Molecular Cell
Article

Intragenic Enhancers Act as Alternative Promoters

Monika S. Kowalczyk,[1,6] Jim R. Hughes,[1,6] David Garrick,[1,7] Magnus D. Lynch,[1,7] Jacqueline A. Sharpe,[1] Jacqueline A. Sloane-Stanley,[1] Simon J. McGowan,[2] Marco De Gobbi,[1] Mona Hosseini,[3] Douglas Vernimmen,[1] Jill M. Brown,[1] Nicola E. Gray,[1] Licio Collavin,[4,5] Richard J. Gibbons,[1] Jonathan Flint,[3] Stephen Taylor,[2] Veronica J. Buckle,[1] Thomas A. Milne,[1] William G. Wood,[1] and Douglas R. Higgs[1,*]

[1]MRC Molecular Haematology Unit
[2]Computational Biology Research Group
Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK
[3]Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK
[4]Laboratorio Nazionale CIB, AREA Science Park, 34012 Trieste, Italy
[5]Dipartimento di Scienze della Vita, Università degli Studi di Trieste, 34129 Trieste, Italy
[6]These authors contributed equally to this work
[7]These authors contributed equally to this work
*Correspondence: doug.higgs@imm.ox.ac.uk
DOI 10.1016/j.molcel.2011.12.021

SUMMARY

A substantial amount of organismal complexity is thought to be encoded by enhancers which specify the location, timing, and levels of gene expression. In mammals there are more enhancers than promoters which are distributed both between and within genes. Here we show that activated, intragenic enhancers frequently act as alternative tissue-specific promoters producing a class of abundant, spliced, multiexonic poly(A)$^+$ RNAs (meRNAs) which reflect the host gene's structure. meRNAs make a substantial and unanticipated contribution to the complexity of the transcriptome, appearing as alternative isoforms of the host gene. The low protein-coding potential of meRNAs suggests that many meRNAs may be byproducts of enhancer activation or underlie as-yet-unidentified RNA-encoded functions. Distinguishing between meRNAs and mRNAs will transform our interpretation of dynamic changes in transcription both at the level of individual genes and of the genome as a whole.

INTRODUCTION

Transcriptional start sites (TSSs) of all currently defined classes of polyadenylated RNA are thought to be defined by one type of *cis* element (generically referred to as promoters) which initiates cell-specific mRNA transcripts. It is now clear that the transcriptome is far more complex than originally proposed (Carninci et al., 2005; Kapranov et al., 2007; Kapranov et al., 2005). A large proportion of the mammalian genome is transcribed in a developmental stage and cell type-specific manner to produce many different classes of RNA (Mattick et al., 2010). Understanding how cell- and developmental-stage-specific expression is determined will be greatly facilitated by identifying the origins of these different classes of RNA. This should allow us to evaluate their potential significance in more detail before embarking on appropriate functional assays.

Recently it was shown that active promoters (whose chromatin is marked by high levels of H3K4me3 and low levels of H3K4me1 [Heintzman et al., 2007, 2009]) are transcribed to generate short bidirectional RNAs centered around TSSs (Core et al., 2008; He et al., 2008; Preker et al., 2008; Seila et al., 2008). Despite this, transcriptional elongation only occurs in the direction of the gene to produce spliced, poly(A)$^+$ RNAs (mRNA). Of interest, it was recently shown that some intergenic, and possibly some intragenic, enhancers (whose chromatin is marked by high levels of H3K4me1 and low levels of H3K4me3) may also be transcribed to produce short bidirectional transcripts, called eRNAs (Kim et al., 2010), which may (Kim et al., 2010) or may not be polyadenylated (De Santa et al., 2010). No elongated poly(A)$^+$ transcripts originating from these enhancers were detected (De Santa et al., 2010; Kim et al., 2010), and it was concluded that, although similar in some respects, promoters and enhancers produce different classes of RNA transcripts: enhancers are not promoters (Kim et al., 2010).

It is estimated that mammalian genomes contain more enhancers than promoters (Bulger and Groudine, 2010), and where analyzed, up to 51% of enhancers lie within the gene body (Heintzman et al., 2007). The gene hosting the enhancers may lie tens to thousands of kilobases away from the genes regulated by the elements and is often unrelated to the target gene. In view of the recent observations on transcription of intergenic enhancers, it was of interest to investigate in detail how activation of intragenic enhancers might affect transcription and expression of their host genes.

To address this, we analyzed a well-characterized set of five erythroid-specific enhancers, four of which lie within the body of the mouse *Nprl3* gene and one which lies upstream from the promoter (R1–R4, Figures 1A and 1B). In erythroid cells, these enhancers coincide with *DnaseI* hypersensitive sites (DHSs), bind erythroid-specific and widely expressed transcription factors, and recruit low levels of RNA polymerase II (RNAP2)

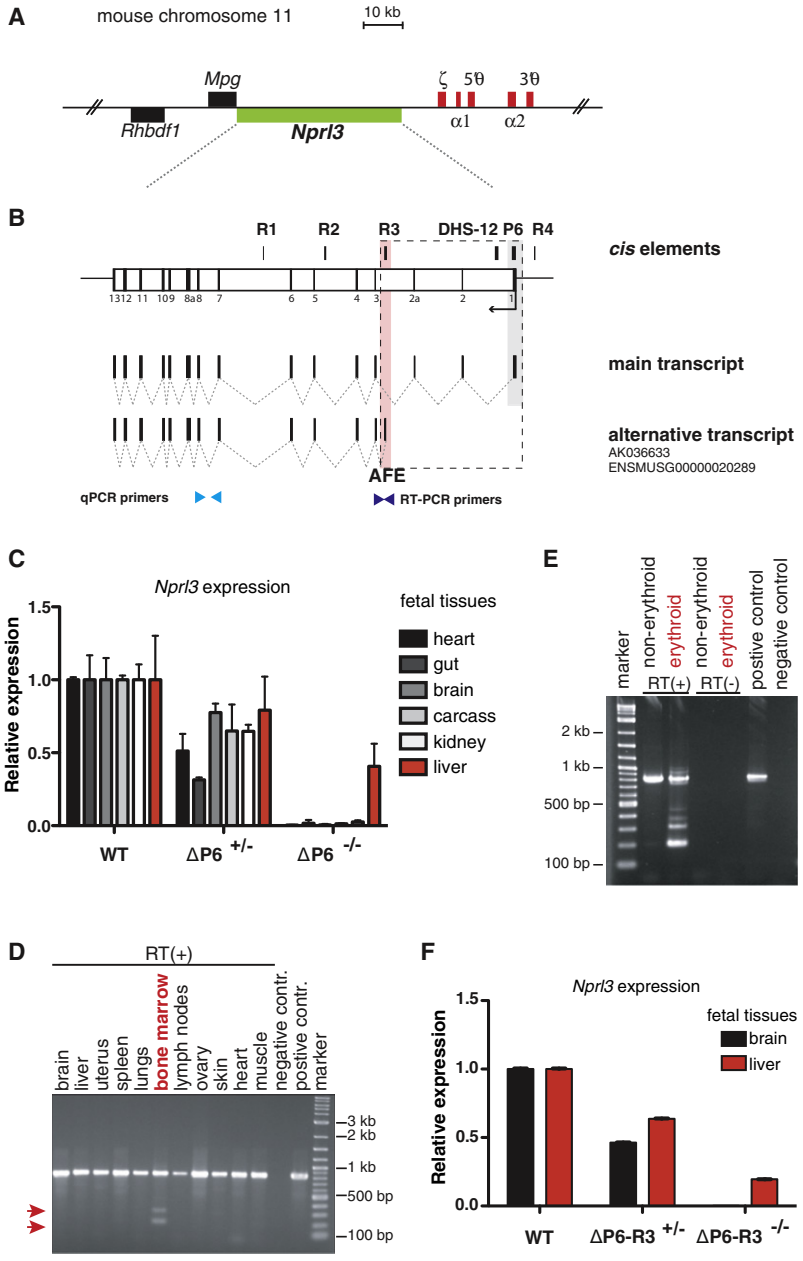

Figure 1. *Nprl3* Promoter Knockout Failed to Abolish *Nprl3* Transcription in Erythroid Cells

(A) The mouse α-globin cluster. The globin genes are shown as red boxes. The *Nprl3* gene containing enhancers is in green. Other genes are shown as black boxes. Genes above the line are transcribed from left to right and those below the line from right to left.

