Letter

Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation

Marcel E. Dinger,[1,6] Paulo P. Amaral,[1,6] Tim R. Mercer,[1,6] Ken C. Pang,[1,2] Stephen J. Bruce,[1] Brooke B. Gardiner,[1,3] Marjan E. Askarian-Amiri,[1] Kelin Ru,[1] Giulia Soldà,[1,4] Cas Simons,[1] Susan M. Sunkin,[5] Mark L. Crowe,[1] Sean M. Grimmond,[1,3] Andrew C. Perkins,[1] and John S. Mattick[1,7]

[1]ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia QLD 4072, Australia; [2]Ludwig Institute for Cancer Research, Melbourne Centre for Clinical Sciences, Heidelberg VIC 3084, Australia; [3]Australian Stem Cell Centre, Monash University, Clayton VIC 3800, Australia; [4]Department of Biology and Genetics for Medical Sciences, University of Milan, Milan 20133, Italy; [5]Allen Institute for Brain Science, Seattle, Washington 98103, USA

The transcriptional networks that regulate embryonic stem (ES) cell pluripotency and lineage specification are the subject of considerable attention. To date such studies have focused almost exclusively on protein-coding transcripts. However, recent transcriptome analyses show that the mammalian genome contains thousands of long noncoding RNAs (ncRNAs), many of which appear to be expressed in a developmentally regulated manner. The functions of these remain untested. To identify ncRNAs involved in ES cell biology, we used a custom-designed microarray to examine the expression profiles of mouse ES cells differentiating as embryoid bodies (EBs) over a 16-d time course. We identified 945 ncRNAs expressed during EB differentiation, of which 174 were differentially expressed, many correlating with pluripotency or specific differentiation events. Candidate ncRNAs were identified for further characterization by an integrated examination of expression profiles, genomic context, chromatin state, and promoter analysis. Many ncRNAs showed coordinated expression with genomically associated developmental genes, such as *Dlx1, Dlx4, Gata6*, and *Ecsit*. We examined two novel developmentally regulated ncRNAs, *Evx1as* and *Hoxb5/6as*, which are derived from homeotic loci and share similar expression patterns and localization in mouse embryos with their associated protein-coding genes. Using chromatin immunoprecipitation, we provide evidence that both ncRNAs are associated with trimethylated H3K4 histones and histone methyltransferase MLL1, suggesting a role in epigenetic regulation of homeotic loci during ES cell differentiation. Taken together, our data indicate that long ncRNAs are likely to be important in processes directing pluripotency and alternative differentiation programs, in some cases through engagement of the epigenetic machinery.

[Supplemental material is available online at www.genome.org. The custom microarray design and microarray expression from this study have been submitted to ArrayExpress under accession nos. A-MEXP-1070 and E-TABM-433, repectively.]

Embryonic stem (ES) cells are immortal cells capable of producing most adult-type lineage-specific cells in vitro (Evans and Kaufman 1981; Martin 1981). As well as offering a model of early development, ES cells have considerable potential to repair and regenerate damaged or genetically defective adult organs (Prelle et al. 2002). Although many recent studies have identified transcription factor networks and epigenetic processes that are fundamental to the maintenance of pluripotency (the stage where cells have the potential to differentiate into any germ layer) and differentiation of ES cells (Boyer et al. 2005, 2006; Wang et al. 2006; Schulz and Hoffmann 2007), the molecular basis for the generation, selection, and behavior of particular stem cell types remains unclear.

Nonprotein-coding RNAs (or noncoding RNAs [ncRNAs]) participate in many processes that coordinate gene expression,

particularly during development (Mattick 2007). MicroRNAs (miRNAs) are *trans*-acting regulatory RNAs that act by regulating the translation or directing the degradation of specific mRNA targets and are known to play central roles in many aspects of development, including ES cell pluripotency (Houbaviy et al. 2003; Lakshmipathy et al. 2007), hematopoiesis (Garzon et al. 2006), CNS development (Schratt et al. 2006), and embryogenesis (Wienholds et al. 2005; Zhao et al. 2005; Mineno et al. 2006). Other, less well characterized classes of ncRNAs are also implicated in developmental regulation and disease in vertebrates (Mattick and Makunin 2005). In particular an increasing number of individually characterized long ncRNAs (>200 nucleotides [nt]) such as *Xist* (Okamoto et al. 2005), *TUG1* (Young et al. 2005), *PINC* (Ginger et al. 2006), *Evf2* (Feng et al. 2006), and *HOTAIR* (Rinn et al. 2007) have important developmental roles.

