4, cMyc, Zfx, and E2f1) at IRSs and PRSs in mouse ESCs.

(E) Relative enrichment of TF sequence recognition motifs within IRSs or matched intragenic control regions in comparison to promoters.

(F) Percentage of IRSs involved in chromatin interaction with ≥1 promoter.

(G) Left: reporter construct used for testing enhancer activity. Right: relative luciferase activity of IRSs or control regions in mouse ESCs is shown. Error bars represent SEM of three biological replicates.

See also Figure S1.

Shearwin et al., 2005). Thus, it is conceivable that transcription at intragenic enhancers may interfere with and attenuate host gene transcription during productive elongation. Additionally, eRNAs from intragenic enhancers may be involved in antisense-mediated regulation of sense mRNA via the formation of transient double-stranded RNA and/or RNAi. Consistent with these scenarios, the expression of intragenic enhancer-containing genes is significantly lower than that of extragenic enhancer-associated genes (Figure 3A).

To directly test these possibilities, we used 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) (Figure S2A). Candidate IRSs (hereinafter referred to as intragenic enhancers) that conformed to many if not all the previously defined characteristic features of bona fide enhancers (bind one or more TFs; are marked by H3K4me1, H3K27ac, and DNase hypersensitivity; undergo active transcription; interact with one or more promoters; and enhance reporter expression) were cloned into the double reporter, either within the intron (pIntragenic) or downstream of the reporter gene (pExtragenic) (Figure 3B), and they were subsequently

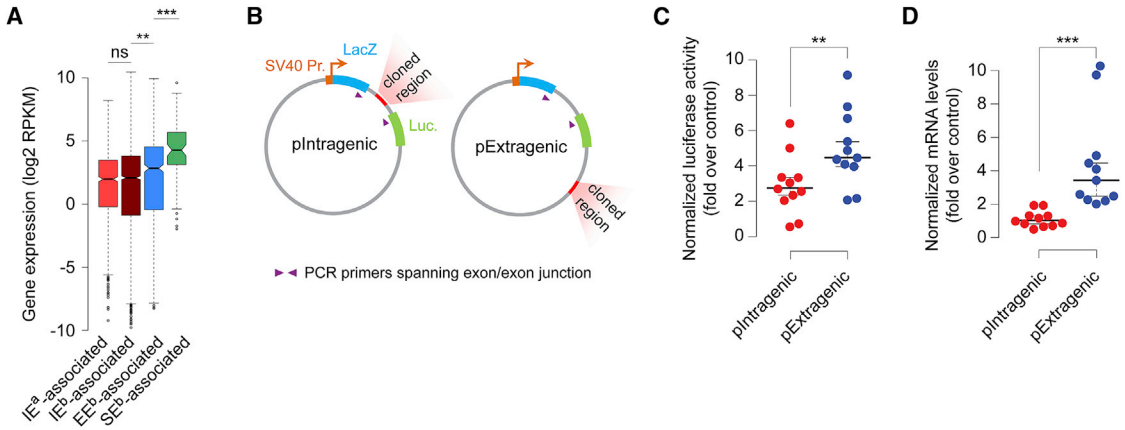



**Figure 3. Enhancers in the Intragenic Position Attenuate Host Gene Expression**

(A) Expression (Brookes et al., 2012) of genes containing intragenic enhancers (IEs; [a]defined here; [b]defined by Whyte et al., 2013) and genes closest to (≤10 kb) extragenic enhancers (EEs) (Whyte et al., 2013) and super enhancers (SEs) (Whyte et al., 2013). **p = 3.46e−5; ***p = 7.63e−8 (Wilcoxon-Mann-Whitney U test, two-sided).

(B) Reporter construct containing two reporter genes *lacZ* (encoding β-galactosidase) and *luciferase*, fused with a recombinant fragment containing two exons and a single intron, driven by SV40 promoter. Regions of interest were cloned either within the intron (pIntragenic) or downstream of the reporter gene (pExtragenic).

(C and D) Normalized luciferase activity (C) and mRNA levels of the reporter gene (D) in mouse ESCs from pIntragenic and pExtragenic reporter constructs cloned with intragenic enhancers. **p = 0.00598; ***p = $1.93 \times 10^{-7}$ (Wilcoxon-Mann-Whitney U test, two-sided). Data points represent mean of n = 5–15 biological replicates. mRNA data, normalized to *Actin*.

See also Figures S2 and S3.

transfected into mouse ESCs. While the cloned intragenic enhancers increased the reporter expression from both intronic and extragenic positions (Figure 3C), which was expected given that enhancers, regardless of their location, can activate/enhance transcription from their target gene promoters (Calo and Wysocka, 2013; Kim and Shiekhattar, 2015; Li et al., 2016; Shlyueva et al., 2014), they generated lower luciferase activity from their intragenic position than from their extragenic position (Figures 3C and S2B). To complement these studies, we asked whether transcriptionally active extragenic enhancers, when cloned as intragenic elements, would also behave similarly. Indeed, we observed a similar phenomenon for established extragenic enhancers for *Nanog* and *Prdm14* genes (Figure S2C), suggesting that the relative reduction in the reporter activity is likely due to enhancer location and not due to any intrinsic property of the enhancer.

Pol II is known to accumulate at splice junctions and has been implicated in co-transcriptional splicing (Jonkers and Lis, 2015; Mayer et al., 2015; Nojima et al., 2015). Although ~90% of intragenic enhancer peaks are at least 100 bp away from splice junctions, we asked if lower luciferase activity in pIntragenic constructs might be due to splicing defects. Because β-gal and luciferase activities were strongly correlated (Figures S2D–S2F), which is unlikely with splicing inefficiencies because luciferase expression is dependent on the removal of the intron sequence containing three translation stop codons (Figure S2A), we concluded that defective splicing is unlikely to be the cause for lower luciferase activity.

Next, to answer whether muted luciferase activity is due to reduced mRNA levels, we performed qRT-PCR analysis, using primers spanning the β-gal and luciferase exon junctions, and we observed lower mRNA levels with the enhancer in the intragenic position than in the extragenic position (Figures 3D, S3A, and S3B). Together, these data suggest that transcriptionally active enhancers in intragenic position, besides enhancing transcription of the host gene and/or other genes, also somehow attenuate host gene expression.

## The Extent of Attenuation Correlates Positively with Nascent eRNA Expression

To gain insight into mechanisms underlying the observed intragenic enhancer-mediated attenuation in gene expression, we examined nascent RNA expression. Having already ruled out splicing defects as a possible cause, we considered two plausible mechanisms: (1) transcriptional activity at intragenic enhancers interfering with host gene transcription and/or (2) antisense-mediated silencing that involves eRNAs originating from intragenic enhancers. Either way, we reasoned that the higher the intragenic enhancer RNA (ieRNA) expression, especially from the antisense strand, the higher the attenuation on host gene transcription (from the sense strand). To test this theory, we calculated the attenuation coefficient, for each intragenic enhancer, as the ratio of normalized mRNA expression of the double-reporter gene with the enhancer cloned in extragenic position (pExtragenic) compared to that with the enhancer cloned in intragenic position (pIntragenic) (Figure 3D). As posited, we observed a strong positive correlation between the nascent ieRNA (antisense) expression, measured from the native genomic locus, and the attenuation coefficient (Figures 4A and S3C), consistent with a role for intragenic enhancer transcription and/or ieRNAs in the negative regulation of host gene expression.

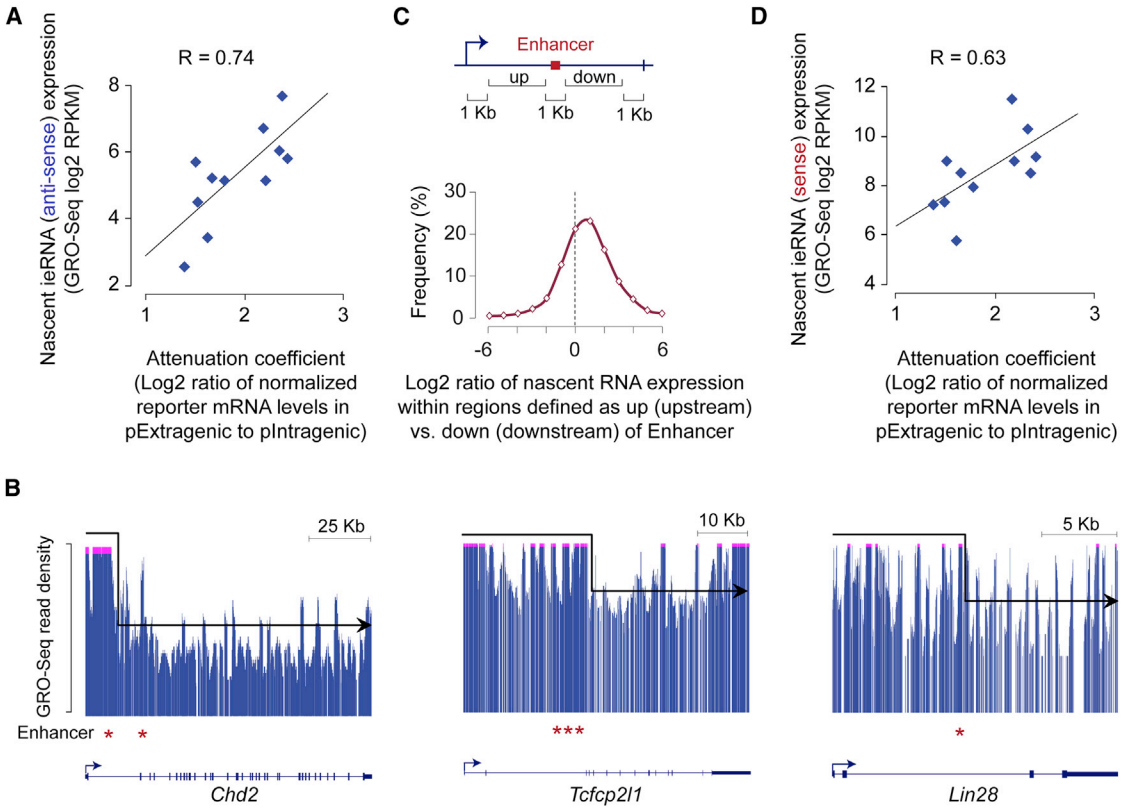



**Figure 4. Transcription at Intragenic Enhancers Attenuates Host Gene Expression**

(A) Correlation between the attenuation coefficient, calculated based on mRNA data shown in Figure S3A (summarized as Figure 3D), and levels of nascent antisense ieRNA expression, assessed from GRO-seq data in mouse ESCs (Jonkers et al., 2014).

(B) GRO-seq profiles of nascent RNA at select genes showing relatively higher levels of GRO-seq signal upstream than downstream of intragenic enhancers (*). GRO-seq signal (y axis) above an arbitrary threshold is truncated (pink streak at top) to highlight the drop-off in the GRO-seq signal downstream of the intragenic enhancer.

(C) Top: schematic showing intragenic regions defined as upstream (up) and downstream (down) of intragenic enhancer. Bottom: distribution of the ratios of GRO-seq (Jonkers et al., 2014) read density within the up region to that within the corresponding down region is shown. Only GRO-seq reads from the sense strand were used.

(D) Correlation between the attenuation coefficient, calculated based on data shown in Figure 3D, and levels of nascent sense ieRNA expression, assessed from GRO-seq data (Jonkers et al., 2014).

See also Figure S3.

## Transcription at Intragenic Enhancers Interferes with and Attenuates Host Gene Transcription

If the attenuation is due to transcriptional interference (Shearwin et al., 2005), we reasoned that at least some of the elongating Pol II transcribing the gene likely must have aborted productive transcription and released truncated RNA transcripts. Indeed, examination of nascent RNA expression across the gene length revealed that host gene expression is higher upstream compared to downstream of intragenic enhancers (Figures 4B and 4C). To ensure that the observed decreases in nascent RNA expression downstream of intragenic enhancers were not due to progressive drop-off in processivity as Pol II traverses through the length of the gene, we repeated this analysis for genes not harboring intragenic enhancers, taking gene center as the reference point. We found that the drop-off in nascent RNA expression is significantly more for genes harboring intragenic enhancers than for those that do not (p value = 1.85e−17, Wilcoxon rank-sum test), consistent with premature termination of transcription by Pol II due to transcriptional interference (Hobson et al., 2012; Prescott and Proudfoot, 2002; Shearwin et al., 2005). We also noted a positive correlation, not as strong, between levels of nascent sense transcripts from intragenic enhancers and the attenuation coefficient (Figures 4D and S3D), which is to be expected given that most functional enhancers, like promoters (Core et al., 2008; Preker et al., 2008; Seila et al., 2008), exhibit bidirectional transcription (Figure 1F) (Heinz et al., 2015; Kim and Shiekhattar, 2015; Li et al., 2016). In vitro studies have shown that rear-end collision of Pol II elongation complexes, transcribing the same DNA strand, results in backtracking (or dislodgement) of trailing (leading, respectively) Pol II (Saeki and Svejstrup, 2009). Thus, it is conceivable that Pol II transcribing intragenic enhancers on the sense strand may also interfere with Pol II transcribing the host gene on the same strand.

To gain insight into the extent of transcription interference, we examined the relationship between the host gene nascent RNA expression and the extent of attenuation (x axis in Figure 4C). Our analysis revealed that the higher the host gene nascent RNA expression upstream of the intragenic enhancer, the higher the attenuation (Figure S3E), suggesting that higher rates of transcription initiation from the host gene promoter perhaps increases the odds of interference with Pol IIs transcribing intragenic enhancers. This result, coupled with a strong positive correlation between nascent ieRNA expression and the attenuation coefficient (Figure 4A), suggests that the extent of transcription interference may range from occasional to frequent depending on many factors, including initiation frequencies at the promoter and the enhancer, elongation rates, and processivity.