(B) The genomic structure of the *Nprl3* gene with associated mRNA isoforms. *cis* elements are shown above the *Nprl3* gene, with R and P representing regulatory and promoter elements. DHS-12 is a mouse-specific DHS (Anguita et al., 2004). The direction of the *Nprl3* gene transcription is shown as a black arrow. *Nprl3* mRNA isoforms are shown below the genomic structure. The P6 deletion (ΔP6) is indicated as a gray rectangle (see also Figure S1A). The deletion between P6 and R3 is shown as a dashed box. The alternative first exon (AFE) of the alternative transcript coincides with enhancer R3 and is highlighted within the red rectangle. Quantitative RT-PCR (qPCR) primers (light blue) used in (C) and (F) span the junction between exon 7 and 8. RT-PCR primers (dark blue) used in (D) and (E) span the junction between AFE and exon 3. Black boxes denote exons and white boxes introns, and splicing of mature transcript is shown by dashed lines.

(C) mRNA expression of *Nprl3* across mouse fetal tissues at embryonic day E14.5. Fetal liver at this stage of development is a source of erythroid cells. qPCR primer positions are depicted in (B) (dark blue arrows). *Nprl3* expression was calculated as in (F). All error bars represent ±SD; $n \geq 3$.

(D and E) (D) Specific products corresponding to the alternative transcript were detectable only in cDNA sample from mouse bone marrow (red arrowheads pointing to the PCR products). Conventional sequencing revealed that the ~300 bp band corresponds to the predicted product size (311 bp), and the ~200 bp band was a splice variant of the alternative transcript. All tissues showed a band (~800 bp) corresponding to the unspliced *Nprl3* transcript. (E) The alternative transcript is present in mouse erythroid cells (Ter119+). Similarly to (D), a band (~800 bp) corresponds to the unspliced *Nprl3* transcript. A ladder of bands specific to erythroid cells corresponds to the PCR products containing splice variants of AFE. For (D) and (E), RT-PCR was performed with primers shown in (B) using cDNA from mouse WT tissues (D) and mouse nonerythroid (Ter119−) and erythroid (Ter119+) cells (E). Mouse genomic DNA was used as a positive control, and negative control excluded polymerase. RT(−) reactions excluded reverse transcriptase.

(F) mRNA expression of *Nprl3* in mouse *ΔP6-R3* fetal tissues at embryonic day E14.5. Fetal liver and brain were used as erythroid and nonerythroid tissue, respectively. For (C) and (F), the qPCR primers used are shown on (B) and span the exon 7–8 junction. *Nprl3* expression was calculated relative to mouse *Gapdh*. For each tissue, the mean of expression in WT is set to 1, and expression in linked samples is expressed relative to this mean. All error bars represent ±SD; $n \geq 3$.

(Anguita et al., 2004). The chromatin associated with these enhancers (in human) is characteristically modified by high levels of H3K4me1 and low levels of H3K4me3 (De Gobbi et al., 2007). Moreover, they have been shown to physically interact with the α-globin promoters (Lower et al., 2009; Vernimmen et al., 2007) in a tissue-specific manner. Finally, it has been shown in transgenic experiments and by natural deletions that these elements are essential for high levels of α-globin expression (Anguita et al., 2002; Higgs and Wood, 2008).

Analyzing the transcription of such intragenic enhancers is obscured by transcription of the host gene (in this case *Nprl3*), and consequently transcription of these elements has been largely ignored. To address this problem, we analyzed transcription of the *Nprl3* gene after deleting its constitutive promoter. This revealed that intragenic enhancers act as highly active, alternative tissue-specific promoters for the gene containing them. Using a genome-wide approach we have shown that many activated intragenic enhancers behave as alternative

promoters producing an abundant class of long poly(A)$^+$ RNAs (referred to here as multiexonic RNAs derived from enhancers, or meRNAs). The expression of meRNAs explains a substantial proportion of the complexity of the transcriptome and how this changes from one cell type to another.

RESULTS

Erythroid-Specific Expression of the *Nprl3* Gene Continues in the Absence of Its Constitutive Promoter

Nprl3 is a widely expressed gene lying adjacent to the α-globin cluster (Figure 1A). It contains four intragenic enhancers and a fifth intergenic enhancer lying 1.4 kb upstream from its constitutive promoter (Figure 1B); four of the five elements are associated with multiple conserved sequences (Hughes et al., 2005). These elements (R1–R4) act as erythroid-specific enhancers of the α-globin genes. This locus thus provides a paradigm for examining in detail the effects of enhancers on transcription.

Since transcription of intragenic enhancers may be masked by transcription of the gene that contains them, we used homologous recombination to delete the *Nprl3* promoter (ΔP6), allowing us to examine any independent transcription originating from the enhancers (Figure 1B; see Figure S1A available online). Using RNA-FISH, we showed that nascent transcription of *Nprl3* was abolished in nonerythroid cells but surprisingly not in erythroid cells (Figures S1B and S1C). Similarly, stable mRNA was detected by qPCR in erythroid cells, but not in a variety of nonerythroid cells. Even in the absence of the *Nprl3* promoter, abundant expression of the *Nprl3* gene (~50% of wild-type level) persisted in erythroid cells (fetal liver in Figure 1C).

We have shown that expression of the *Nprl3* gene is upregulated in human (Figure S1D) (Lower et al., 2009) and mouse erythroid cells (Figure S1E), and the observations presented here suggest that a substantial proportion of the expression of *Nprl3* mRNA seen in erythroid cells is derived from an erythroid-specific alternative promoter(s) distinct from the canonical *Nprl3* promoter. An EST in the mouse genome (AK036633), annotated as a gene isoform by Ensembl (ENSMUSG00000020289), has a start site lying downstream from the main promoter of the *Nprl3* gene. Annotation of this alternative transcript shows a unique alternative first exon (AFE) (Figure 1B), which is formed from within intron 2 of the *Nprl3* gene. Beyond this AFE the isoform structure appears identical to the full-length transcript. Interestingly, the AFE coincides with one of the known intragenic enhancers (R3).

First, we tested wild-type mouse tissues by RT-PCR to verify the existence of this alternative transcript. Primers spanning the unique AFE-exon3 junction generated two specific PCR products only in bone marrow cDNA (Figure 1D). Sequence analysis confirmed both to be specific products capturing AFE-exon3 splice junctions. Although it seemed likely that these are erythroid-specific transcripts, since bone marrow contains a mixture of cells from all hematopoietic lineages, we tested purified populations of erythroid (Ter119+) and nonerythroid (Ter119−) cells. Several erythroid-specific PCR products were amplified, and sequence analysis showed a series of related exonic junctions, where different donor splice sites within the AFE are used (Figure 1E). These observations show that although R3 has the hallmarks of an enhancer, when activated in erythroid cells, it can also behave as an alternative, independent, internal promoter of the *Nprl3* gene.

Transcription of *Nprl3* Persists in the Absence of Both the Constitutive Promoter and the R3 Enhancer

In a further attempt to abolish transcription of the *Nprl3* gene in erythroid cells, we deleted both the canonical *Nprl3* promoter and the R3 enhancer (ΔP6-R3) (Figure S2A). Analyzing expression of *Nprl3* mRNA in this mouse model again showed no expression in the brain (as expected), but surprisingly still showed about 20% (with respect to WT) of *Nprl3* mRNA in erythroid cells (Figure 1F).

Since deletion of both the constitutive promoter and the alternative erythroid-specific promoter still failed to abolish *Nprl3* transcription in erythroid cells, we next set out to determine the origin(s) of the remaining transcription in the *Nprl3* locus. We analyzed total RNA from erythroid and nonerythroid cells of wild-type, *ΔP6*$^{-/-}$, and *ΔP6-R3*$^{-/-}$ mice using custom tiled arrays. Again, expression of the *Nprl3* gene in both *ΔP6*$^{-/-}$ and *ΔP6-R3*$^{-/-}$ was undetectable in brain (Figure S2B). By contrast, in erythroid cells of both knockouts the only nontranscribing region of the *Nprl3* gene coincided with the extent of each deletion.

Of interest, we noted that in both of these knockout mice there was a clear peak of erythroid-specific transcription upstream of the *Nprl3* promoter, which coincides with the conserved intergenic enhancer R4. Transcription associated with such intergenic enhancers (eRNAs) was previously described as short and bidirectional, as occurs from promoters (De Santa et al., 2010; Kim et al., 2010).

Enhancers in and around the *Nprl3* Locus Produce eRNAs

One possibility is that transcription from the enhancers contributes to the erythroid-specific expression of mRNA from the *Nprl3* gene, although short eRNAs could not account for expression extending throughout the gene (Figure S2B). To reveal the mechanism by which transcription of the *Nprl3* gene continues despite deleting two of its known TSSs, we analyzed transcription using high-throughput sequencing (RNA-Seq). Since previous attempts failed to resolve whether eRNAs are polyadenylated (De Santa et al., 2010; Kim et al., 2010), we analyzed both poly(A)$^+$ and poly(A)$^-$ RNA from wild-type erythroid cells.