Long ncRNA transcription is prevalent throughout the mammalian genome (Engstrom et al. 2006), and transcriptomic studies in mouse show that the number of distinct long ncRNAs is comparable to that of mRNAs (Carninci et al. 2005). The limited number of functional studies of long ncRNAs reveal that

they act via a diverse range of mechanisms in many regulatory processes, including transcription (Feng et al. 2006), splicing (Yan et al. 2005), translation (Wang et al. 2005), nuclear factor trafficking (Willingham et al. 2005), imprinting (Sleutels et al. 2002; Thakur et al. 2004), genome rearrangement (Nowacki et al. 2007), and chromatin modification (Bernstein and Allis 2005; Rinn et al. 2007). Comparative analysis of mouse long ncRNAs indicates that their promoters, primary sequence, and splice sites are under purifying selection (Ponjavic et al. 2007). Given the tissue- and cell-type specific (Kapranov et al. 2007; Nakaya et al. 2007; Mercer et al. 2008) and dynamically regulated expression (Ravasi et al. 2006) of long ncRNAs, it appears likely that many more of the vast numbers of mammalian long ncRNAs are intrinsically functional.

In light of the diversity and abundance of long ncRNAs, the functional characterization of this transcript class is a considerable challenge, and functional screens using cell-based assays have met with limited success (Willingham et al. 2005). Unlike protein-coding genes where sequence motifs are usually indicative of function, at least in the biochemical sense, ncRNA sequence information is currently uninformative for predicting function. However, many long ncRNAs have been found to originate from complex transcriptional loci, in which the ncRNAs are coordinately transcribed with their associated protein-coding transcripts (Engstrom et al. 2006), and several recent examples of characterized ncRNAs, such as *Evf2* (Feng et al. 2006), *HOTAIR* (Rinn et al. 2007), *Kcnq1ot1* (Thakur et al. 2004), and *Air* (Sleutels et al. 2002), support a functional relationship between the ncRNA and the associated or related protein-coding gene(s). Therefore, by examining the genomic context of ncRNAs relative to protein-coding genes of known function, in conjunction with expression data, it may be possible to predict a related role for the associated nonprotein-coding transcript.

In this paper, we describe the developmentally regulated expression of hundreds of long ncRNAs during the differentiation of mouse ES cells. By examining the genomic context in combination with their expression profiles, we identify candidates likely to have roles in pluripotency and differentiation. To further understand the potential roles of these transcripts, we characterized two novel ncRNAs and find evidence of their association with chromatin and chromatin-modifying factors. Our data suggest that long ncRNAs are likely to play an important role in the regulation of both pluripotency and lineage commitment and therefore need to be considered to further understand these fundamental biological processes.

Results

Expression profiling of ncRNAs during EB differentation

To examine the expression profiles of noncoding and protein-coding RNAs during mouse ES cell differentiation, we interrogated a custom microarray with RNA isolated at 11 time points from differentiating embryoid bodies (EBs) over a 16-d period (see Methods; Table 1). Consistent with previous reports (Zambrowicz et al. 1998; Ramalho-Santos et al. 2002), we found that 58% of protein-coding transcripts were expressed above background (see Methods) during EB differentiation and 24% (2103 out of 8625) of these were significantly differentially expressed (B-statistics > 3; fold-change > 2) between one or more time points. From the ncRNA subset, we found that 26% were expressed above background and 18% (174 out of 945) of these

Table 1. Summary of microarray expression results

Probe class	Total	Expressed above background	Differentially expressed[a]
Coding	14,827	8,625 (58%)	2,103 (24%)
Noncoding	3,659	945 (26%)	174 (18%)
Combined	18,486	9,570 (52%)	2,277 (24%)

[a]Significant differential expression was defined as probes with B-statistics > 3 and fold-change > 2.

were significantly differentially expressed (Supplemental Table S1). It should be noted that some known ncRNAs, such as *Evf2* (Feng et al. 2006), were not detected above the conservative background cutoff levels used in this analysis even though the presence and differential expression of *Evf2* were shown by quantitative real-time PCR (qRT-PCR; see below). Similar to previous observations (Ravasi et al. 2006; Nakaya et al. 2007; Mercer et al. 2008), the mean expression intensity of ncRNAs was lower than for mRNAs (Supplemental Fig. S1A).