## Role of Intragenic Enhancer RNAs and RNAi Machinery in Host Gene Transcription Attenuation

Next, to address whether eRNAs originating from intragenic enhancers might also attenuate host gene expression through antisense-mediated regulation, we took advantage of ESCs deficient in *Exosc3*, the catalytic subunit of the RNA exosome complex. Compared to wild-type ESCs, *Exosc3*-null ESCs exhibit elevated levels of eRNAs, long non-coding RNAs (lncRNAs), and upstream antisense RNAs, but not mRNAs (Pefanis et al., 2015). Despite the elevated levels of ieRNAs in *Exosc3*-deficient ESCs (Figures 5A and 5B), we observed no more attenuation (Figure 5C), suggesting that ieRNAs per se may not be sufficient to explain attenuation. Since RNA exosome-mediated regulation of eRNA levels could occur via two distinct mechanisms, namely, via post-transcriptional eRNA degradation or possibly through the repression of eRNA synthesis by promoting early transcription termination (Andrulis et al., 2002; Lemay et al., 2014; Pefanis et al., 2015; Preker et al., 2008), to distinguish between the two, we biochemically fractionated chromatin-associated, nucleoplasmic, and cytpoplasmic transcripts in wild-type and *Exosc3*-null ESCs. Contrary to mRNAs, which are abundant in the cytoplasmic and nucleoplasmic fractions, eRNAs and lncRNAs were abundant in the chromatin-associated and nucleoplasmic fractions (Figure 5D), consistent with their short half-lives due to rapid degradation. Most if not all of the increases in the total eRNA levels in *Exosc3*-null ESCs was due to increases in the nucleoplasmic and cytoplasmic fractions (Figures 5D and 5E), suggesting that RNA exosome-mediated regulation of ieRNAs is through post-transcriptional degradation and not through repression of ieRNA synthesis. Consequently, the lack of increased attenuation in *Exosc3*-deficient ESCs (Figures 5B and 5C), despite the increase in ieRNA level, suggests ieRNA transcripts are unlikely to be involved in attenuating their cognate host gene expression.

Next, to determine whether the observed attenuation might involve the endo-small interfering RNA (siRNA) pathway, wherein synthesis of ieRNAs yields siRNAs/microRNAs (miRNAs) that can suppress host gene expression post-transcriptionally, we analyzed published Argonaute (Ago2) cross-linking immunoprecipitation followed by deep sequencing of RNAs (CLIP-seq) data from ESCs, and we observed no enrichment for RNAs originating from intragenic enhancers (Figure 5F), suggesting that ieRNAs are not loaded into Ago2-containing RNA-induced silencing complexes. Consistent with this notion, we observed no less attenuation in Dicer-knockout ESCs, defective in RNAi, compared to wild-type ESCs (Figure 5G).

Besides their key roles in siRNA/miRNA biogenesis, RNAi factors also promote transcription termination (Skourti-Stathaki et al., 2014) and degradation of nascent transcripts (Knuckles et al., 2017). Antisense transcription, induced by R-loops, may result in the formation of transient double-stranded RNA, which in turn can attract Dicer-Argonaute machinery and promote H3K9me2 deposition and HP1$\gamma$ recruitment, effectively creating localized patches of heterochromatin (Skourti-Stathaki et al., 2014). The resulting heterochromatin prolongs and reinforces Pol II pausing at R-loop-associated regions, leading to transcription termination (Skourti-Stathaki et al., 2014). To answer whether premature termination of host gene transcription (Figures 4B and 4C) might involve such a mechanism, we analyzed published datasets, and we found no evidence of enrichment for R-loops (Chen et al., 2015) or H3K9me2 (Kurimoto et al., 2015) at intragenic enhancers (Figure S4). Nor did we find evidence of enrichment for RNAi factors Drosha and Dgcr8, known to physically associate with chromatin and promote the degradation of nascent mRNA (Knuckles et al., 2017), at intragenic enhancers (Figure 5H). Collectively, these data (Figures 4, 5, and S4) suggest that ieRNA transcripts per se are not sufficient to explain transcription interference, nor attenuation. Given that nascent ieRNA expression correlates positively with the extent of attenuation (Figures 4A and 4D), ruling out ieRNA transcripts leaves the act of transcription at intragenic enhancers itself as a plausible mechanism that can explain the attenuation of host gene expression.

## Role for Intragenic Enhancer-Mediated Attenuation in Cell Fate Determination

To determine whether the intragenic enhancer-mediated attenuation has any physiological relevance, we used the CRISPR/Cas9 system to delete candidate intragenic enhancers in ESCs. Removal of intragenic enhancer from highly expressed *Tet2* or *Ino80* genes, not surprisingly, decreased the host gene expression (Figures S5A and S5B), consistent with its known interaction with host gene promoter (Schoenfelder et al., 2015; Zhang et al., 2013). Because deletion of the enhancer eliminates its attenuation as well as activating functions, we suspected that the loss of any attenuation may have been masked by the presumably dominant enhancer activation function, resulting in net loss in gene expression. In an attempt to uncouple the enhancer and attenuator functions, we focused on candidate intragenic enhancer-containing genes with low-to-moderate expression in ESCs (Figure S1B), with the reasoning that the enhancer's dominant role at these genes is presumably as an attenuator. As expected, removal of the intragenic enhancer from *Meis1* or *Mapk4* resulted in derepression (~2- to 3-fold) of respective host genes (Figures 6A and S5C), but not of other nearby genes for which the deleted sequence may be an extragenic enhancer (Figures 6D, 6E, and S5D–S5F).

Notably, ESCs lacking the *Meis1* intragenic enhancer exhibited severe morphological defects, loss of alkaline phosphatase activity, decreased expression of pluripotency-associated genes, and increased expression of several differentiation-associated genes (Figures 6B and 6C)—all consistent with ESC

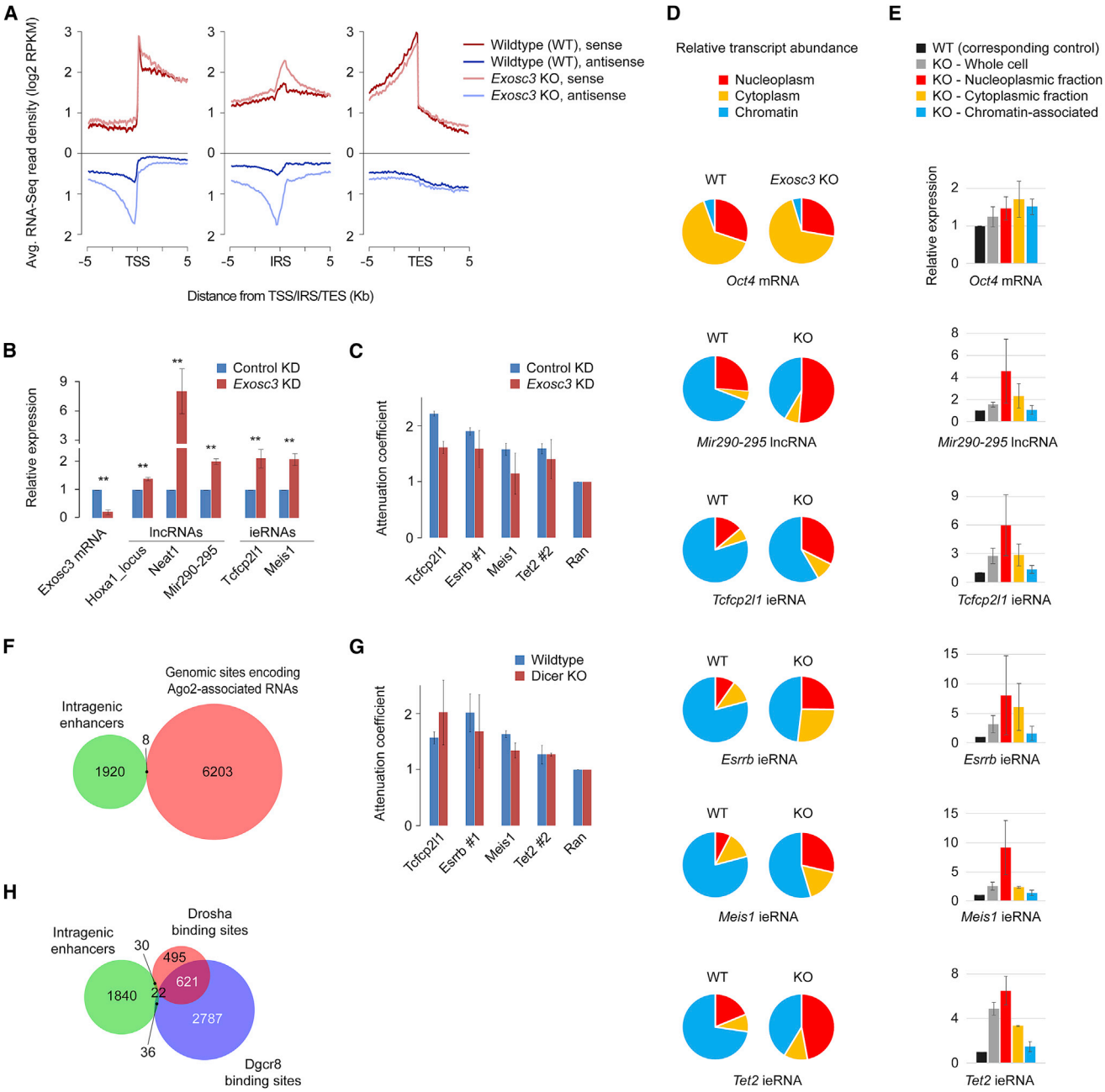



**Figure 5. The Role of eRNAs from Intragenic Enhancers in Host Gene Transcription Attenuation**

(A) Read density plot showing average expression (RPKM) profile in wild-type (WT) and *Exosc3*-knockout (KO) mouse ESCs, as measured using RNA-seq (Pefanis et al., 2015), near transcription start site (TSS), intragenic enhancer (IE), and transcription end site (TES).

(B) qRT-PCR analysis of RNA expression in wild-type and *Exosc3*-knockdown (KD) mouse ESCs. Data are normalized to *Actin*. Error bars represent SEM of three biological replicates. **p < 0.01 (Student's t test, two-sided).

(C) Normalized luciferase activity from pIntragenic and pExtragenic reporter constructs, cloned with intragenic enhancers, in control KD and *Exosc3*-KD mouse ESCs. Data are normalized to control sequence from Ran locus. Error bars represent SEM of three biological replicates.

(D) Proportion of transcripts within chromatin-associated, nucleoplasmic, and cytoplasmic fractions of WT (left) or *Exosc3*-KO (right) mouse ESCs. Data, normalized to *Actin* expression in the whole cell, represent mean of three biological replicates.

(E) RNA expression in chromatin-associated, nucleoplasmic, or cytoplasmic fractions of WT or *Exosc3*-KO mouse ESCs. Expression in WT cells (whole cell or various fractions) is set to 1 (black bar). Expression in KO cells (whole cell or various fractions) was calculated relative to that in WT cells (corresponding whole cell or respective fractions). Data are normalized to *Actin* expression in the whole cell. Error bars represent SEM of three biological replicates.

(F) Overlap between genomic sites encoding Ago2-associated RNAs, as determined using CLIP-seq in mouse ESCs (Leung et al., 2011), and intragenic enhancers.

*(legend continued on next page)*

differentiation. Silencing *Meis1* in ESCs lacking the *Meis1* intragenic enhancer, at least partially, restored the differentiation phenotype (Figures S5G and 6F), suggesting that *Meis1* derepression is the cause for the observed phenotype. Indeed, *Meis1* overexpression in ESCs recapitulates the ESC differentiation phenotype observed in ESCs lacking the *Meis1* intragenic enhancer (Figures S5H–S5J). Furthermore, subjecting ESCs lacking the *Meis1* enhancer to undergo embryoid body differentiation revealed their propensity to preferentially differentiate toward mesoderm lineage, as evidenced by a dramatic increase in the levels of mesoderm specification genes (*Brachyury*, *Eomes*, and *Gsc*), presumably at the expense of endoderm and ectoderm lineage markers (Figures 6G and S5K). Our findings, consistent with *Meis1*'s mesoderm-centric essential roles in definitive hematopoiesis and normal cardiac development (Azcoitia et al., 2005; Mahmoud et al., 2013), highlight, at least in the case of *Meis1*, a physiological role for intragenic enhancer-mediated attenuation in determining cell fate choice during early embryonic development.

### Enhancer Strength and Attenuation

To gain insight into why intragenic enhancers within low-to-moderately expressed *Meis1* and *Mapk4* exhibit strong attenuator activity whereas those within highly expressed *Tet2* and *Ino80* do not, we grouped intragenic enhancers into four bins based on their host gene expression, and we found that enhancers within lowly expressed genes have significantly lower TF occupancy and enhancer-associated marks (Figures S6A–S6D), suggesting that they may be relatively weak enhancers with presumably limited activation potential. Consistent with this observation, a smaller percentage of enhancers within lowly expressed genes are involved in enhancer-promoter interactions (Figure S6E); even those that interact make fewer and weaker contacts (Figures S6E and S6F), with a vast majority interacting with host gene promoters. Derepression of *Meis1* and *Mapk4* upon respective intragenic enhancer deletions suggests that any gains in host gene activation from a weak intragenic enhancer are insufficient to overcome losses from enhancer transcription interfering with the host gene transcription; the enhancer, in effect, serves the role of a repressor. In contrast, robust host gene activation by strong intragenic enhancers, as could have been the case for *Tet2* and *Ino80* (their intragenic enhancers designated as part of a super-enhancer), may overwhelm and mask any attenuation by enhancer transcription, but not without enhancer transcription attenuating some of the host gene transcription. In such cases, the net effect would still be gene activation, with enhancer deletion resulting in a net loss in host gene expression.