First we analyzed the poly(A)$^-$ RNA of the WT erythroid cells and compared this with high-resolution chromatin maps (Figure 2). Strand-specific analysis showed that there is abundant primary transcription in the direction of *Nprl3* transcription (bottom strand), but also prominent and discrete antisense RNA transcripts (~1 kb in length, top strand) associated with each of the enhancers including the intergenic element (R1–R4, Figure 2). The intergenic enhancer R4 was associated with a separate peak of antisense transcription, which is distinct from that associated with the *Nprl3* constitutive promoter (see red column, Figure 2). The intergenic location of this element allowed us to see bidirectional poly(A)$^-$ eRNAs originating from this element, as previously noted for other intergenic enhancers

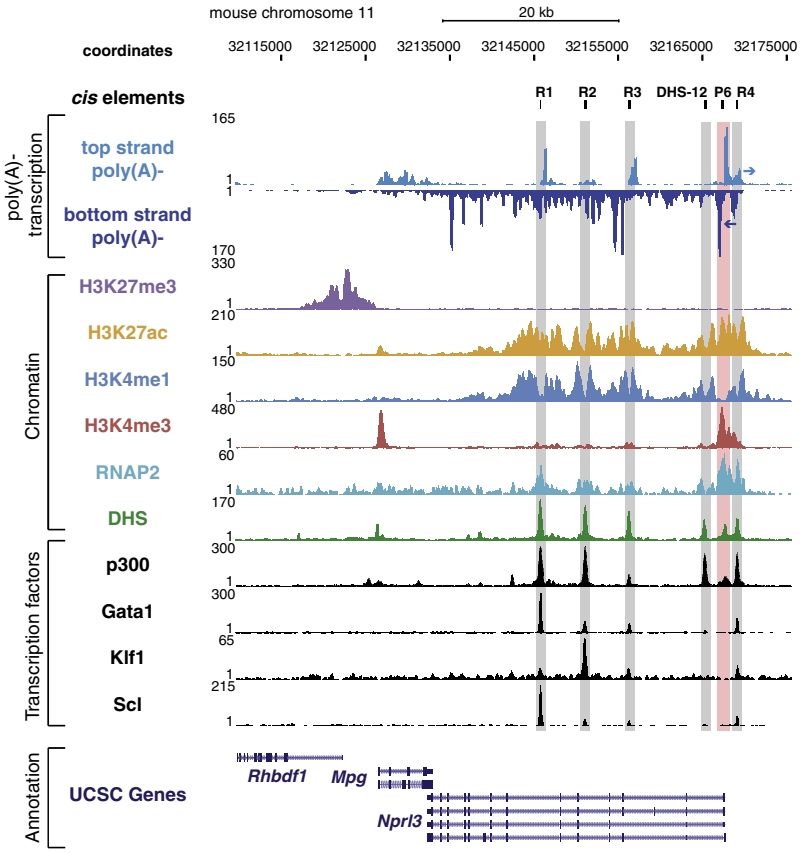

Figure 2. The Epigenetic and Transcriptional Landscape of the *Nprl3* Locus in Mouse Erythroid Cells

The mouse *Nprl3* locus with high-resolution maps of poly(A)$^-$ transcription, chromatin states, and erythroid-specific transcription factor binding. Poly(A)$^-$ RNA-Seq data was split into top (light blue) and bottom strand (dark blue). Arrows indicate bidirectional transcription from the center of enhancer R4. The y axis represents read density. The enhancers and the *Nprl3* promoter are highlighted as gray and red columns, respectively. UCSC Genome annotation is shown at the bottom.

(Kim et al., 2010). The remainder of the enhancers are embedded in the body of the *Nprl3* gene, and any sense transcription originating from these elements was masked by transcription of the host gene.

Activated Enhancers Direct Expression of Abundant Poly(A)$^+$ Isoforms

We next analyzed the poly(A)$^+$ RNA. To account for any alternatively spliced transcripts, we used algorithms (TopHat and Cufflinks), which do not depend on gene annotation (Trapnell et al., 2009, 2010) to reconstruct gene isoforms from the spliced poly(A)$^+$ RNA-Seq data (Figure 3A). Unexpectedly, this predicted four isoforms extending from the intragenic enhancers (R2 and R3, DHS-12) and also from the nearby intergenic enhancer (R4) to the poly(A) site of the *Nprl3* gene. These long transcripts encode unique AFEs spliced onto an adjacent annotated exon with the remainder of the transcript spliced and polyadenylated in the same manner as the host gene.

Each of the predicted, alternative isoforms were verified by RT-PCR between the unique 5′ exons (AFEs) and distal *Nprl3* exons confirming their spliced structure and erythroid specificity in wild-type cells (Figure S3A) and in the two knockout models ($\Delta P6^{-/-}$ and $\Delta P6\text{-}R3^{-/-}$) (Figures S3B and S3C).

To further characterize the enhancer-derived polyadenylated transcripts and their abundance with respect to the levels of RNA from the intact gene, we performed RNA-Seq in erythroid cells derived from $\Delta P6\text{-}R3^{-/-}$ mice (Figure 3B). Even in the absence of the constitutive *Nprl3* promoter and two intragenic enhancers (R3 and DHS-12), transcripts extending from the remaining enhancers (R2 and R4) to the poly(A) addition site of the *Nprl3* gene are still detected (Figures S3B and S3C). Normalized RNA-Seq data show that these two enhancer isoforms (in total) represent ~20% of the poly(A)$^+$ RNA from the intact *Nprl3* gene (Figure 3B), consistent with previously established qPCR data (Figure 1F). None of these transcripts produced proteins that could be detected by western blot analysis (Figures S3D and S3E).

Together these data show that the intragenic enhancers (R2, R3, and DHS-12) and an upstream intergenic enhancer (R4) act as alternative erythroid-specific promoters. Significant levels of divergent transcription from these enhancers occur independently of the *Nprl3* promoter and resemble that seen from canonical promoters.

Genome-wide Identification of Intragenic Enhancers

Having identified alternative poly(A)$^+$ RNA isoforms with start sites coinciding with the α-globin enhancers, we next determined whether this phenomenon occurs at other intragenic enhancers throughout the genome. To identify enhancers and to facilitate their correlation with RNA-Seq data, we used previously defined chromatin signatures that distinguish enhancers from promoters. Enhancers are marked by DHSs with a high level of H3K4me1 and low levels of H3K4me3 (Heintzman et al., 2007, 2009), whereas promoters are associated with DHSs marked by high levels of H3K4me3 and low levels of H3K4me1. By considering only those elements that unequivocally display the H3K4me1 high and H3K4me3 low signature (Figure 4A), we identified 3,358 erythroid enhancers, of which 1,794 lie within gene bodies (~54%). TSSs based on current gene annotation were used to define promoters independently of their chromatin marks. The composite profiles (including RNAP2, H3K4me3, and H3K4me1) comparing predicted enhancers and annotated promoters (Figures 4B and 4C) showed that these two classes of elements have different chromatin and transcription factor signatures.

To ensure that these selection criteria had identified bona fide enhancers, we performed further analysis to confirm that these elements were clearly distinguished from canonical

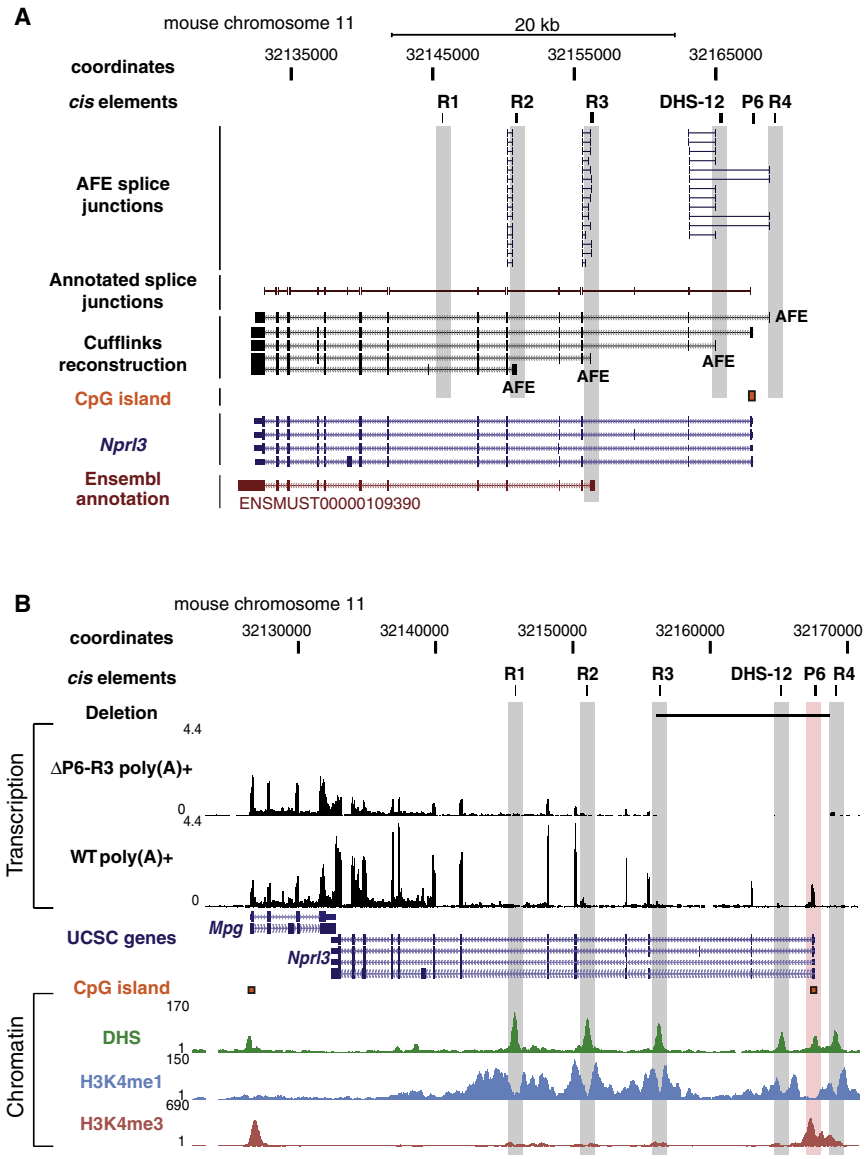

Figure 3. Long Poly(A)$^+$ RNAs within the *Nprl3* Locus in WT and Mutant, *ΔP6-R3*, Erythroid Cells

(A) Overview of poly(A)$^+$ transcription within the *Nprl3* locus in WT erythroid cells. Spliced reads from RNA-Seq data are displayed split into two classes: reads associated with AFE splice junctions and reads associated with annotated splice junctions. Cufflinks transcript reconstruction (black) compared to the UCSC Gene annotation (purple). The Cufflinks isoform associated with enhancer R3 was found in the Ensembl genome annotation. The enhancers (gray columns) and the *Nprl3* CpG island are shown.

(B) The mouse *Nprl3* locus with high-resolution maps of normalized poly(A)$^+$ RNA-Seq data for WT and *ΔP6-R3* erythroid cells (top two tracks). The y axes represent fragments per base pairs per million reads aligned. The enhancers and the *Nprl3* promoter are highlighted as gray and red columns, respectively. UCSC Gene annotation is shown in purple, and UCSC CpG Island annotation is shown as orange boxes. High-resolution maps for *DnaseI* hypersensitivity, H3K4me1, and H3K4me3 are shown below. The y axes represent read density.

promoters. This showed that sequences in the enhancer set are predominantly bound by tissue-specific transcription factors (in this case, Gata1, Scl, Klf1, and Ldb1) and the cofactor p300; the associated chromatin is also modified by H3K27ac (Figure 4A). These observations are consistent with previous criteria used to identify active enhancers (Creyghton et al., 2010; Heintzman et al., 2007, 2009; Rada-Iglesias et al., 2011). Further bioinformatic analysis clearly showed characteristic differences in DNA sequence and coding potential between the enhancer and promoter sets (Figures S4A–S4E).

To extend the analysis to another species and a well-characterized nonerythroid cell type, we used the same approach to identify enhancers using previously published data from primary fetal human lung fibroblasts (Bernstein et al., 2010). In these cells, 21,374 intragenic and 17,664 intergenic enhancers were identified (Figure S5A).

Many Intragenic Enhancers Produce Poly(A)$^-$ eRNAs

Genome-wide analysis of the poly(A)$^-$ RNAs in erythroid cells showed variable levels of antisense transcription originating from intragenic enhancers (Figure 4D), and similar results were obtained from human primary lung fibroblasts using nascent transcription (global run-on, GRO) (Core et al., 2008) (Figure S5E). In erythroid cells, 876 (58%) intragenic enhancers were transcribed at detectable levels (Figures 4D and 4E), and similarly, in human fibroblasts transcription could be detected at 8,775 (41%) intragenic enhancers (Figures S5E and S5F). Since intergenic enhancers are transcribed in both directions (Kim et al., 2010), we also analyzed sense transcription for intragenic enhancers. Sense poly(A)$^-$ transcription from the intragenic enhancers, normally masked by transcription of the host gene, was weakly seen in erythroid (mouse) (Figure 4F) and more clearly in nonerythroid (human) cells (Figure S5G). These findings show that many intragenic enhancers are transcribed in both sense and antisense directions into short poly(A)$^-$ eRNA. Of note, the differences in the levels of eRNA expression are reflected by different degrees of activating histone modification marks (Figure 4G and Figure S5H).

Many Intragenic Enhancers Contribute to a Class of Full-Length Poly(A)$^+$ RNAs

Based on the observations made at the *Nprl3* locus, we next determined whether intragenic enhancers might frequently

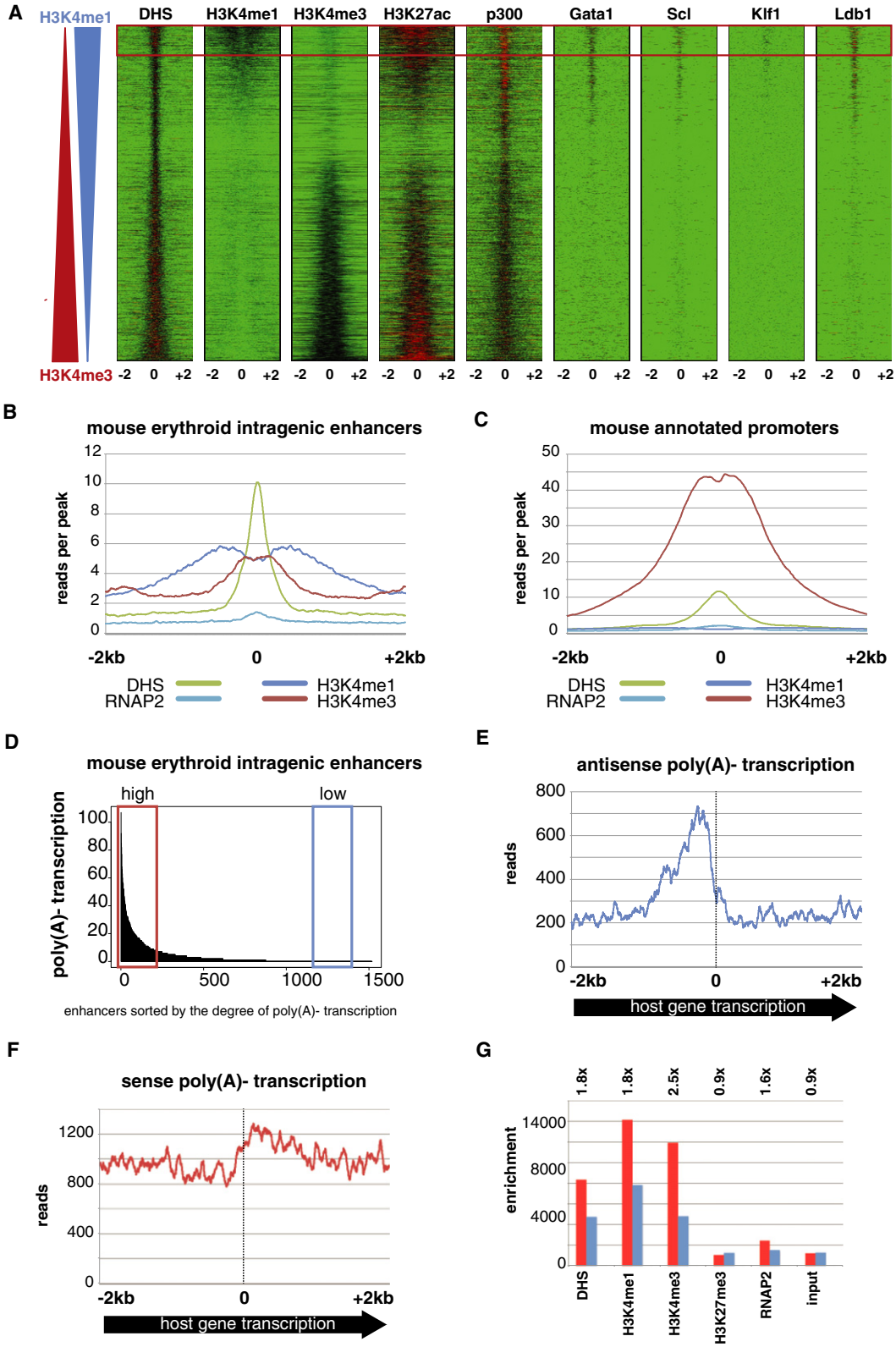

Figure 4. Genome-wide Identification of Enhancers in Erythroid Cells

(A) All detected mouse erythroid DHS sites were sorted based on the difference in enrichment of H3K4me1 and H3K4me3. The same sort order was used for all panels displayed here, and levels of H3K4me3 and H3K4me1 are depicted as red and blue triangles, respectively. Analysis of the *DnaseI* hypersensitivity,

generate alternative poly(A)$^+$ isoforms. Within the body of active genes, it would only be possible to detect enhancer driven isoforms which produced AFEs. Hence we developed a stringent pipeline to identify previously unannotated AFEs, and as expected, it detected all the confirmed AFEs within the *Nprl3* gene. In erythroid cells, we detected 176 enhancers producing AFE transcripts, analysis of which showed they use conventional splice signals. We verified 13 junctions between the identified AFEs and the appropriate exon of the host gene by RT-PCR (Figures S3A and S6).