### Functional Characterization of Intragenic Enhancer-Containing Genes

Characterization of genes containing transcriptionally active intragenic enhancers in ESCs revealed an enrichment for ESC identity genes, including those associated with ESC self-renewal, differentiation, and early embryonic development (Figure S7A), suggesting a potential role for intragenic enhancers in fine-tuning the transcription of key cell identity genes. To explore this further, we used data from murine bone marrow-derived macrophages (Ostuni et al., 2013) to annotate putative intragenic enhancers (Table S2). As in ESCs, macrophage enhancers are enriched for H3K4me1, H3K27ac, and sequence motifs bound by macrophage-specific master TFs (PU.1, JunB, and Stat1), but not ubiquitously expressed TFs (Myc, Zfx, and CTCF) (Figures S7B and S7C). Comparison of intragenic enhancer-associated genes in ESCs and macrophages revealed a minimal overlap (Figures S7D and S7E). We extended our analyses to other mammalian cells (Calabrese et al., 2012; Ghamari et al., 2013; Tippmann et al., 2012), and we found that intragenic enhancers are largely cell type specific (Figure S7F; Tables S3, S4, and S5), consistent with gene ontology analysis revealing an enrichment for genes associated with cell-type-specific biological processes (Figure S7G).

## DISCUSSION

Enhancers are thought to activate or increase transcription from the target gene promoter by facilitating the recruitment of co-activators and other accessory factors to the promoter to potentiate transcription initiation and/or sustain transcription (Calo and Wysocka, 2013; Heinz et al., 2015; Kim and Shiekhattar, 2015; Li et al., 2016; Shlyueva et al., 2014). Recent genome-wide studies have established that enhancers themselves recruit Pol II and undergo Pol II-mediated transcription to generate short non-coding eRNAs. Several lines of evidence support a functional role for eRNAs in perhaps all stages of gene activation, from facilitating chromatin accessibility, enhancer-promoter loop formation, Pol II recruitment, and promoter-proximal Pol II pause release to modulating TF-DNA binding (Kim and Shiekhattar, 2015; Li et al., 2016; Schaukowitch et al., 2014). However, whether the act of enhancer transcription, rather than the eRNA transcript itself, has any functional role in regulating target gene expression remains enigmatic. Here we report an unanticipated role for transcriptionally active intragenic enhancers in attenuating host gene expression. Our studies suggest that the act of transcription at intragenic enhancers interferes with and attenuates host gene transcription. Evidence from our analyses of nascent RNA supports premature termination of host gene transcription at or near the intragenic enhancer (Figures 4B and 4C), consistent with enhancer transcription (or other enhancer property) interfering with host gene transcription during productive elongation.

A recent study reported that convergent antisense transcription (CAT), occurring immediately downstream of gene TSSs (~150 ± 50 bp), is a characteristic feature of lower-expressed genes (with Pol II density in the gene body used as a proxy for

(G) Normalized luciferase activity from pIntragenic and pExtragenic reporter constructs (Figure 3B), cloned with intragenic enhancers, in wild-type and *Dicer1*-deficient mouse ESCs. Data are normalized to control sequence from Ran locus. Error bars represent SEM of three biological replicates.
(H) Overlap between intragenic enhancers and Drosha- and Dgcr8-binding sites in mouse ESCs (Knuckles et al., 2017).
See also Figure S4.

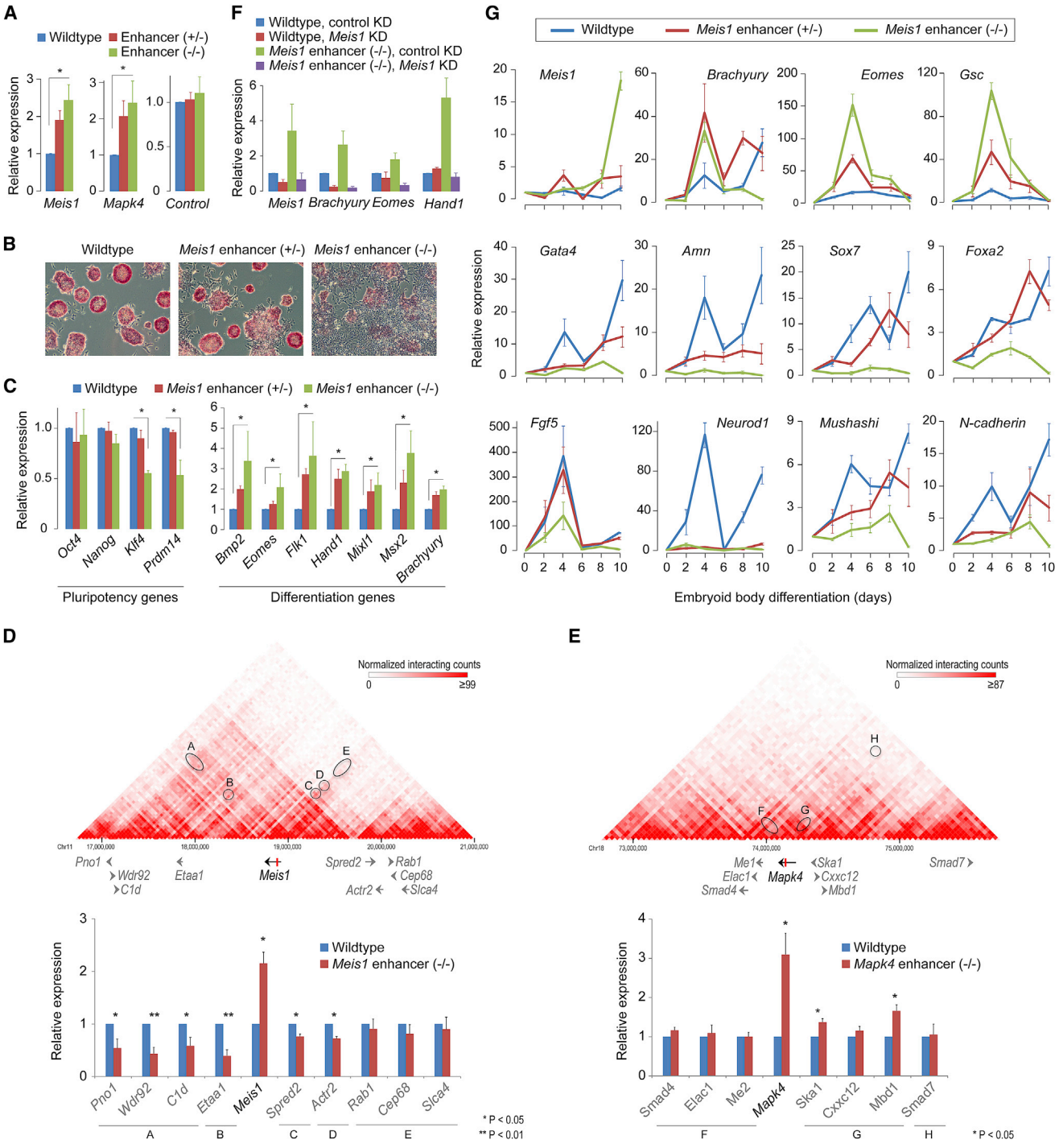



**Figure 6. A Role for Intragenic Enhancer-Mediated Attenuation in Cell Fate Determination**

(A) qRT-PCR analysis of mRNA levels of intragenic enhancer-containing genes *Meis1* and *Mapk4* in corresponding wild-type, enhancer heterozygous (+/−), and enhancer knockout (−/−) mouse ESCs. As a control, an intronic region (showing no signs of an enhancer) from *Cdk6* gene was deleted and *Cdk6* expression was measured. Data are normalized to *Actin*. Error bars represent SEM of three biological replicates. *p < 0.01 (Student's t test, two-sided).

(B) Morphology and alkaline phosphatase staining of wild-type, *Meis1* enhancer (+/−), and *Meis1* enhancer (−/−) ESCs.

(C) mRNA levels of pluripotency-associated ESC identity genes (left) and differentiation/developmental genes (right) in wild-type, *Meis1* enhancer (+/−), and *Meis1* enhancer (−/−) ESCs. Data are normalized to *Actin*. Error bars represent SEM of three biological replicates. *p < 0.05 (Student's t test, two-sided).

(D and E) Top: chromosomal view of Hi-C interaction maps of Meis1 (D) and Mapk4 (E) loci in mouse ESCs (Dixon et al., 2012). Hi-C maps capture all genome-wide interactions, functional as well as structural. Black/gray arrows represent gene structures and the red vertical bar within Meis1 and Mapk4 genes represents

*(legend continued on next page)*

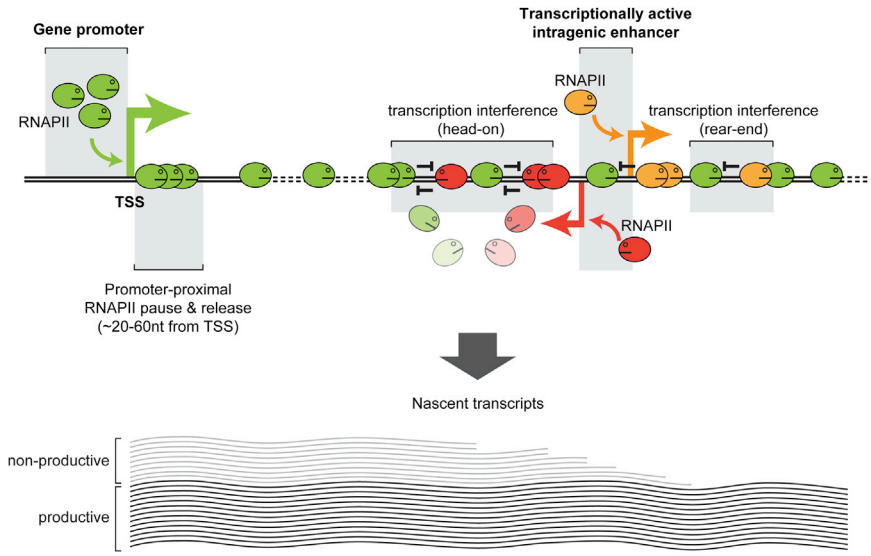



**Figure 7. Model for Transcription at Intragenic Enhancers Interfering with and Attenuating Host Gene Transcription during Productive Elongation**

Pol II transcribing intragenic enhancers (shown as red and orange oval faces) interferes with and attenuates the host gene transcription (Pol II shown as green oval faces) during productive elongation, with the extent of interference ranging from occasional to frequent depending on many factors, including initiation frequencies at the promoter and the enhancer, elongation rates, and processivity. Transcription interference may involve one or more events, including excessive torsional stress due to positive DNA supercoiling (Chong et al., 2014; Ma et al., 2013), Pol II stalling/collision, modifications to the chromatin structure/architecture, and antisense-mediated regulation, all eventually leading to premature termination of host gene transcription and non-productive transcripts. The net effect of intragenic enhancer-mediated regulation of host gene expression may range from repression to fine-tuning or activation depending on the difference between gains from enhancer-mediated host gene activation and losses due to enhancer-mediated interference and attenuation of host gene transcription.

gene expression), and the authors surmised that it may be contributing to the genes' lower expression, likely via interference with transcription initiation and/or release of Pol II from promoter-proximal pausing (Mayer et al., 2015). However, another study investigating CAT within 2 kb downstream of TSSs reported no correlation between levels of CAT and gene expression (Lavender et al., 2016), and those authors concluded that CAT is not inhibitory. Besides the differences in the cell types, CAT definition, data types, and thresholds used for analyses, the discrepancy between these two studies is explainable by the apparent requirement for CAT to traverse past the gene TSS for it to have a negative impact on gene expression (Chen et al., 2016). Given that the intragenic enhancer transcription we studied is located thousands, if not tens of thousands, of base pairs downstream of the host gene promoter (~98% and ~90% of the enhancers are at least 2 and 5 kb from TSSs, respectively) and that enhancer transcription (on the antisense strand) does not seem to extend very far (Figure 1D), it is highly unlikely that enhancer transcription directly interferes with transcription initiation and/or Pol II pause release at the host gene promoter. It is conceivable, however, that transcription at intragenic enhancers disrupts enhancer-promoter interaction and, thus, transcription initiation and/or Pol II pause release at the host gene promoter; but, such a mechanism would have to be exclusive to (or, at least, have a bias toward) intragenic enhancers, because the same enhancer has higher activation potential from an extragenic as opposed to an intragenic position (Figure 3). Thus, we infer that any interference by intragenic enhancer transcription (or other enhancer property) on host gene transcription occurs likely during productive elongation, except when the enhancer is near the host gene TSS (within ~200 bp), in which case direct interference with transcription initiation and/or Pol II pause release may also be in play.