We next analyzed the distribution of all identified erythroid AFEs around the intragenic enhancers (Figure 5A). This showed a highly significant enrichment of AFEs within 1 kb downstream (but not upstream) of the midpoint of enhancers. These findings show that although many intragenic enhancers are transcribed from both sense and antisense strands of DNA, RNA from the sense strand is transcribed and spliced to produce long poly(A)$^+$ mRNA transcripts. We refer to this class of RNA transcripts as multiexonic enhancer RNAs (meRNAs). Therefore, as in the *Nprl3* gene, a large proportion of enhancers throughout the genome behave as alternative intragenic promoters, although they retain the chromatin signature of an enhancer rather than a promoter.

Enhancer-Driven Poly(A)$^+$ RNAs Are Abundant Full-Length Transcripts

By deleting the promoter of the *Nprl3* gene, we showed that the intragenic enhancers act as promoters independently of the canonical *Nprl3* promoter. This principle is also clearly exemplified by other genes whose canonical promoters are inactive in erythroid cells, unmasking transcription from intragenic erythroid enhancers (e.g., see *D18Ertd653e* [Figure 5B], *Acmsd* [Figure 5C], *Abat* [Figure 5D], and *Tg* [Figure 6A]). These examples also allowed us to assess the levels of expression from these enhancers by comparing the expression of enhancer-driven RNA to expression of the closest neighboring gene. This revealed that the amount of meRNA produced is as variable as that produced from canonical promoters. For example, at the mouse *D18Ertd653e* and *Acmsd* loci, meRNA isoforms are expressed at levels comparable to their neighboring genes, *4933403F05Rik* and *Ccnt2*, respectively (Figures 5B and 5C). A similar meRNA isoform in the *Abat* locus is expressed at 40% relative to that of *Tmem186* gene (Figure 5D). By deleting the canonical promoter of the *Nprl3* gene, we have shown that meRNAs may account for up to 50% of the mRNA derived from this gene in erythroid cells (Figure 1C).

Having identified these meRNAs by RNA-Seq, we have also identified intact meRNAs by northern blotting (e.g., *Tg* [Figures 6A and 6B] and *Znfx1* [Figures 6C and 6D]), confirming that they are abundant full-length transcripts extending from the intragenic enhancers to the poly(A) addition site of the host gene. The levels of meRNA expression determined by northern blots correspond to those determined by RNA-Seq. We have therefore shown that intragenic enhancers may act as TSSs, producing abundant, full-length, polyadenylated transcripts. meRNA transcripts arising from enhancers (as opposed to canonical promoters) represent an important class of RNA.

meRNAs Add to the Complexity of the Transcriptome

With this understanding that abundant, full-length, polyadenylated transcripts may originate both from promoters and enhancers, we reviewed the current Refseq annotation. We noted that some enhancer-driven transcripts have already been annotated as isoforms of their associated protein-coding genes. For example, *D18Ertd653e* has two isoforms, one from its canonical promoter and one from an erythroid-specific enhancer (Figure 7A). Surprisingly, in erythroid cells, we found 139 "active TSSs" with the chromatin signature of an enhancer rather than a promoter (Figure 7B). Similar analysis of primary human lung fibroblasts (Figure 7C), K562 (erythroleukemia), and GM12878 (B cell) showed that between 1% and 7.5% of active TSSs correspond to enhancers. Although the number of TSSs with an enhancer signature varies from tissue to tissue, a relatively small fraction of TSSs overlap between tissues; most are specific to each cell type (Figure 7D).

The complexity of meRNAs in the transcriptome will be much greater than we have detected in this study. We could only identify meRNAs via their association with AFEs. Many enhancers may initiate meRNAs by transcribing an existing exon, rather than an AFE. This principle is illustrated by the meRNAs derived from the *Abat* gene in erythroid cells (in which the canonical promoter is inactive) (Figure S7). Given that there may be many more enhancers than protein-coding genes determining cell identity (Bulger and Groudine, 2010; Heintzman et al., 2009) in hundreds of different cell types, enhancer-driven meRNAs will account for a substantial degree of the complexity and abundance of poly(A)$^+$ RNA in the transcriptome.

DISCUSSION

Enhancers and promoters are currently considered as two distinct classes of functional *cis* elements. Promoters are

H3K4me1, H3K4me3, H3K27ac, p300, Gata1, Scl, Klf1, and Ldb1 data segregate the DHS sites into H3K4me1 (enhancer) and H3K4me3 (promoter) enriched populations. Most erythroid-specific transcription factors are bound to enhancers. The red rectangle indicates the cut-off used for enhancers. Each panel shows the distribution of signal in a 4 kb window centered in the middle of each DHS.

(B and C) Chromatin profiles normalized for number of peaks for the mouse erythroid intragenic enhancers are shown in (B), and all annotated mouse TSSs (UCSC Genes) is shown in (C). Color coding for each chromatin mark is shown below.

(D) Mouse erythroid enhancers were sorted based on the level of antisense poly(A)$^-$ transcription. Red and blue rectangles contain high and low transcribing enhancers, respectively.

(E and F) The cumulative poly(A)$^-$ transcription associated with intragenic enhancers in the antisense (E) and sense (F) direction is seen to originate close to the midpoint of the enhancers and extend ~1 kb in the antisense direction in mouse erythroid cells. The poly(A)$^-$ transcription in the sense direction is masked by the host gene transcription (see relatively higher background in F in comparison to E).

(G) The comparison between high (red) and low (blue) transcribing enhancers is displayed as enrichment of various factors. The high and low transcribing enhancers are as indicated in (D). The fold difference for each factor is indicated above the graphs.