While we could not find evidence supporting a role for ieRNAs in host gene transcription attenuation, we cannot rule out a role for ieRNAs in antisense-mediated regulation of sense RNA through RNAi or other mechanisms (Faghihi and Wahlestedt, 2009; Pelechano and Steinmetz, 2013; Skourti-Stathaki et al., 2014). While ieRNAs themselves may have additional regulatory functions, our findings suggest that the act of transcription at intragenic enhancers, by itself, can negatively regulate host gene transcription during productive elongation through transcription interference (Figure 7). Although our conclusions are based on candidate intragenic enhancers we investigated, the attenuation function, in theory, can exist for all actively transcribed intragenic enhancers. Further studies are required to investigate, on a genome-wide scale, the net impact of intragenic enhancer-mediated attenuation on host gene expression, which may entail uncoupling the enhancers' activation and attenuation functions. Nevertheless, our findings suggest that

intragenic enhancers. Ovals highlight interaction strengths between the regions containing intragenic enhancer and gene promoter(s). Bottom: mRNA levels of genes within *Meis1* (or *Mapk4*) locus in wild-type or *Meis1* (*Mapk4*, respectively) enhancer (−/−) ESCs are shown. Data are normalized to *Actin*. Error bars represent SEM of three biological replicates.

(F) Depletion of Meis1 in *Meis1* enhancer (−/−) ESCs restores the expression of differentiation genes. Data are normalized to *Actin*. Error bars represent SEM of three biological replicates. *p < 0.05 (Student's t test, two-sided).

(G) mRNA levels of early differentiation markers (top row, mesoderm; middle row, endoderm; bottom row, ectoderm) during embryoid body formation of wild-type, *Meis1* enhancer (+/−), and *Meis1* enhancer (−/−) ESCs. Data are normalized to *Actin*. Error bars represent SEM of three biological replicates.

See also Figures S5 and S6.

intragenic enhancers serve more regulatory functions than previously appreciated and that intragenic enhancer-mediated transcription attenuation could represent a general mechanism to attenuate and fine-tune host gene transcription, much like a transcriptional rheostat, during productive elongation.

While the attenuating function of intragenic enhancer transcription is perhaps an unintended by-product or unavoidable consequence of enhancer activation, given that enhancers can activate transcription of their target gene promoters independently of their position or location relative to their targets (Calo and Wysocka, 2013; Kim and Shiekhattar, 2015; Li et al., 2016; Shlyueva et al., 2014), evolution could have selected for enhancers to exist outside of protein-coding genes and still have them perform the same function. Our *Meis1* studies, which reveal an essential role for intragenic enhancer-mediated attenuation in maintaining ESCs in their pluripotent state, highlight that intragenic sites may have been selected to function as enhancers so that not only can they activate transcription of one or more genes from a distance but also fine-tune transcription of their host gene through transcription interference. Given that the precise level of master TF Oct4 alone can determine three distinct fates of ESCs (Niwa et al., 2000), it is conceivable that refinement of expression of key cell identity genes through intragenic enhancer-mediated attenuation may prevent abrupt changes in their expression levels, which could influence cell fate decisions.

We propose a highly generalizable model wherein transcription at enhancers, weak and strong, interferes with and attenuates host gene expression. The net effect of intragenic enhancer-mediated regulation may range from repression (as in the case of *Meis1* and *Mapk4*) to fine-tuning or activation (as may be the case for *Tet2* and *Ino80*), depending on the difference between gains from enhancer-mediated host gene activation and losses due to enhancer-mediated interference and attenuation of host gene transcription. Under this model, the attenuation function of transcriptionally active intragenic enhancers within highly expressed genes would provide for a fine-tuned enhancement of host gene expression by preventing excessive transcription activation. Such an intrinsic negative feedback mechanism would keep the enhancer and the promoter from activating each other in an exaggerated manner. Intragenic enhancer-mediated attenuation, in addition to providing a potential mechanism to fine-tune transcription of key cell identity genes, may also inhibit expression of genes with inherently leaky transcription while keeping them poised for activation. An obvious implication of our finding is that antisense transcription upstream of TSSs (Core et al., 2008; Preker et al., 2008; Seila et al., 2008) may play an active role in attenuating readthrough sense transcription, originating from enhancers/genes located immediately upstream, via transcription interference, as previously shown (Nguyen et al., 2014). Such a mechanism would explain why genes with divergent transcription, as a group, have higher expression than those without it (Lavender et al., 2016).

In summary, our studies suggest that transcriptionally active intragenic enhancers not only enhance transcription of one or more genes from a distance but also fine-tune transcription of the host gene through transcription interference, facilitating differential utilization of the same regulatory element for disparate functions.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mouse ES Cell Culture
- METHOD DETAILS
  - Transient Transfection
  - Double-Reporter Plasmids and Assay
  - CRISPR/Cas9-Mediated Targeted Deletion
  - Quantitative RT-PCR
  - Alkaline Phosphatase (AP) Staining
  - Embryoid Body Differentiation
  - Western Blot
  - Identification of Intragenic Sites of Pol II Enrichment
  - RNA-Seq, GRO-Seq, Start-Seq, and DRIP-RNA-Seq Data Analysis
  - ChIP-Seq Data Analysis
  - Chromatin Interaction Map Analysis
  - Ago2 CLIP-Seq Data Analysis
  - Expressed Sequence Tag (EST) Analysis
  - Sequence Analysis
  - Functional Enrichment Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and six tables can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2017.09.010.

## AUTHOR CONTRIBUTIONS

R.J. conceived the study. S.C., P.Y., and R.J. designed the study. S.C. performed all the experiments. P.Y. performed all the bioinformatics analyses. J.P.K. generated CRISPR clones. D.K. contributed to bioinformatics analyses. A.E.C. contributed to cloning. S.C., P.Y., J.P.K., D.K., A.E.C., A.J.O., K.A., and R.J. analyzed the data. R.J. wrote the manuscript with input from all authors.

## ACKNOWLEDGMENTS

We thank T.K. Archer, T. Henriques, L. Ho, G. Hu, D. Fargo, T.A. Kunkel, S. Peddada, and P.A. Wade for useful discussions and comments on the manuscript. This work was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (1ZIAES102625 to R.J.).

Received: February 24, 2017
Revised: July 10, 2017
Accepted: September 6, 2017
Published: October 5, 2017

## REFERENCES

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Andrulis, E.D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., and Lis, J.T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. Nature *420*, 837–841.

Azcoitia, V., Aracil, M., Martínez-A, C., and Torres, M. (2005). The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. Dev. Biol. *280*, 307–320.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Bintu, L., Ishibashi, T., Dangkulwanich, M., Wu, Y.Y., Lubkowska, L., Kashlev, M., and Bustamante, C. (2012). Nucleosomal elements that control the topography of the barrier to transcription. Cell *151*, 738–749.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157–170.

Calabrese, J.M., Sun, W., Song, L., Mugford, J.W., Williams, L., Yee, D., Starmer, J., Mieczkowski, P., Crawford, G.E., and Magnuson, T. (2012). Site-specific silencing of regulatory elements as a mechanism of X inactivation. Cell *151*, 951–963.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazzio, T.G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nat. Struct. Mol. Biol. *22*, 999–1007.

Chen, Y., Pai, A.A., Herudek, J., Lubas, M., Meola, N., Järvelin, A.I., Andersson, R., Pelechano, V., Steinmetz, L.M., Jensen, T.H., and Sandelin, A. (2016). Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. Nat. Genet. *48*, 984–994.

Cheung, A.C., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249–253.

Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014). Mechanism of transcriptional bursting in bacteria. Cell *158*, 314–326.

Cinghu, S., Yellaboina, S., Freudenberg, J.M., Ghosh, S., Zheng, X., Oldfield, A.J., Lackford, B.L., Zaykin, D.V., Hu, G., and Jothi, R. (2014). Integrative framework for identification of key cell identity genes uncovers determinants of ES cell identity and homeostasis. Proc. Natl. Acad. Sci. USA *111*, E1581–E1590.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L., Lis, J.T., Siepel, A., and Kraus, W.L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol. Cell *50*, 212–222.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. *8*, e1000384.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Faghihi, M.A., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol. *10*, 637–643.

Ghamari, A., van de Corput, M.P., Thongjuea, S., van Cappellen, W.A., van Ijcken, W., van Haren, J., Soler, E., Eick, D., Lenhard, B., and Grosveld, F.G. (2013). In vivo live imaging of RNA polymerase II transcription factories in primary cells. Genes Dev. *27*, 767–777.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc. Natl. Acad. Sci. USA *106*, 5187–5191.

Ho, L., Miller, E.L., Ronan, J.L., Ho, W.Q., Jothi, R., and Crabtree, G.R. (2011). esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. Nat. Cell Biol. *13*, 903–913.

Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. Mol. Cell *48*, 365–374.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 167–177.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife *3*, e02407.

Kim, T.K., and Shiekhattar, R. (2015). Architectural and functional commonalities between enhancers and promoters. Cell *162*, 948–959.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kireeva, M.L., Hancock, B., Cremona, G.H., Walter, W., Studitsky, V.M., and Kashlev, M. (2005). Nature of the nucleosomal barrier to RNA polymerase II. Mol. Cell *18*, 97–108.

Knuckles, P., Carl, S.H., Musheev, M., Niehrs, C., Wenger, A., and Bühler, M. (2017). RNA fate determination through cotranscriptional adenosine methylation and microprocessor binding. Nat. Struct. Mol. Biol. *24*, 561–569.

Kowalczyk, M.S., Hughes, J.R., Garrick, D., Lynch, M.D., Sharpe, J.A., Sloane-Stanley, J.A., McGowan, S.J., De Gobbi, M., Hosseini, M., Vernimmen, D., et al. (2012). Intragenic enhancers act as alternative promoters. Mol. Cell *45*, 447–458.

Kurimoto, K., Yabuta, Y., Hayashi, K., Ohta, H., Kiyonari, H., Mitani, T., Moritoki, Y., Kohri, K., Kimura, H., Yamamoto, T., et al. (2015). Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells. Cell Stem Cell *16*, 517–532.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lavender, C.A., Cannady, K.R., Hoffman, J.A., Trotter, K.W., Gilchrist, D.A., Bennett, B.D., Burkholder, A.B., Burd, C.J., Fargo, D.C., and Archer, T.K. (2016). Downstream antisense transcription predicts genomic features that define the specific chromatin environment at mammalian promoters. PLoS Genet. *12*, e1006224.

Lemay, J.F., Larochelle, M., Marguerat, S., Atkinson, S., Bähler, J., and Bachand, F. (2014). The RNA exosome promotes transcription termination of backtracked RNA polymerase II. Nat. Struct. Mol. Biol. *21*, 919–926.

Leung, A.K., Young, A.G., Bhutkar, A., Zheng, G.X., Bosson, A.D., Nielsen, C.B., and Sharp, P.A. (2011). Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. Nat. Struct. Mol. Biol. *18*, 237–244.

Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. Cell *145*, 502–511.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat. Rev. Genet. *17*, 207–223.

Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A., and Wysocka, J. (2011). Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. Nat. Struct. Mol. Biol. *18*, 120–127.

Ma, J., Bai, L., and Wang, M.D. (2013). Transcription under torsion. Science *340*, 1580–1583.

Mahmoud, A.I., Kocabas, F., Muralidhar, S.A., Kimura, W., Koura, A.S., Thet, S., Porrello, E.R., and Sadek, H.A. (2013). Meis1 regulates postnatal cardiomyocyte cell cycle arrest. Nature *497*, 249–253.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Martin, R.M., Rino, J., Carvalho, C., Kirchhausen, T., and Carmo-Fonseca, M. (2013). Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. Cell Rep. *4*, 1144–1155.

Mayer, A., Iulio, J.D., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell *161*, 541–554.

Nasim, M.T., Chowdhury, H.M., and Eperon, I.C. (2002). A double reporter assay for detecting changes in the ratio of spliced and unspliced mRNA in mammalian cells. Nucleic Acids Res. *30*, e109.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annu. Rev. Genet. *46*, 1–19.

Nguyen, T., Fischl, H., Howe, F.S., Woloszczuk, R., Serra Barros, A., Xu, Z., Brown, D., Murray, S.C., Haenni, S., Halstead, J.M., et al. (2014). Transcription mediated insulation and interference direct gene cluster expression switches. eLife *3*, e03635.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372–376.

Nojima, T., Gomes, T., Grosso, A.R., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell *161*, 526–540.

Oldfield, A.J., Yang, P., Conway, A.E., Cinghu, S., Freudenberg, J.M., Yellaboina, S., and Jothi, R. (2014). Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. Mol. Cell *55*, 708–722.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*, 157–171.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P., et al. (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell *161*, 774–789.

Pelechano, V., and Steinmetz, L.M. (2013). Gene regulation by antisense transcription. Nat. Rev. Genet. *14*, 880–893.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat. Rev. Mol. Cell Biol. *16*, 190–202.

Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W., and Sandelin, A. (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res. *38*, D105–D110.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851–1854.

Prescott, E.M., and Proudfoot, N.J. (2002). Transcriptional collision between convergent genes in budding yeast. Proc. Natl. Acad. Sci. USA *99*, 8796–8801.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Qian, Z., Cai, Y.D., and Li, Y. (2006). Automatic transcription factor classifier based on functional domain composition. Biochem. Biophys. Res. Commun. *347*, 141–144.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Saeki, H., and Svejstrup, J.Q. (2009). Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. Mol. Cell *35*, 191–205.

Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. Mol. Cell *56*, 29–42.

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., et al. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. *25*, 582–597.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference–a crash course. Trends Genet. *21*, 339–345.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436–439.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Teves, S.S., Weber, C.M., and Henikoff, S. (2014). Transcribing through the nucleosome. Trends Biochem. Sci. *39*, 577–586.

Thomas-Chollier, M., Hufton, A., Heinig, M., O'Keeffe, S., Masri, N.E., Roider, H.G., Manke, T., and Vingron, M. (2011). Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nat. Protoc. *6*, 1860–1869.

Tippmann, S.C., Ivanek, R., Gaidatzis, D., Schöler, A., Hoerner, L., van Nimwegen, E., Stadler, P.F., Stadler, M.B., and Schübeler, D. (2012). Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels. Mol. Syst. Biol. *8*, 593.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., and Adelman, K. (2015). Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol. Cell *58*, 311–322.

Zhang, Y., Wong, C.H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., et al. (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature *504*, 306–310.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Anti-Meis1, rabbit polyclonal | Abcam | Ab124686; RRID: AB_10973740 |
| Anti-Ran, mouse monoclonal | BD Biosciences | 610341; RRID: AB_397731 |
| Chemicals, Peptides, and Recombinant Proteins | | |
| 2-Mercaptoethanol | Sigma | M3148 |
| ApaI Enzyme | NEB | R0114S |
| DMEM | Thermo Fisher Scientific | 11965-084 |
| EmbryoMax nucleosides | Millipore | ES-008-D |
| EmeraldAmp GT PCR Master Mix | Takara | RR310A |
| ESGRO Complete PLUS Clonal Grade Medium | Millipore | SF001-500 |
| Fetal Bovine Serum (FBS) | Gemini | 100-125 |
| Gelatin | Sigma | G1890 |
| Inactivated MEFs | GIBCO | A24903 |
| LIF | Millipore | ESG1107 |
| Lipofectamine 2000 | Invitrogen | 1168019 |
| Meis1 FlexiTube siRNA | QIAGEN | SI01303610 |
| NaCl, 5M | Sigma | S5150 |
| Negative Control siRNA | QIAGEN | 1027310 |
| Non-Essential Amino Acids Solution | Thermo Fisher Scientific | 11140050 |
| Protease Inhibitors | Roche | 4693159001 |
| QIAzol Lysis Reagent | QIAGEN | 79306 |
| SalI Enzyme | NEB | R3138S |
| Sodium Deoxycholate | Sigma | 30970 |
| SsoFast EvaGreen supermix | Bio-Rad | 1725201 |
| Tris HCL, pH 8.0 | Sigma | T2663 |
| Critical Commercial Assays | | |
| Alkaline Phosphatase Detection Kit | Stemgent | 00-0055 |
| Dual-Light Luciferase and β-Galactosidase Reporter Gene Assay System | Thermo Fisher Scientific | T1003 |
| iScript cDNA Synthesis Kit | Bio-Rad | 1708891 |
| QIAquick Gel Extraction Kit | QIAGEN | 28706 |
| Rneasy Mini Kit | QIAGEN | 74104 |
| TOPO TA Cloning Kit | Invitrogen | 450641 |
| Experimental Models: Cell Lines | | |
| Mouse ESCs (E14Tg2a) | ATCC | CRL-1821 |
| Mouse ESCs, Dicer knockout (KO) | Novus Biologicals | NBP1-96751 |
| Mouse ESCs, Exosc3 knockout (KO) | Prof. Uttiya Basu | N/A |
| Mouse ESCs, Meis1 intragenic enhancer (+/−) | This paper | N/A |
| Mouse ESCs, Meis1 intragenic enhancer (−/−) | This paper | N/A |
| Mouse ESCs, Mapk4 intragenic enhancer (+/−) | This paper | N/A |
| Mouse ESCs, Mapk4 intragenic enhancer (−/−) | This paper | N/A |
| Mouse ESCs, Cdk6 intronic control region (+/−) | This paper | N/A |
| Mouse ESCs, Cdk6 intronic control region (−/−) | This paper | N/A |
| Mouse ESCs, Tet2 intragenic enhancer (−/−) | This paper | N/A |
| Mouse ESCs, Ino80 intragenic enhancer (−/−) | This paper | N/A |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Oligonucleotides | | |
| Primers used for PCR amplification of enhancer and control sequences (used in the double-reporter plasmid assay) | This paper | Table S6 |
| Guide RNA sequences used for CRISPR/Cas9-mediated targeted deletions | This paper | Table S6 |
| Screening primers used to test CRISPR/Cas9-mediated targeted deletions | This paper | Table S6 |
| Gene-specific primers used for RT-qPCR analysis | This paper | Table S6 |
| Recombinant DNA | | |
| Plasmid: maxGFP | Lonza | VDF-1012 |
| Plasmid: pExtragenic | This paper | N/A |
| Plasmid: pGL3-Basic | Promega | E1751 |
| Plasmid: pIntragenic | This paper | N/A |
| Plasmid: pRL-TK | Promega | E2231 |
| Plasmid: pTN23 | Prof. I.C. Eperon | N/A |
| Plasmid: pX330-U6-Chimeric_BB-CBh-hSpCas9 | Addgene | 42230 |
| Other | | |
| Mouse ESCs, Pol II ChIP-seq | Brookes et al., 2012 | GEO: GSE34520 |
| Mouse ESCs, Pol II ChIP-seq | Rahl et al., 2010 | GEO: GSE20485 |
| Mouse ESCs, Pol II ChIP-seq | Shen et al., 2012 | GEO: GSE29184 |
| Mouse ESCs, Pol II ChIP-seq | Tippmann et al., 2012 | GEO: GSE33252 |
| Mouse ESCs, Pol II-S2P ChIP-seq | Rahl et al., 2010 | GEO: GSE20485 |
| Mouse ESCs, Pol II-S5P ChIP-seq | Rahl et al., 2010 | GEO: GSE20485 |
| Mouse ESCs, H3K4me1 ChIP-seq | Creyghton et al., 2010 | GEO: GSE24164 |
| Mouse ESCs, H3K4me3 ChIP-seq | ENCODE | GEO: GSE31039 |
| Mouse ESCs, H3K9me2 ChIP-seq | Kurimoto et al., 2015 | GEO: GSE60204 |
| Mouse ESCs, H3K27ac ChIP-seq | ENCODE | GEO: GSE31039 |
| Mouse ESCs, H3K27me3 ChIP-seq | Ho et al., 2011 | GEO: GSE27708 |
| Mouse ESCs, H3K36me3 ChIP-seq | ENCODE | GEO: GSE31039 |
| Mouse ESCs, Brg1 ChIP-seq | Ho et al., 2009 | GEO: GSE14344 |
| Mouse ESCs, cMyc ChIP-seq | Chen et al., 2008 | GEO: GSE11431 |
| Mouse ESCs, Dgcr8 ChIP-seq | Knuckles et al., 2017 | GEO: GSE92257 |
| Mouse ESCs, Drosha ChIP-seq | Knuckles et al., 2017 | GEO: GSE92257 |
| Mouse ESCs, E2f1 ChIP-seq | Chen et al., 2008 | GEO: GSE11431 |
| Mouse ESCs, Klf4 ChIP-seq | Chen et al., 2008 | GEO: GSE11431 |
| Mouse ESCs, Nanog ChIP-seq | Marson et al., 2008 | GEO: GSE11724 |
| Mouse ESCs, nMyc ChIP-seq | Chen et al., 2008 | GEO: GSE11431 |
| Mouse ESCs, Oct4 ChIP-seq | Marson et al., 2008 | GEO: GSE11724 |
| Mouse ESCs, p300 ChIP-seq | Creyghton et al., 2010 | GEO: GSE24164 |
| Mouse ESCs, Prdm14 ChIP-seq | Ma et al., 2011 | GEO: GSE42616 |
| Mouse ESCs, Sox2 ChIP-seq | Marson et al., 2008 | GEO: GSE11724 |
| Mouse ESCs, Stat3 ChIP-seq | Ho et al., 2011 | GEO: GSE27708 |
| Mouse ESCs, Zfx ChIP-seq | Chen et al., 2008 | GEO: GSE11431 |
| Mouse ESCs, Dnase-seq | ENCODE | GEO: GSE37074 |
| Mouse ESCs, GRO-seq | Jonkers et al., 2014 | GEO: GSE48895 |
| Mouse ESCs, Start-seq | Williams et al., 2015 | GEO: GSE43390 |
| Mouse ESCs, RNA-seq | Brookes et al., 2012 | GEO: GSE34520 |
| Mouse ESCs, DRIP-RNA-seq | Chen et al., 2015 | GEO: GSE67581 |
| Mouse ESCs, Hi-C | Dixon et al., 2012 | GEO: GSE35156 |
| Mouse ESCs, Capture Hi-C | Schoenfelder et al., 2015 | ArrayExpress E-MTAB-2414 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Mouse ESCs, Pol II ChIA-PET | Zhang et al., 2013 | GEO: GSE44067 |
| Mouse ESCs, Ago2 CLIP-seq | Leung et al., 2011 | http://rowley.mit.edu/pubs/Ago2_CLIP/Ago2_CLIP.html |
| Mouse Exosc3 WT ESCs, RNA-seq | Pefanis et al., 2015 | SRA: SRP042355 |
| Mouse Exosc3 KO ESCs, RNA-seq | Pefanis et al., 2015 | SRA: SRP042355 |
| Mouse BMDM, Pol II ChIP-seq | Ostuni et al., 2013 | GEO: GSE38377 |
| Mouse BMDM, JunB ChIP-seq | Ostuni et al., 2013 | GEO: GSE38377 |
| Mouse BMDM, PU.1 ChIP-seq | Ostuni et al., 2013 | GEO: GSE38377 |
| Mouse BMDM, Stat6 ChIP-seq | Ostuni et al., 2013 | GEO: GSE38377 |
| Mouse Terminal Neuron, Pol II ChIP-seq | Tippmann et al., 2012 | GEO: GSE33252 |
| Mouse Trophoblast Stem Cell, Pol II ChIP-seq | Calabrese et al., 2012 | GEO: GSE39406 |
| Murine Erythroleukemia, Pol II ChIP-Seq | Ghamari et al., 2013 | GEO: GSE46849 |

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact, Dr. Raja Jothi (jothi@nih.gov).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Mouse ES Cell Culture

Mouse ESCs (E14Tg2a, ATCC) were maintained on gelatin-coated plates in the ESGRO complete plus clonal grade medium (Millipore), as previously described (Cinghu et al., 2014; Oldfield et al., 2014). For experiments, ESCs were cultured on gelatin-coated plates in M15 medium: DMEM (Invitrogen) supplemented with 15% FBS, 10μM 2-mercaptoethanol, 0.1 mM nonessential amino acids (Invitrogen), 1x EmbryoMax nucleosides (Millipore), and 1U/ml of ESGRO mLIF (Millipore). Dicer KO mouse ESCs (Novus Biologicals) and Exosc3 KO mouse ESCs, which was a kind gift from U. Basu (University of Columbia) (Pefanis et al., 2015), were grown on inactivated MEFs (GIBCO) using the M15 medium. All cells used in the study were routinely tested for mycoplasma contamination.

## METHOD DETAILS

### Transient Transfection

For transfections, ESCs were cultured in M15 medium and transfected with 50nM siRNA or 100ng plasmid DNA using Lipofectamine 2000 (Invitrogen) as per manufacturer protocol. Gene-specific siRNAs used: Meis1 (QIAGEN) and non-targeting negative control (QIAGEN).

### Double-Reporter Plasmids and Assay

The double-reporter construct was based upon pTN23 (Nasim et al., 2002) plasmid, which was a kind gift from I.C. Eperon (University of Leicester). The construct encodes β-gal and luciferase proteins within a single reading frame but on separate exons, with an intervening intron sequence containing stop codons. Using site directed mutagenesis, multiple point mutations were introduced in the intronic and extragenic (downstream of the reporter gene) regions to facilitate the cloning of DNA segments of interest (intragenic enhancer, extragenic enhancer, or control sequences). Newly created multiple cloning site (MCS) contains restriction sites for restriction enzymes Sal1, Xma1/Sma1, Apa1 and Cla1. Enhancer and control DNA sequences of interest were PCR amplified using mouse genomic DNA as a template, digested with SalI and ApaI (NEB), and cloned into the intronic or extragenic MCS. See Table S6 for the list of primers used for PCR amplification. Mouse ESCs were transfected with 200ng double-reporter plasmids along with 50ng of maxGFP internal control plasmids using lipofectamine 2000 transfection reagent (Invitrogen). Forty-eight hours after transfection, cells were lysed using the lysis buffer from the Dual-Light Luciferase & β-Galactosidase Reporter Gene Assay System (Thermo Fisher). *Firefly* luciferase, β-gal and GFP activities were measured using the Synergy 2 Microplate reader (BioTek). Normalized luciferase or β-galactosidase (β-gal) activity was calculated as the ratio of the luminescence signal divided by the maxGFP fluorescence signal. To measure reporter activities in Exosc3-depleted ESCs, the double-reporter plasmids were co-transfected with Exosc3 or non-targeting negative control siRNAs (QIAGEN), at 50nM concentration, using lipofectamine 2000 (Invitrogen). Forty-eight hours after transfection, the reporter assay was performed as described above.