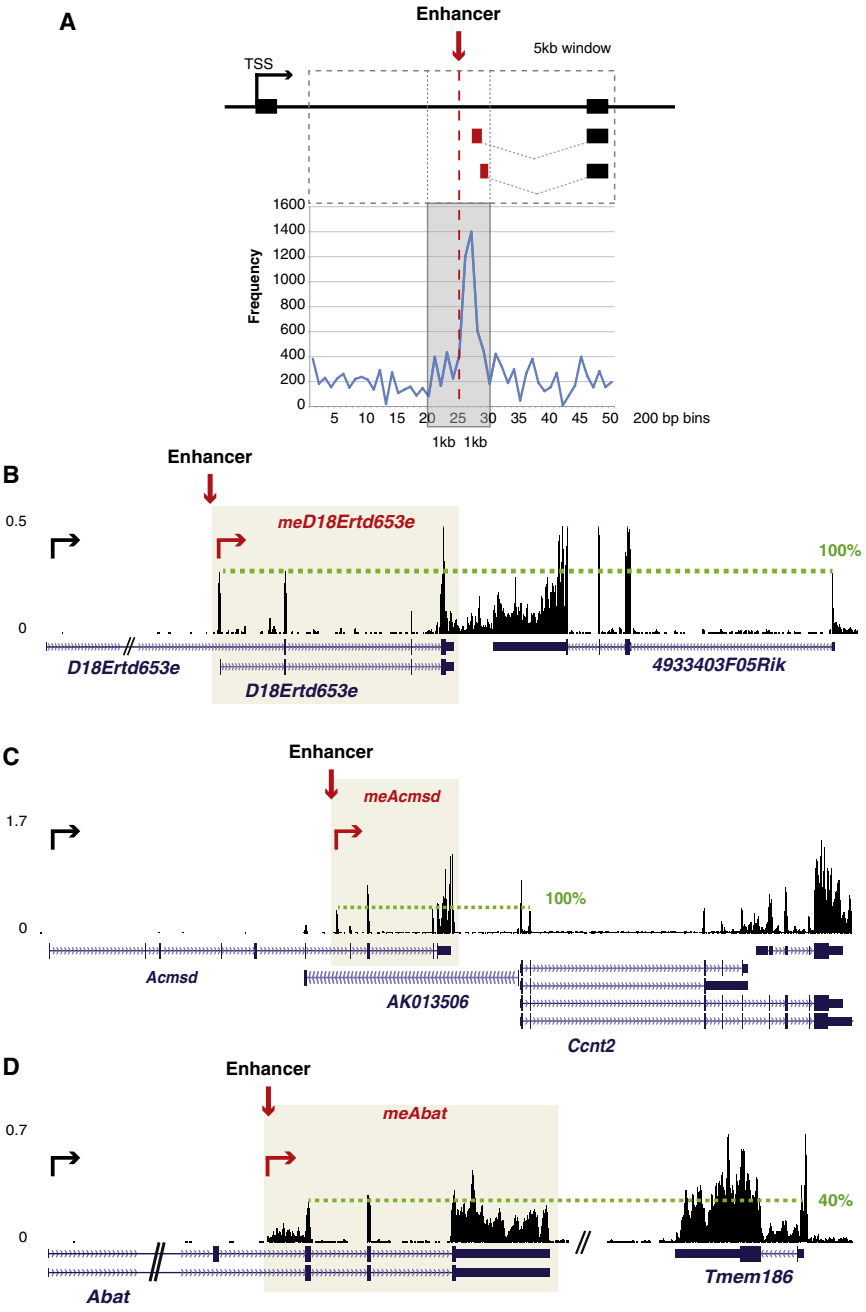

Figure 5. AFEs Are Associated with Enhancers, and meRNAs Are Expressed at Levels Similar to Those of Neighboring Protein-Coding Genes

(A) The idealized structure of a gene containing an enhancer is shown. The TSS is shown as a black arrow. The midpoint of the enhancer is represented as a red arrow and aligned with other data. mRNA forms are represented below the gene; the AFE is shown in red. Black boxes denote exons and white boxes introns, and splicing of mature transcripts is shown by dashed lines. The frequency of AFEs is shown in 200 bp bins relative to the midpoint of intragenic enhancers over a 10 kb window. (B)–(D) give examples of meRNAs that are expressed from within genes with inactive canonical promoters (*D18Ertd653e*, *Acmsd*, *Abat*). In each panel, normalized poly(A)$^+$ RNA-Seq data in WT erythroid cells is displayed at the top and UCSC Gene annotation below. TSSs are indicated as black (promoter) and red arrows (enhancer). The extent of the enhancer transcripts (*meD18Ertd653e* in B, *meAcmsd* in C, and *meAbat* in D) is highlighted within beige rectangles. On each panel a green dashed line connects the AFE of each meRNA to the first exon of the neighboring gene to compare meRNA expression levels to mRNA from protein-coding genes. In (C) the expression of *meAcmsd* is compared to the second exon of *Ccnt2* gene due to the overlapping signal from first exons of *Ccnt2* and *AK013506*.

defined as regions which initiate gene transcription (associated with TSSs), whereas enhancers are distally positioned elements which regulate transcription from canonical promoters temporally and spatially. Although these functional definitions remain correct, data presented here show that bona fide intragenic enhancers frequently behave as alternative promoters (alternative TSS) producing a class of abundant, full-length poly(A)$^+$ mRNA, which we refer to as meRNAs.

Transcription from intragenic enhancers closely resembles that from promoters, since engaged RNAP2 produces short bidirectional poly(A)$^-$ RNA transcripts but also produces abundant, full-length, spliced poly(A)$^+$ transcripts extending in the direction of the host gene (meRNAs). Although they initiate at intragenic enhancers, meRNAs still reflect the host gene's structure using the same splicing and polyadenylation signals, although a proportion use cryptic splice signals within the intron containing the enhancer to produce noncoding AFEs. As for conventional promoters, the mechanism underlying the directionality of meRNAs remains to be determined (Seila et al., 2009).

The existence of this class of meRNAs has been obscured by transcripts driven from the canonical promoters, which overlap meRNAs. Because of this confounding problem, researchers have focused on RNAs originating from intergenic regions (e.g., long intergenic noncoding RNAs or lincRNAs [Guttman et al., 2009; Khalil et al., 2009]). meRNAs that produce an AFE can readily be identified even when the host gene is being transcribed (e.g., *Nprl3* [Figure 3A] and *Znfx1* [Figure 6D]). The expression and processing of meRNAs can, however, be seen most clearly when the canonical promoter has been deleted (e.g., *Nprl3* gene [Figure 3B]) or when the canonical promoter is naturally inactive in erythroid cells (e.g., *D18Ertd653e* [Figure 5B], *Acmsd* [Figure 5C], and *Abat* [Figure 6]). The expression level of both eRNAs and meRNAs

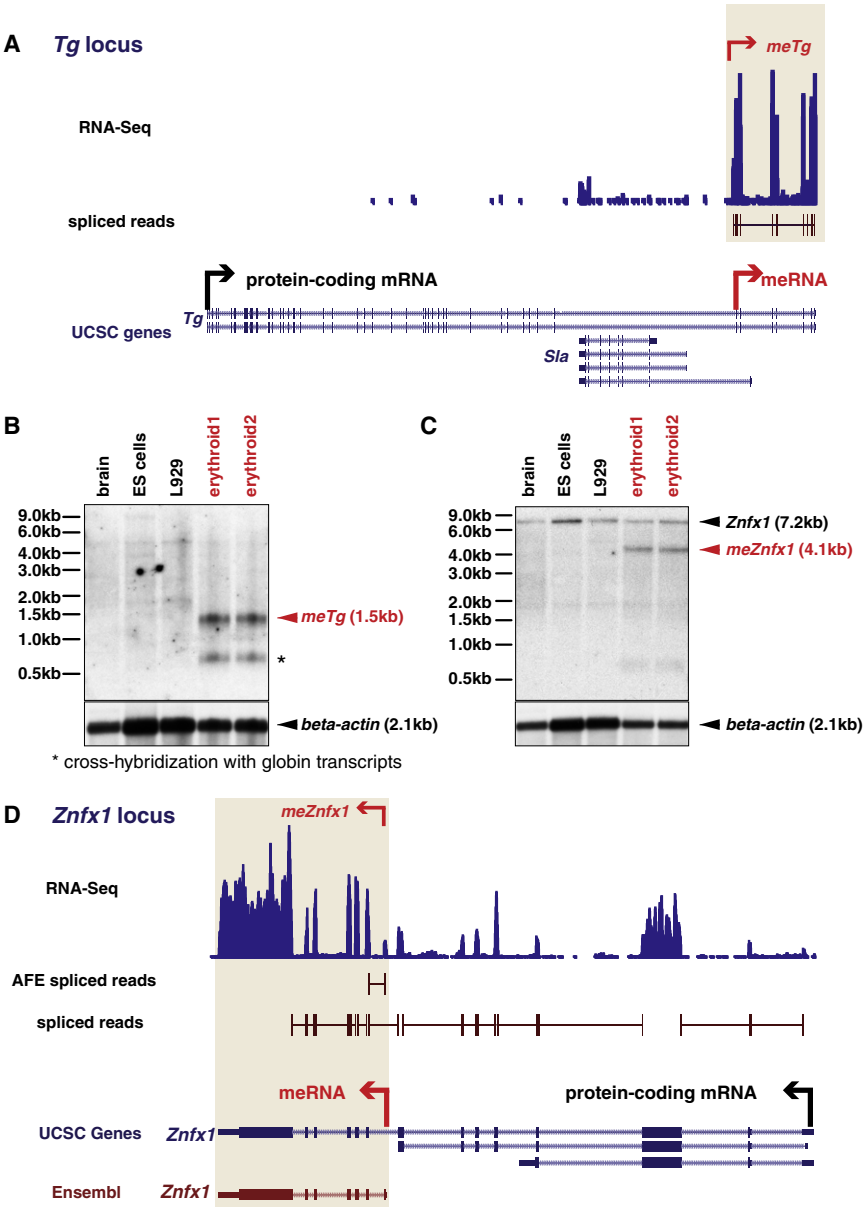

Figure 6. Erythroid Enhancers Give Rise to Full-Length Poly(A)$^+$ RNA

(A) The mouse *Tg* locus with high-resolution maps of normalized poly(A)$^+$ RNA-Seq data in WT erythroid cells. The y axis represents the normalized expression value. Splice reads detected in erythroid cells are displayed below the RNA-Seq track. UCSC Gene annotation is shown. The canonical promoter of the *Tg* gene is inactive in erythroid cells, but black arrow indicates the TSS from this promoter (encoding for protein-coding mRNA). A beige rectangle highlights the extent of the enhancer transcript (*meTg*). The red arrow indicates the TSS of this enhancer (encoding for meRNA).