### CRISPR/Cas9-Mediated Targeted Deletion

Genomic regions of interest (~1100 bp) were deleted in mouse ESCs by co-transfecting two Cas9 containing plasmids (pX330), each carrying a unique guide RNA (gRNA) to direct a double strand break (DSB) at a specific genomic locus. gRNAs were designed with the assistance of the CRISPR Design Tool (crispr.mit.edu) to minimize off-target effects and cloned into the pX330 plasmid, then sequenced verified prior to transfection. See Table S6 for the list of guide RNA sequences. After transfection using lipofectamine 2000 (Invitrogen), ESCs were seeded at low density to allow for selection of individual colonies. Colonies were individually expanded and split for future culture or genomic DNA isolation. 100ng of genomic DNA from these colonies or mouse genomic DNA (control) was screened in a PCR reaction (EmeraldAmp, Takara) with primer pairs annealing to the region outside the DSB sites. See Table S6 for the list of screening primers used to confirm targeted deletions. PCR reactions were separated on 1% agarose gels. ESC DNA gave one of three results: one band migrating the same as control DNA (wild-type/WT), one band migrating faster than control DNA (null), or two bands (heterozygote/het), one corresponding to the wild-type allele and one corresponding to the null allele. Bands were excised from the gel and purified using the QIAquick Kit (QIAGEN), cloned using the TOPO TA cloning kit (Invitrogen), and sequenced (Genewiz) to confirm the status of the deletion. Additionally, genotypes were confirmed by screening with a PCR reaction containing a primer pair annealing to the genomic region inside the location of the two DSBs. Wild-type and heterozygotes gave a single product matching mouse genomic DNA, while nulls gave no PCR product. See Table S6 for the list of guide RNA sequences and screening primers used.

### Quantitative RT-PCR

Quantitative RT-PCR analysis was performed in biological triplicates or quintuplicates. Total RNAs were prepared from cells with the RNeasy Mini kit (QIAGEN). Subcellular RNA fractions were prepared as described previously (Bhatt et al., 2012). Briefly, cytoplasmic and nucleoplasmic RNA was purified using QIAGEN RNeasy columns and chromatin RNA was isolated using QiAzol (QIAGEN), followed by further purification with RNeasy columns. All samples were eluted into 100 uL RNase-free water. cDNAs were generated using the iScript kit (Bio-Rad) according to the manufacturer's instructions. For each biological replicate, quantitative PCR reactions were performed in technical triplicates using the SsoFast EvaGreen supermix (Bio-Rad) on the Bio-Rad CFX-384 or CFX-96 Real-Time PCR System, and the data normalized to *Actin*. Data from individual control/WT samples were normalized to 1, and data from individual mutant/KO/KD samples were normalized to the corresponding control/WT sample. Data from biological replicates are plotted as mean ± SEM. See Table S6 for the list of gene-specific primers used. In the case of double-reporter RNA analysis, primers spanning the Lacz-Luc exon-exon junction were used to quantify the reporter expression, which was normalized to maxGFP expression to control for transfection efficiency. See Table S6 for primers spanning the Lacz-Luc exon-exon junction and GFP primers to control for transfection. The attenuation coefficient, for each enhancer, is calculated as the ratio of normalized mRNA expression of the double-reporter gene in pExtragenic compared to that in pIntragenic.

### Alkaline Phosphatase (AP) Staining

AP staining was performed using the Alkaline Phosphatase Detection Kit (Stemgent), according to the manufacturer's instructions.

### Embryoid Body Differentiation

Wild-type, Meis1-enhancer heterozygote (+/−) and Meis1-enhancer null (−/−) mouse ESCs were grown on non-adhering plates, using the M15 media without the LIF component, and allowed to form cell aggregates in suspension culture for 10 days. Embryoid bodies (EBs) were collected on days 0, 2, 4, 6, 8, and 10, and RNA was isolated as described below. Quantitative RT-PCR reactions were performed, as described below, to assess the expression of various pluripotency- and differentiation-associated genes.

### Western Blot

Cell pellets, lysed in RIPA buffer (25 mM Tris HCl, pH 7.4, 150 mM NaCl, 1% Nonidet P-40, 1% sodium deoxycholate) with protease inhibitors, were sonicated using Bioruptor (Diagenode) for three cycles (30 s on and 50 s off). The lysate was boiled with SDS/PAGE sample buffer, loaded onto a NuPAGE gel, and transferred to 0.22 μM PVDF membranes. Each membrane was treated with appropriate primary and secondary antibodies. The membrane was then incubated with a horseradish peroxidase- conjugated secondary antibody and developed with enhanced chemilluminescence PLUS reagent (Amersham). Loading was normalized based on Ran.

### Identification of Intragenic Sites of Pol II Enrichment

Intragenic sites of Pol II enrichment were determined based on RefSeq gene annotations. "Gene body" is defined as the region beginning 1 Kb downstream of transcription start site (TSS) to transcription end site (TES). For every RefSeq-annotated gene, a 1 Kb sliding window, that slides by 10 bp each time, was used to record the number of Pol II ChIP-seq reads mapping to 1 Kb regions within the gene body; "median gene body Pol II signal" was then calculated as the median of the recorded Pol II read counts across all such 1 Kb windows within the "gene body." The 1 Kb sliding window begins its slide from 1 Kb downstream of TSS, with the window centered at 1 Kb downtream of TSS, and ends at TES, with the window centered at TES. For every 1 Kb region within the entire length of the gene, Pol II Pausing Index (PI) (Adelman and Lis, 2012), defined as the relative ratio of Pol II density within that 1 Kb region to median "gene body" Pol II density, is computed. Regions (1Kb in length) with PI $\geq$ 10 were defined as intragenic

sites of Pol II enrichment, with the mid-points defined as peaks. Those peaks that fall anywhere between TSS and 500 bp downstream of TSS were defined as promoter-proximal Pol II sites (PRSs), provided that they are not within the "gene body" of any known/predicted genes (UCSC known genes). Those peaks that fall near TES (within *D* bp upstream of TES, where *D* = max [1000, min(2000, 5% of the gene length)]) were discarded as cleavage/polyadenylation-related Pol II accumulation near 3′ end of the genes (Core et al., 2008; Kwak et al., 2013; Nojima et al., 2015). Those peaks that fall anywhere between *D* bp downstream of TSS and *D* bp upstream of TES were designated as intragenic Pol II sites (IRSs), provided that they are at least 1 Kb away from TSSs of all known/predicted genes (UCSC known genes). Peaks that fall within the region between 500 and *D* bp downstream of TSS were left unclassified and discarded. Consequently, IRSs were defined only for genes that are at least 2 Kb in length. These definitions resulted in 7,530 PRSs and 1,928 IRSs in ESCs.

### RNA-Seq, GRO-Seq, Start-Seq, and DRIP-RNA-Seq Data Analysis

Sequence reads from RNA-seq experiments were aligned to the mouse reference genome (mm9 assembly) and UCSC annotated transcripts using Tophat (Kim et al., 2013) version v2.0.4. Gene expression was calculated using Cufflinks version v2.0.2, and represented as reads/fragments per kilobase of transcript per million mapped reads (RPKM/FPKM). See the Key Resources Table for the list of publicly available RNA-seq datasets used for analysis. For GRO-Seq, Start-Seq, and DRIP-RNA-seq analysis, mapped reads from previously published spike-in control-normalized GRO-Seq (GEO: GSE48895) (Jonkers et al., 2014), Start-Seq (GEO: GSE43390) (Williams et al., 2015), and DRIP-RNA-seq (GEO: GSE67581) (Chen et al., 2015) data were used.

### ChIP-Seq Data Analysis

Sequence reads generated from ChIP-seq experiments, by various groups, were processed uniformly by aligning to the mouse reference genome (mm9 assembly) using Bowtie (Langmead et al., 2009) version 0.12.8. See the Key Resources Table for the list of publicly available ChIP-seq datasets used for analysis. Only reads that mapped to unique genomic regions with at most two mismatches were retained for follow up analysis. For visualization and generation of screenshots from the UCSC Genome Browser, each ChIP-seq dataset was first normalized by the total number of reads and then to one million reads. ChIP-seq read density plots were generated by calculating the number of reads within ± 5 kb upstream and downstream of sites of interest in 100 bp windows, and normalized to reads per base/kilobase per million reads (RPM/RPKM) and plotted as histograms. Data for heatmaps were generated similarly.

### Chromatin Interaction Map Analysis

Chromatin interaction between IRSs and promoters were determined based on published Capture Hi-C (E-MTAB-2414) (Schoenfelder et al., 2015) and Pol II ChIA-PET (GEO: GSE44067) (Zhang et al., 2013) datasets in ESCs. An IRS is considered to interact with a promoter if the 1 Kb region centered around the IRS center and 1 Kb region centered around transcription start site (promoter) has been observed to interact in the Capture Hi-C and/or and Pol II ChIA-PET interaction maps.

### Ago2 CLIP-Seq Data Analysis

To assess whether RNA synthesized from intragenic enhancers might be involved in RNAi-mediated gene silencing, we compared intragenic enhancers (IRSs) to the list of published genomic sites encoding Ago2-associated RNAs in ESCs (Leung et al., 2011), as determined using Ago2-CLIP-Seq. RNA from an intragenic enhancer is considered to be loaded into Ago2-containing RNA-induced silencing complexes if one or more genomic sites encoding Ago2-associated RNAs are located within the 1000 nt region of the intragenic enhancer.

### Expressed Sequence Tag (EST) Analysis

ESTs are short sub-sequences of cDNA sequences. Mouse ESTs (source: dbEST; URL: https://www.ncbi.nlm.nih.gov/dbEST/) aligned to the mouse reference genome (mm9 assembly), obtained from UCSC Genome Browser, were used for analysis. About 4.37 million mapped strand-specific mouse ESTs were used to generate EST density plots by calculating the number of EST starts within ± 5 kb upstream and downstream of sites of interest (TSS/TES/intragenic enhancer) in 100 bp windows, and normalized to tags per kilobase per million tags and plotted as histograms.

### Sequence Analysis

TRAP (Thomas-Chollier et al., 2011) was used to search for known TF motifs, obtained from Jaspar (Portales-Casamar et al., 2010) and TRANSFAC (Qian et al., 2006) databases, within DNA sequences spanning 500 nucleotides around the centers of IRSs or matched intragenic control sequences (defined as the DNA sequences spanning 500 nucleotides around the gene mid-point). Statistical significance for enrichment of sequence motifs within IRSs or intragenic control regions were calculated in reference to sequences from promoter regions. Benjamini-Hochberg method was used for multiple-testing correction. CpG island annotations were downloaded from the UCSC genome browser.

### Functional Enrichment Analysis

Intragenic enhancer-containing genes in ESCs were analyzed in relation to the rank-ordered list of all mouse genes (Cinghu et al., 2014) that are likely to be associated with ESC maintenance versus ESC differentiation.

## QUANTIFICATION AND STATISTICAL ANALYSIS

See the STAR Methods Details for details of quantification and statistical analysis.

**Molecular Cell, Volume *68***

## Supplemental Information

# Intragenic Enhancers Attenuate Host Gene Expression

**Senthilkumar Cinghu, Pengyi Yang, Justin P. Kosak, Amanda E. Conway, Dhirendra Kumar, Andrew J. Oldfield, Karen Adelman, and Raja Jothi**

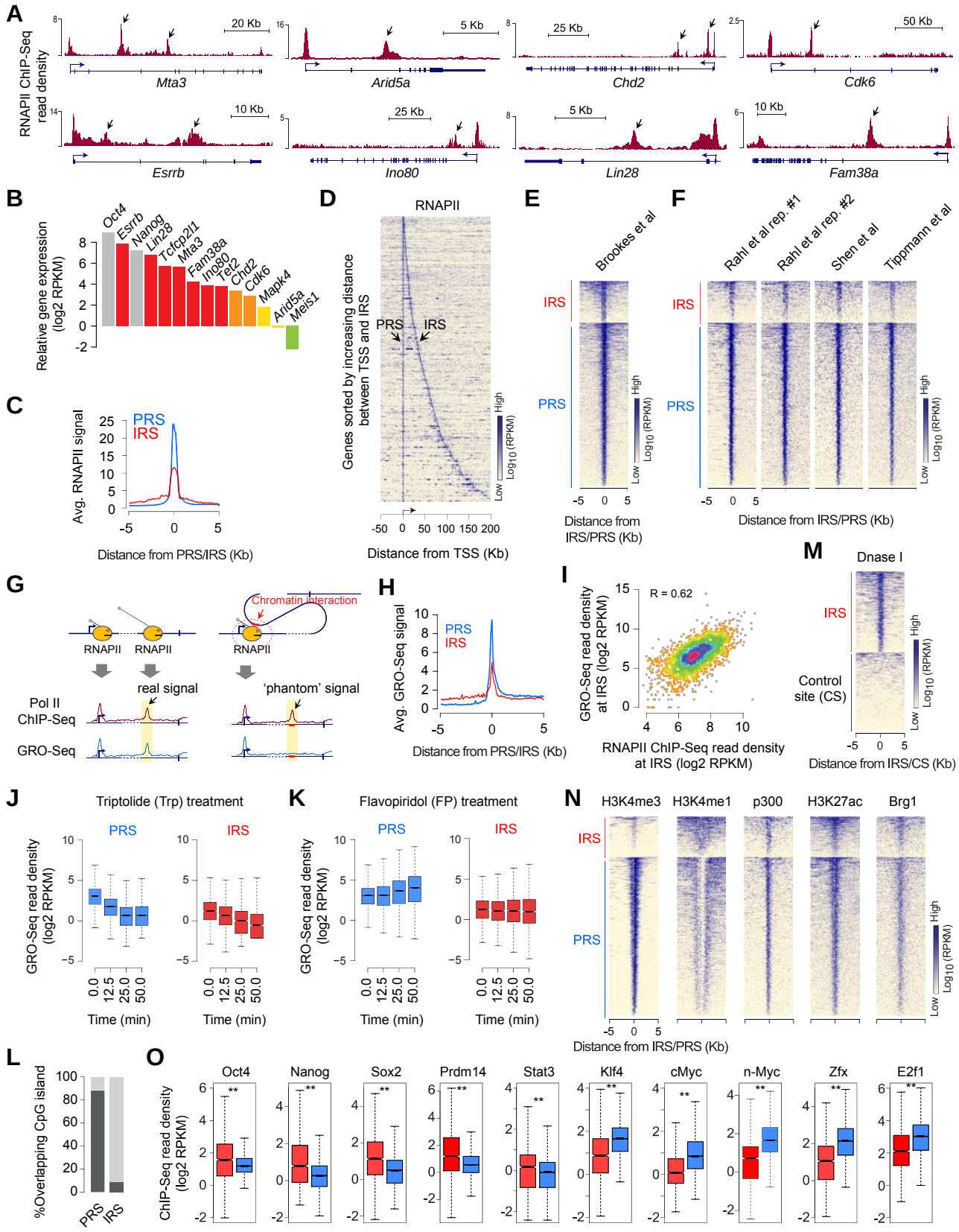

A
RNAPII ChIP-Seq read density
Mta3
Arid5a
Chd2
Cdk6
Esrrb
Ino80
Lin28
Fam38a
B
Relative gene expression (log2 RPKM)
Oct4
Esrrb
Nanog
Lin28
Tcfcp2l1
Mta3
Fam38a
Ino80
Tet2
Chd2
Cdk6
Mapk4
Arid5a
Meis1
C
Avg. RNAPII signal
PRS
IRS
Distance from PRS/IRS (Kb)
D
RNAPII
Genes sorted by increasing distance between TSS and IRS
PRS
IRS
Distance from TSS (Kb)
E
Brookes et al
IRS
PRS
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
PRS
IRS
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
IRS
Control site (CS)
Distance from IRS/CS (Kb)
J
Triptolide (Trp) treatment
PRS
IRS
GRO-Seq read density (log2 RPKM)
Time (min)
K
Flavopiridol (FP) treatment
N
H3K4me3
H3K4me1
p300
H3K27ac
Brg1
IRS
PRS
Distance from IRS/PRS (Kb)
L
%Overlapping CpG island
PRS
IRS
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 RNA Pol II enrichment mark transcriptionally active intragenic enhancers**

(A) ChIP-Seq profiles of RNA Pol II (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).