(B and C) Northern blots show full-length *meTg* RNA (B) and *meZnfx1* RNA (C) of expected sizes. Both meRNAs are present in mouse erythroid cells (two biological replicates) but are absent in non-erythroid cells (brain cells, ESCs, L929 cells). The *Tg* mRNA from the canonical promoter is not expressed in any of the cells tested (B). The full-length *Znfx1* mRNA from the canonical promoter is present in all cells tested (C). β-actin mRNA is shown as a loading control.

(D) The mouse *Znfx1* locus with high-resolution maps of normalized poly(A)$^+$ RNA-Seq data in WT erythroid cells. The y axis represents the normalized expression value. Reads spanning splice junctions containing an unannotated exon (new spliced reads) are displayed separately from spliced reads containing annotated reads, and both are shown below the RNA-Seq track. UCSC Gene annotation is shown. Ensembl annotation shows the alternative *Znfx1* transcript, which we found here to be *meZnfx1*.The canonical promoter of the *Znfx1* gene (black arrow) is active in erythroid cells and to some degree masks transcription from the erythroid enhancer. The beige rectangle highlights the extent of the enhancer transcript (*meZnfx1*). The red arrow indicates the start and the direction of transcription from this enhancer (encoding for meRNA).

from activated enhancers is just as variable as those seen from canonical promoters. In some cases expression of meRNAs may be as great as that directed from canonical promoters of neighboring genes (Figures 5B–5D) and readily detectable on northern blots (Figures 6B and 6C).

An important question is why intragenic enhancers act as alternative promoters. Physical interaction via looping has now been demonstrated for many enhancers and promoters in different cell types (Dekker, 2008; Splinter and de Laat, 2011). It is possible that the juxtaposition between these two classes of *cis* elements, which is associated with high levels of transcription at the canonical promoter, may also promote transcription from the interacting enhancer. When the enhancer is located in intergenic regions, this could produce short poly(A)$^-$ transcripts or perhaps in some cases lincRNAs (Cabili et al., 2011). By contrast, as shown here, an intragenic enhancer will produce a meRNA transcript which is spliced and polyadenylated in a similar manner to the host gene.

What is the biological significance of this class of RNA? There are currently more than ten subclasses of RNA. Although the roles of some RNAs (e.g., mRNA, rRNA, tRNA) are fully or partially established, in many cases (e.g., eRNAs) their role is unclear (Kim et al., 2010). meRNAs resemble isoforms of the host gene, but in the case of *Nprl3*, the meRNAs are not translated into any detectable protein. Global analysis of meRNAs also suggests that, in general, meRNAs have relatively low protein-encoding potential. meRNAs are abundant and complex and show tissue- and developmental-stage specificity. Therefore, rather than simply producing transcriptional noise, it would be surprising if evolution has not used some meRNAs,

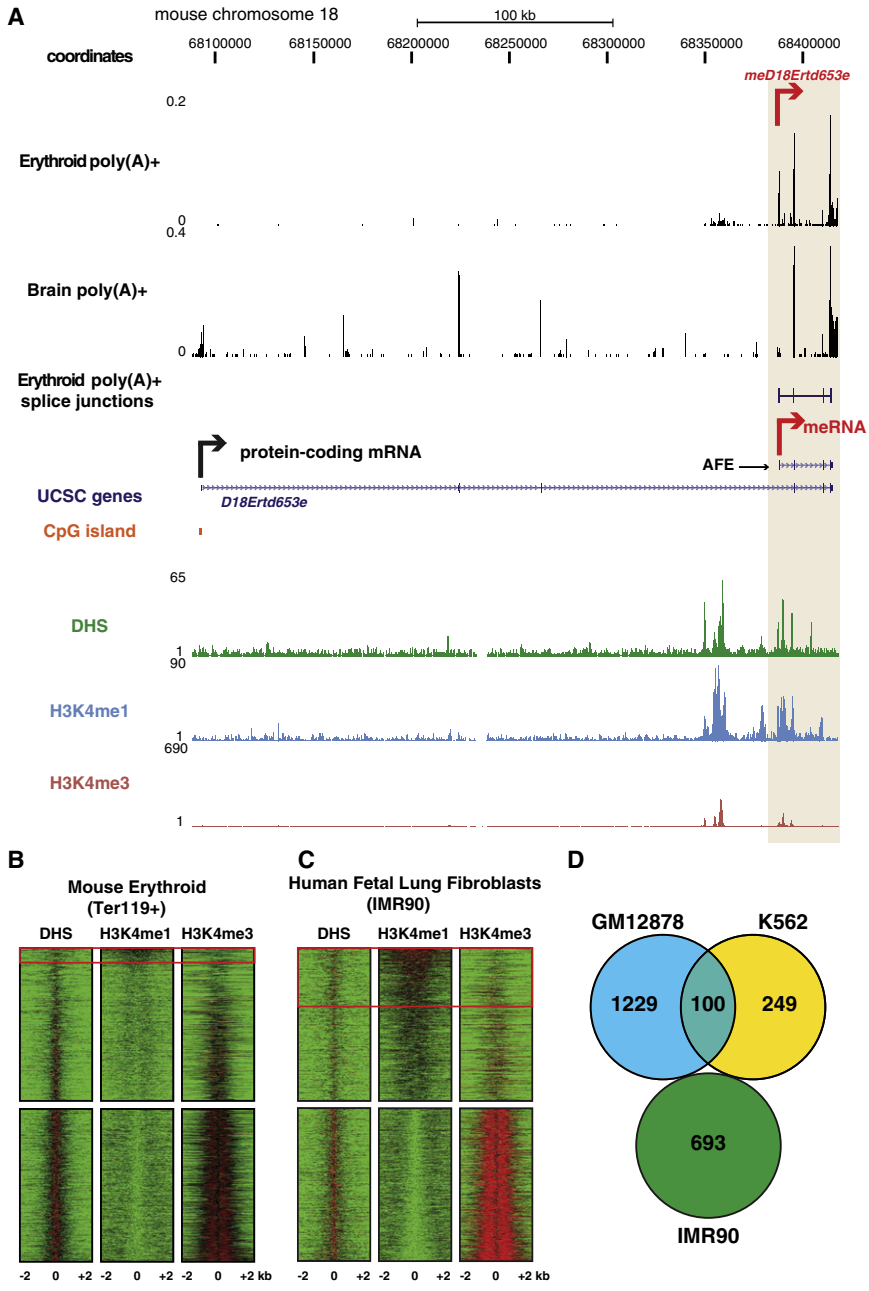

Figure 7. Tissue-Specific Enhancer Transcripts within Current Transcriptome Annotation

(A) The mouse *D18Ertd653e* locus with high-resolution maps of normalized poly(A)$^+$ RNA-Seq data for WT erythroid and brain cells. The y axes represent the normalized expression value. The canonical promoter of the *D18Ertd653e* gene is associated with a CpG island, and a black arrow indicates the start and the direction of transcription from this promoter. Transcription from this promoter is present in the brain but absent in erythroid cells. The beige column highlights the extent of the enhancer transcript (*meD18Ertd653e*). The red arrow indicates the start and the direction of transcription from this enhancer. *meD18Ertd653e* transcript is present in erythroid but absent in brain cells. The spliced reads detected in erythroid cells are displayed below the RNA-Seq tracks. UCSC Gene annotation including the annotated enhancer transcript (the annotated AFE is shown with an arrow) is shown. UCSC CpG Island annotation is shown as an orange box. High-resolution maps for the chromatin mark H3K27me3, *DnaseI* hypersensitivity (DHS), H3K4me1, and H3K4me3 are shown below. The y axes represent read density.

(B) Mouse transcription start sites (from UCSC Genes and Refseq gene annotations), which overlap with a single DHS site in Ter119+ cells (13,506 sites), were sorted based on the difference in enrichment of H3K4me1 and H3K4me3. Analysis of chromatin marks, as in Figure 4A, shows that 139 TSSs active in erythroid cells resemble an enhancer chromatin signature (indicated by the red rectangle).

(C) The human transcription start sites (from UCSC Genes and Refseq gene annotations), which overlap a single DHS site in IMR90 cells (16,508 sites), were sorted based on the difference in enrichment of H3K4me1 and H3K4me3. Analysis of the chromatin marks (as in Figure S5A) shows that 693 TSSs active in IMR90 cells resemble an enhancer chromatin signature (indicated by the red rectangle). In both (B) and (C), the upper panels show the 2,000 most H3K4me1-enriched TSSs, and the lower panels show the 2,000 most H3K4me3-enriched TSSs.

(D) A Venn diagram of the active TSSs in human IMR90, K562, and GM12878 cells with an enhancer chromatin signature shows the highly tissue-specific nature of meRNAs.

or their processed RNA products, for important biological functions which can now be evaluated.