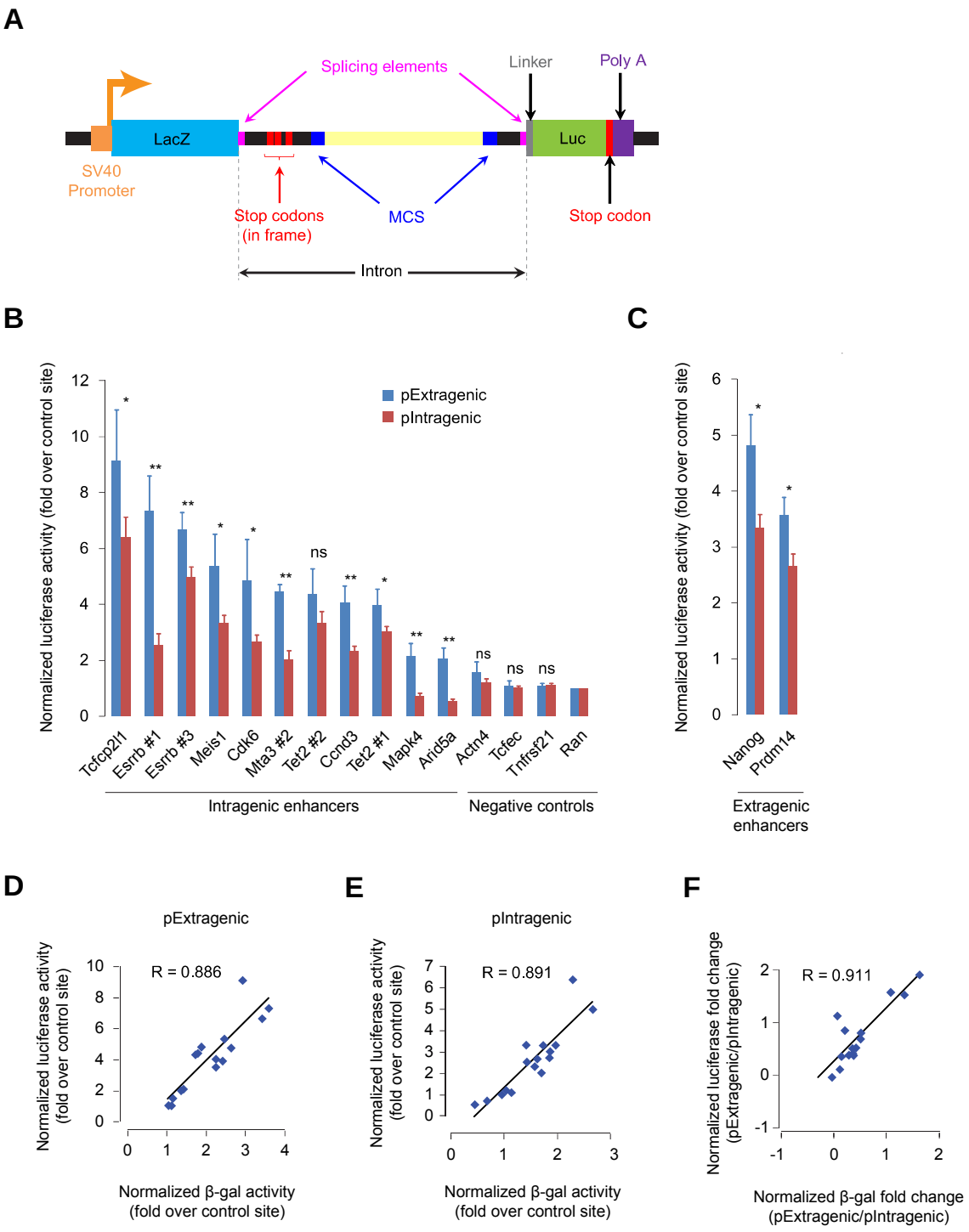



**Figure S2 (related to Figure 3). Enhancers in intragenic position attenuate reporter activity**

(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).

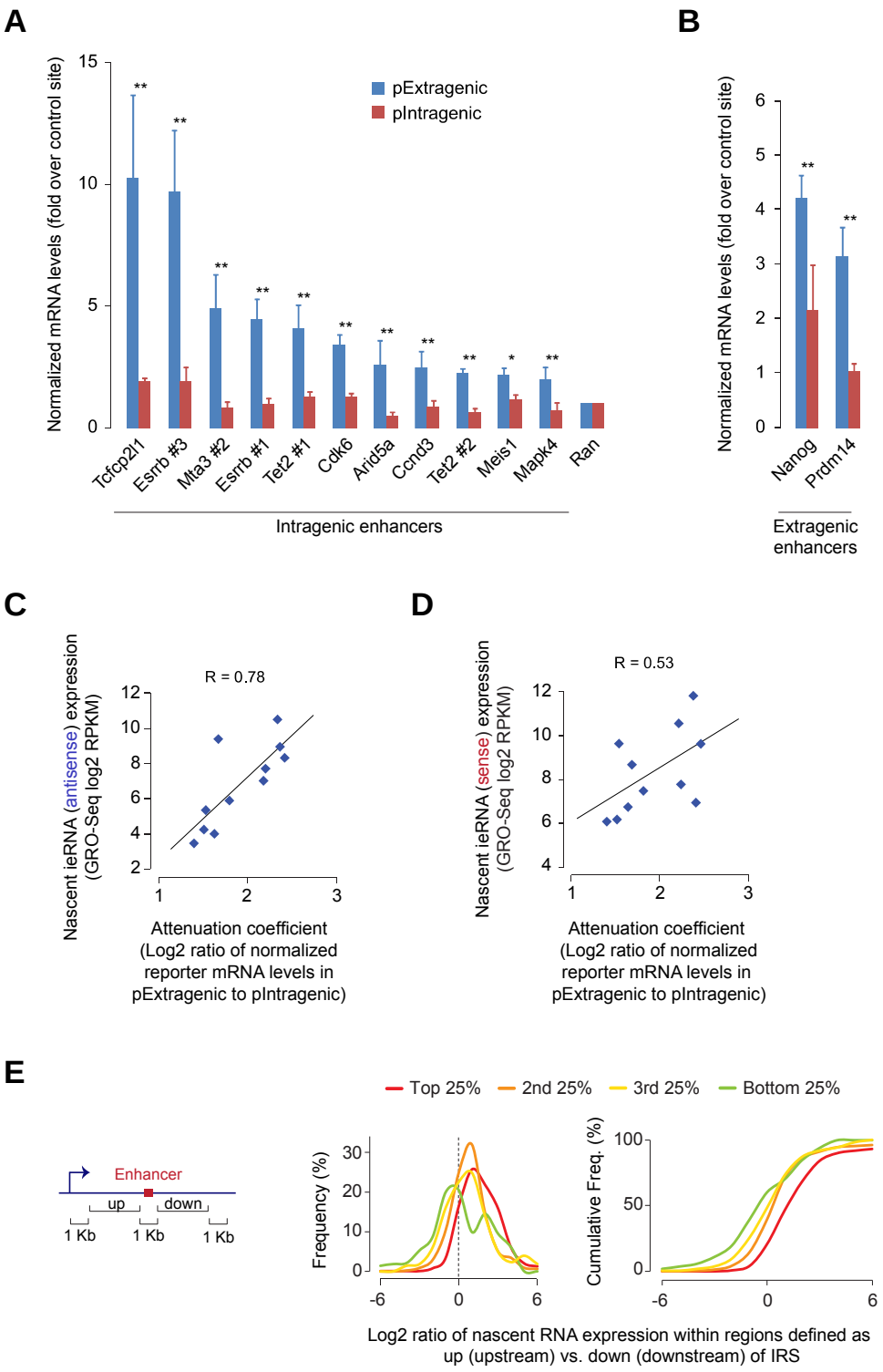



**Figure S3 (related to Figure 3 and 4). Enhancers in intragenic position attenuate host gene transcription**

(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.

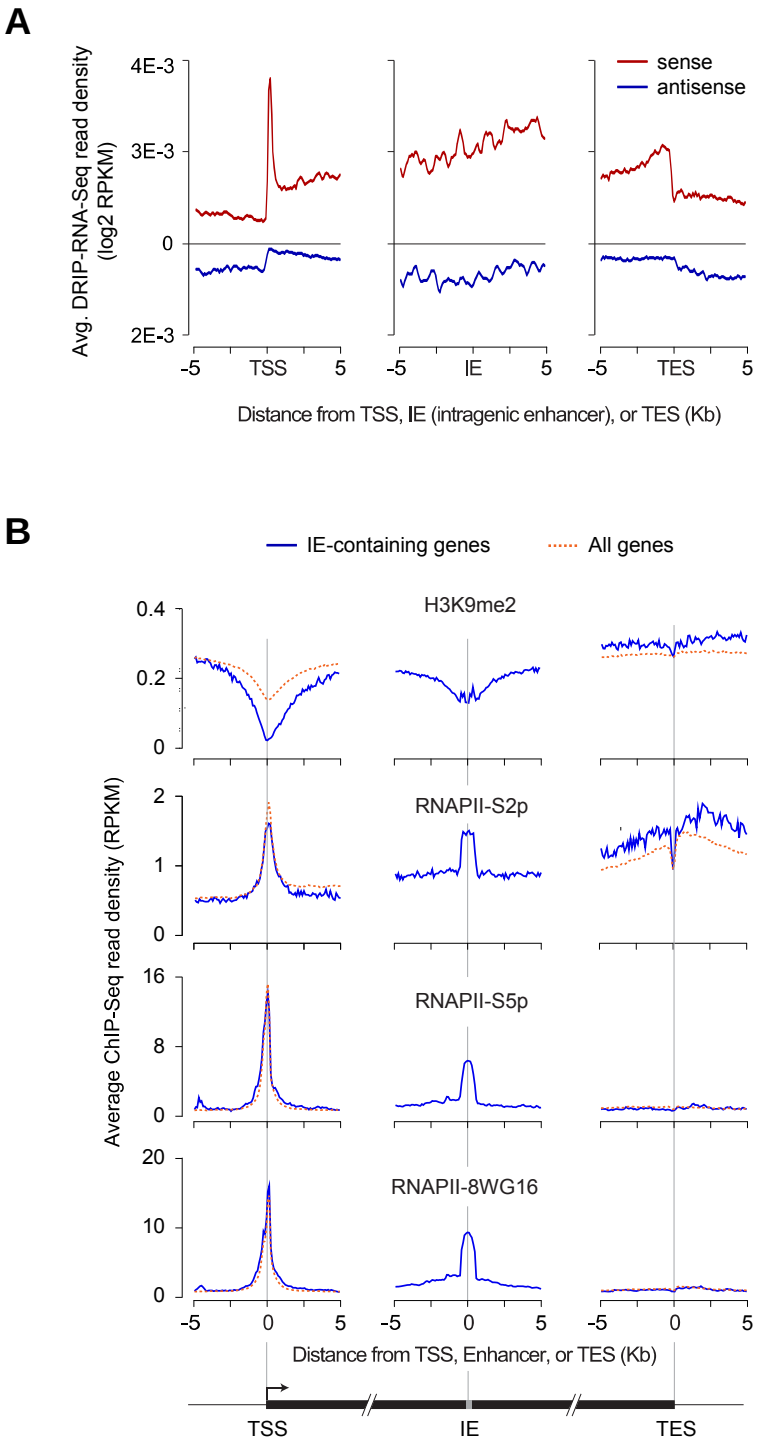



**Figure S4 (related to Figure 5). Intragenic enhancers are not enriched for R-loops or R-loop-mediated RNAi-dependent H3K9me2**

(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, RNA Pol II (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.