A major task in understanding the transcriptome is to identify the different classes of RNA so that the functional role of each subclass can be determined. Whatever their role, this study shows that meRNAs constitute a complex and abundant class of RNA. Using stringent criteria we identified 1,794 intragenic enhancers in erythroid cells, of which at least 876 express eRNA and 179 produce meRNAs identified via their associated AFEs. Due to the high stringency with which we identified enhancers and the unknown frequency with which enhancer-driven transcripts use the donor splice sites of the host gene rather than produce AFEs (Figure S7), we have greatly underestimated the full contribution that meRNAs make to the transcriptome of erythroid cells. Nevertheless, even using this very stringent analysis, in just one tissue (erythroid cells) we have already demonstrated that 139 annotated, active TSSs correspond to enhancers rather than promoters (Figure 5B).

Taking all cell types into account, it is clear that mammalian genomes contain many more enhancers than promoters (Bulger and Groudine, 2010; Heintzman et al., 2009), and up to 51% of these enhancers are intragenic. Analysis of just three cell types (IMR90, K562, GM12878) in this study showed very little overlap in enhancer-driven meRNAs between cell types. meRNAs may

account for a substantial proportion of transcriptome complexity and how this changes from one cell type to another. This insight will require the field to reannotate the transcriptome.

Clearly, distinguishing enhancer-driven meRNAs from the protein-coding mRNAs of host genes containing the enhancers will discriminate between genes whose expression is increased in a specific cell type from those genes whose expression may simply increase as a result of containing enhancers for other genes (bystander effect) (Cajiao et al., 2004). Identifying meRNAs together with eRNAs may solve a long-standing problem of analyzing when and where enhancers are activated during commitment, differentiation, and development at a single-cell level.

EXPERIMENTAL PROCEDURES

Primary Cells

Mouse primary erythroid cells were sorted from the spleens of acetylphenylhydrazine-treated (to induce acute hemolytic anemia) mice based on the expression of Ter119 antigen (Vernimmen et al., 2009). Mouse erythroid cells were grown from fetal livers (Dolznig et al., 2001).

Deletional Constructs

The ΔP6-deleted segment spans 2,315 bp between coordinates chr11:32,166,133-32,168,448 (mm9); the ΔP6-R3 deleted segment spans 12,403 bp between coordinates chr11:32,156,045-32,168,448 (mm9). Characteristics of deleted elements, embryonic stem cells (ESCs) targeting, screening, and generation of mouse models are described in the Supplemental Experimental Procedures.

RNA Blot

RNA was poly(A) selected with the PolyATract mRNA isolation system (Promega). Northern blots were performed using NorthernMax-Gly Kit (Ambion).

RNA Sequencing

For RNA-Seq library, total RNA was split into poly(A)$^+$ and poly(A)$^-$ RNA using the PolyATract mRNA isolation system (Promega). Poly(A)$^+$ RNA libraries were produced using the Illumina mRNA-Seq paired-end kit after globin depletion using GlobinClear (Ambion). Poly(A)$^-$ RNA libraries were produced using the Illumina DGE Small RNA Sample Prep kit, after depletion of ribosomal transcripts with RiboMinus Eukaryote Kit for RNA sequencing (Invitrogen).

Chromatin Immunoprecipitation and ChIP Sequencing

Chromatin immunoprecipitation (ChIP) was performed as described (De Gobbi et al., 2007). ChIP sequencing (ChIP-Seq) libraries were prepared and sequenced using the standard Illumina protocol.

DnaseI Assay and *DnaseI* Sequencing

Nuclei from primary erythroid cells (Ter119+) were digested with eight increasing concentrations of *DnaseI* (Roche) (Higgs et al., 1990). DNA (1.5 mg) from the mid-phase digestions was blunt-ended with T4 DNA Polymerase (NEB) and prepared for Illumina GAII sequencing. The *DnaseI*-digested material was amplified using PCR primer PE1.0 and PE2.0 (Illumina).

Antibodies

The following antibodies were used for ChIP: anti-H3K4me1, anti-H3K4me3, anti-H3K27me3 (Upstate 07-436, 07-473, 07-449), anti-H3K27ac (Abcam ab4729), and anti-RNA-PolII (Santa Cruz Biotechnology H224). Protein blots were performed using anti-Nprl3 (C16B) (C-terminal antibody) (Lunardi et al., 2009), anti-GST (Santa Cruz Biotechnology sc-138), and anti-Gapdh (Cell Signaling, 3683).

ACCESSION NUMBERS

Sequencing data sets described in this study have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus (GSE27921).

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at doi:10.1016/j.molcel.2011.12.021.

ACKNOWLEDGMENTS

We thank N.J. Proudfoot, V.G. Sankaran, D. Clynes, and S. Hänni for technical advice and for critically reading the manuscript. This work was supported by Marie Curie RTN EUrythron (MRTN-CT-2004-005499) (M.S.K.), Medical Research Council (UK), and the National Institute for Health Research (NIHR) Biomedical Research Centre Oxford, and assisted by the Computational Biology Research Group (CBRG) (Oxford).

Received: May 9, 2011
Revised: October 14, 2011
Accepted: December 6, 2011
Published online: January 19, 2012

REFERENCES

Anguita, E., Sharpe, J.A., Sloane-Stanley, J.A., Tufarelli, C., Higgs, D.R., and Wood, W.G. (2002). Deletion of the mouse alpha-globin regulatory element (HS −26) has an unexpectedly mild phenotype. Blood *100*, 3450–3456.

Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G., and Higgs, D.R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. EMBO J. *23*, 2841–2852.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.

Bulger, M., and Groudine, M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. Dev. Biol. *339*, 250–257.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Cajiao, I., Zhang, A., Yoo, E.J., Cooke, N.E., and Liebhaber, S.A. (2004). Bystander gene activation by a locus control region. EMBO J. *23*, 3854–3863.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the Mamm. genome. Science *309*, 1559–1563.

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.

De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J.A., Sharpe, J.A., Koch, C.M., Dunham, I., Gibbons, R.J., Wood, W.G., and Higgs, D.R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. Blood *110*, 4503–4510.

Dekker, J. (2008). Gene regulation in the third dimension. Science *319*, 1793–1794.

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. 10.1371/journal.pbio.1000384.

Dolznig, H., Boulme, F., Stangl, K., Deiner, E.M., Mikulits, W., Beug, H., and Mullner, E.W. (2001). Establishment of normal, terminally differentiating mouse erythroid progenitors: molecular characterization by cDNA arrays. FASEB J. *15*, 1442–1444.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. Science *322*, 1855–1857.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Higgs, D.R., and Wood, W.G. (2008). Long-range regulation of alpha globin gene expression during erythropoiesis. Curr. Opin. Hematol. *15*, 176–183.

Higgs, D.R., Wood, W.G., Jarman, A.P., Sharpe, J., Lida, J., Pretorius, I.M., and Ayyub, H. (1990). A major positive regulatory region located far upstream of the human alpha-globin gene locus. Genes Dev. *4*, 1588–1601.

Hughes, J.R., Cheng, J.F., Ventress, N., Prabhakar, S., Clark, K., Anguita, E., De Gobbi, M., de Jong, P., Rubin, E., and Higgs, D.R. (2005). Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc. Natl. Acad. Sci. USA *102*, 9830–9835.

Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S., and Gingeras, T.R. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res. *15*, 987–997.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484–1488.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. USA *106*, 11667–11672.

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.

Lower, K.M., Hughes, J.R., De Gobbi, M., Henderson, S., Viprakasit, V., Fisher, C., Goriely, A., Ayyub, H., Sloane-Stanley, J., Vernimmen, D., et al. (2009). Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. Proc. Natl. Acad. Sci. USA *106*, 21771–21776.

Lunardi, A., Chiacchiera, F., D'Este, E., Carotti, M., Dal Ferro, M., Di Minin, G., Del Sal, G., and Collavin, L. (2009). The evolutionary conserved gene C16orf35 encodes a nucleo-cytoplasmic protein that interacts with p73. Biochem. Biophys. Res. Commun. *388*, 428–433.

Mattick, J.S., Taft, R.J., and Faulkner, G.J. (2010). A global view of genomic information—moving beyond the gene and the master regulator. Trends Genet. *26*, 21–28.

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.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

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.

Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell Cycle *8*, 2557–2564.

Splinter, E., and de Laat, W. (2011). The complex transcription regulatory landscape of our genome: control in three dimensions. EMBO J. *30*, 4345–4355.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J. *26*, 2041–2051.

Vernimmen, D., Marques-Kranc, F., Sharpe, J.A., Sloane-Stanley, J.A., Wood, W.G., Wallace, H.A., Smith, A.J., and Higgs, D.R. (2009). Chromosome looping at the human {alpha} globin locus is mediated via the major upstream regulatory element (HS −40). Blood *114*, 4253–4260.