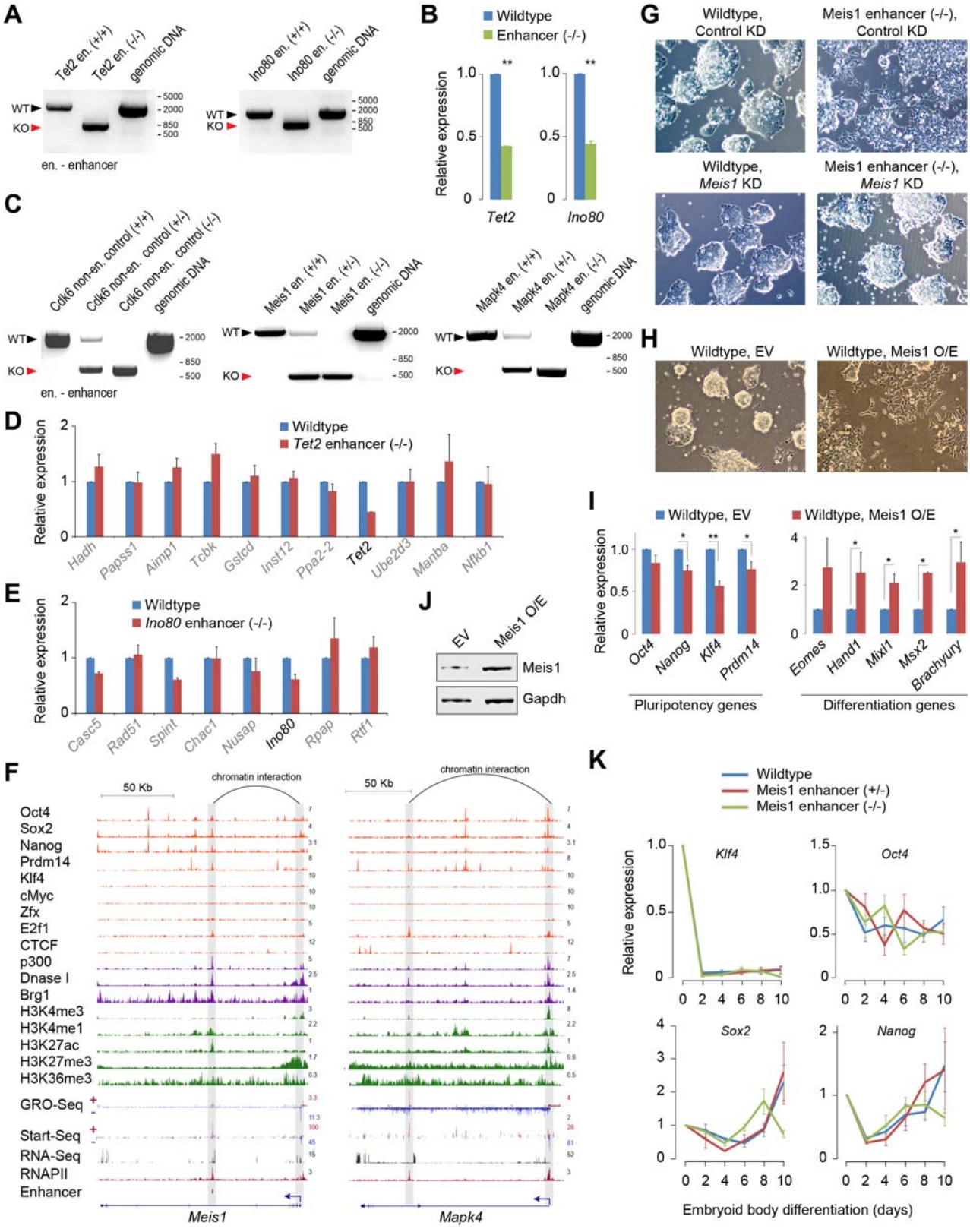

A
Tet2 en. (+/+)
Tet2 en. (-/-)
genomic DNA
Ino80 en. (+/+)
Ino80 en. (-/-)
genomic DNA
WT
KO
5000
2000
850
500
en. - enhancer
B
Wildtype
Enhancer (-/-)
Relative expression
Tet2
Ino80
C
Cdk6 non-en. control (+/+)
Cdk6 non-en. control (+/-)
Cdk6 non-en. control (-/-)
genomic DNA
Meis1 en. (+/+)
Meis1 en. (+/-)
Meis1 en. (-/-)
Mapk4 en. (+/+)
Mapk4 en. (+/-)
Mapk4 en. (-/-)
D
Wildtype
Tet2 enhancer (-/-)
Hadh
Papss1
Aimp1
Tcbk
Gstcd
Inst12
Ppa2-2
Tet2
Ube2d3
Manba
Nfkb1
E
Wildtype
Ino80 enhancer (-/-)
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
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
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 RNA Pol II (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.

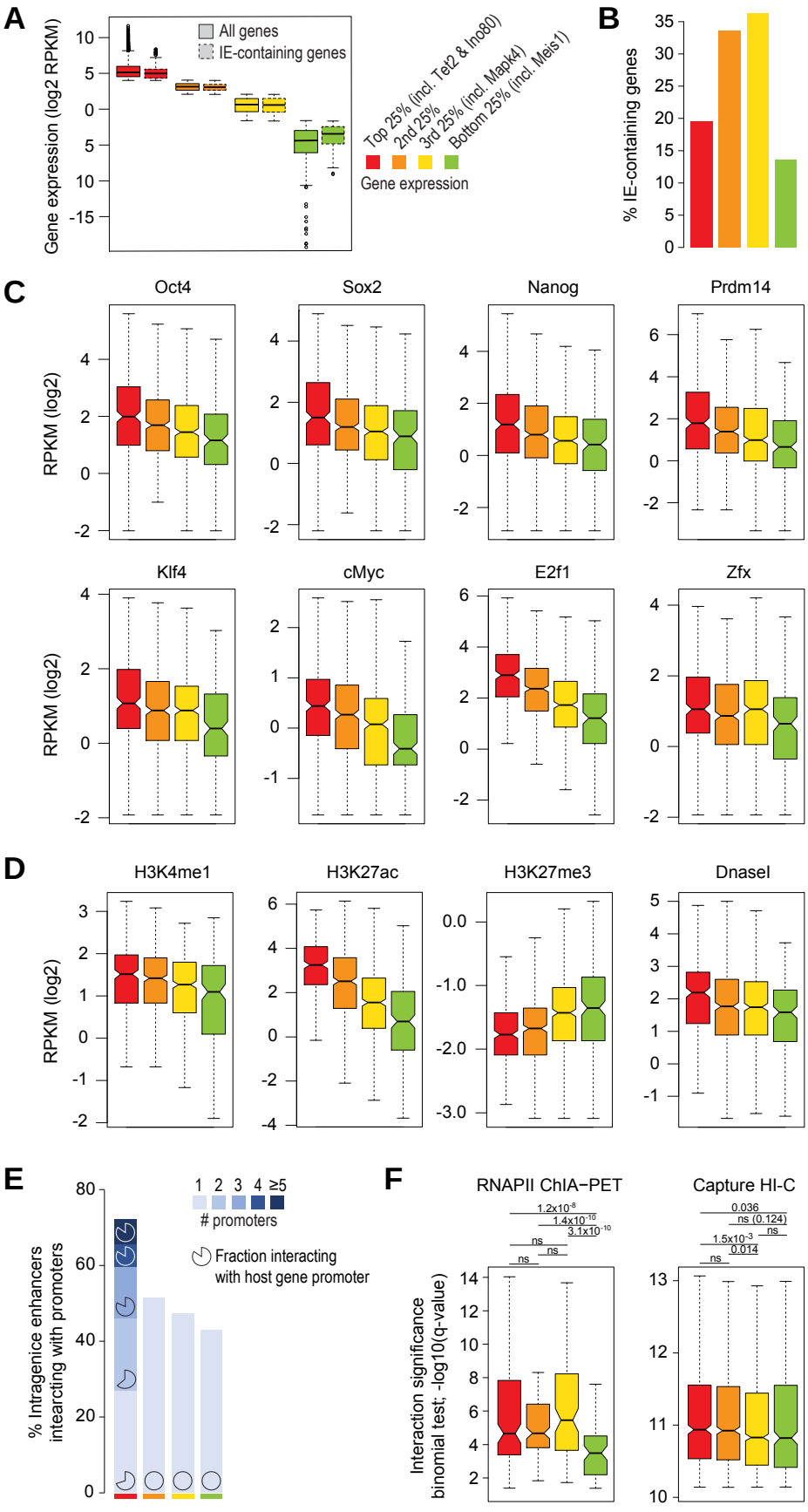

A
Gene expression (log2 RPKM)
All genes
IE-containing genes
Top 25% (incl. Tet2 & Ino80)
2nd 25%
3rd 25% (incl. Mapk4)
Bottom 25% (incl. Meis1)
Gene expression
B
% IE-containing genes
C
Oct4
Sox2
Nanog
Prdm14
Klf4
cMyc
E2f1
Zfx
RPKM (log2)
D
H3K4me1
H3K27ac
H3K27me3
DnaseI
RPKM (log2)
E
% Intragenice enhancers
intearcting 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
ns
ns
0.036
ns (0.124)
ns
1.5x10-3
0.014
ns

**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 RNA Pol II (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).

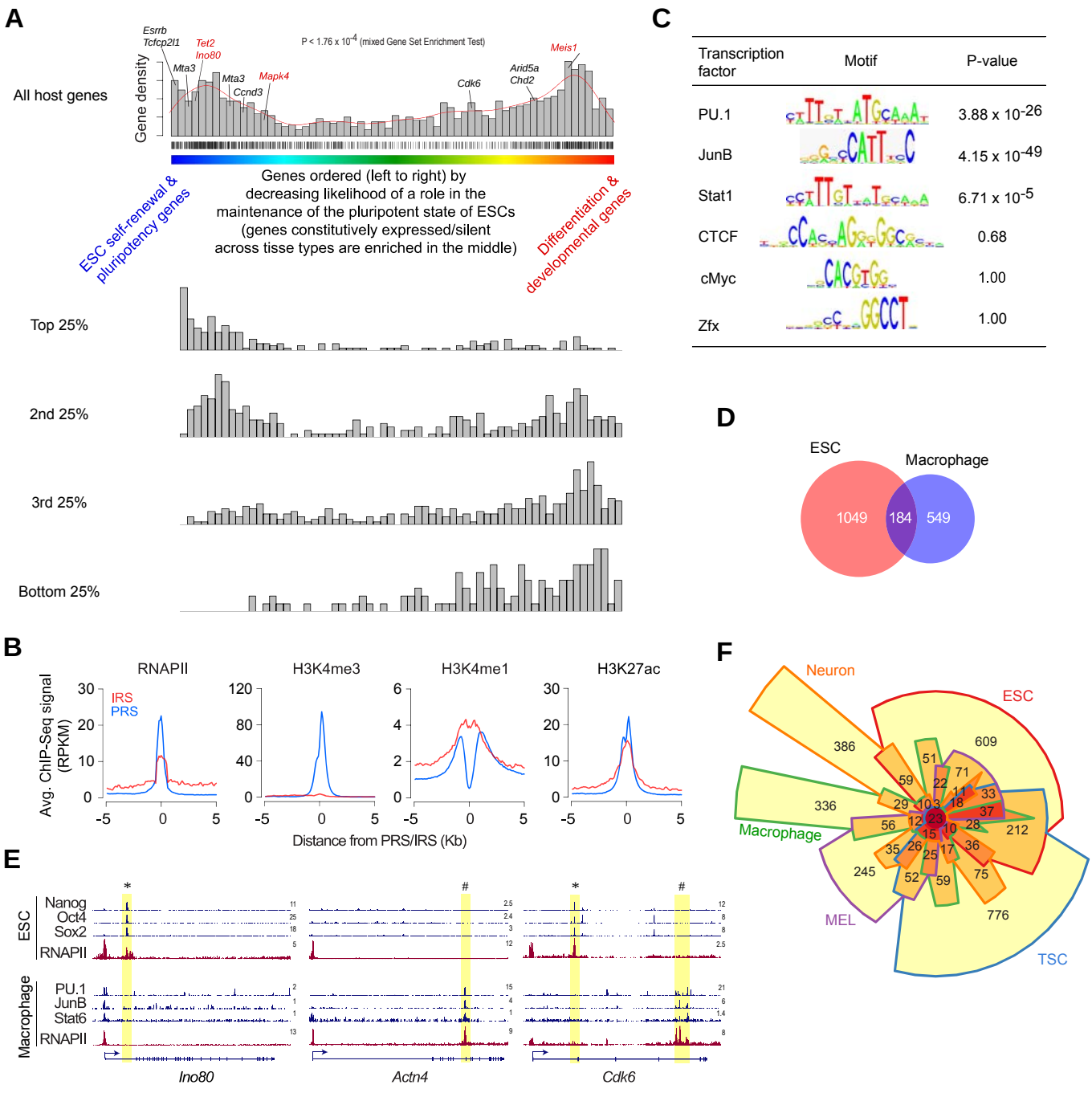

A
All host genes
Gene density
Esrrb
Tcfcp2l1
Tet2
Ino80
Mta3
Mta3
Ccnd3
Mapk4
Cdk6
Arid5a
Chd2
Meis1
P < 1.76 x 10-4 (mixed Gene Set Enrichment Test)
Genes ordered (left to right) by decreasing likelihood of a role in the maintenance of the pluripotent state of ESCs (genes constitutively expressed/silent 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.