The number of transcripts (both ncRNAs and mRNAs) that exhibit differential expression was not uniform across the entire 16-d time course during EB differentiation. Almost twice as many ncRNAs and mRNAs were expressed during the first two days of EB differentiation than at any subsequent time points (Supplemental Fig. S1B). This is also consistent with previous observations of a greater abundance of expression during pluripotent stages of EB differentiation (Ivanova et al. 2002; Ramalho-Santos et al. 2002; Bruce et al. 2007a). The high rate of differential expression in the first two days may reflect dramatic changes in genetic programs associated with pluripotency and the initial specification of differentiation trajectories. The similarity of these trends between ncRNAs and mRNAs suggest they are subject to similar modes of regulation during EB differentiation, and that they both participate actively in these processes.

Given the generality of EB differentiation programs (Keller 1995; Smith 2001), we may expect ncRNAs expressed during this formative stage to be evolutionarily conserved within vertebrates. A previous study defined highly conserved regions, termed phastCons elements, in vertebrate genomes (Siepel et al. 2005). We found that ncRNAs expressed during EB differentiation were enriched for phastCons elements (twofold) relative to the genome average, and this enrichment was even more pronounced in differentially expressed ncRNAs (3.2-fold). It has been previously shown that phastCons elements in noncoding sequences are enriched for predicted RNA secondary structures (Siepel et al. 2005). Therefore, we used genome-wide maps of RNAz-predicted conserved RNA secondary structures (Washietl et al. 2005a; Mercer et al. 2008) to identify expressed ncRNAs that contained conserved RNA secondary structures (see Methods). We found that 29% (267 out of 945) of expressed transcripts contained conserved RNA secondary structures ($P > 0.5$), which were slightly enriched in differentially expressed transcripts (1.5-fold; Supplemental Table S1). The enrichment of conserved elements and predicted secondary structures supports a functional role for these ncRNAs during EB differentiation.

Identification of ncRNAs that are coregulated in pluripotency, primitive streak formation, and mesoderm differentiation

Because various differentiation programs occur simultaneously during EB cell differentiation (Smith 2001), standard clustering

approaches are limited in their ability to discriminate distinct groups of expression patterns. Therefore, to identify ncRNAs associated with specific ES cell differentiation processes, we searched for dynamic expression patterns that closely resemble those of well-characterized marker genes whose expression can be used to demarcate different stages of EB differentiation (Bruce et al. 2007a). Using this approach, it is possible to define three distinct classes of expression profiles, which can roughly be considered to correspond to the EB differentiation phases of pluripotency, primitive streak formation, and mesoderm differentiation. Although these distinctions are somewhat arbitrary, they are nonetheless representative of significant biological transitions that occur during EB differentiation. We identified ncRNAs associated with each of these classes (Fig. 1A), which are discussed in further detail below.

Pluripotency

Pluripotency is governed by a few key genes that are highly expressed in undifferentiated ES cells and rapidly down-regulated upon differentiation (Bruce et al. 2007a). Approximately 200 protein-coding genes characterize this stem cell state (Ivanova et al. 2002; Ramalho-Santos et al. 2002) and many have established functions in maintenance of pluripotency (Ivanova et al. 2006). To identify putative ncRNAs involved in pluripotency, we looked for ncRNA expression profiles that correlated with the expression profiles of *Pou5f1* (Nichols et al. 1998), *Nanog* (Chambers et al. 2003), and *Sox2* (Loh et al. 2006), which are core components of the transcriptional network for maintaining pluripotency (Kim et

al. 2008). We identified 12 ncRNAs (Fig. 1B), many of which are genomically associated with protein-coding genes that exhibited expression profiles correlated to *Pou5f1* or *Sox2* expression (Pearson's correlation coefficient, $R^2 > 0.9$; Supplemental Fig. S2A). In addition, promoters of two of these ncRNAs are occupied by *Pou5f1* and *Nanog* in ES cells according to a recent genome-wide chromatin immunoprecipitation (ChIP) study (Boyer et al. 2005), indicating that ncRNA expression may be regulated by these transcription factors.

Primitive streak formation

At the onset of gastrulation, epiblast cells from the primitive streak undergo a mesenchymal transition, giving rise to the mesoderm and definitive endoderm. This process is reproduced during ES cell differentiation into EBs and is marked by the expression of genes such as *Evx1* and brachyury (*T*) (Fig. 1C) (Robertson et al. 2000; Hirst et al. 2006; Bruce et al. 2007a). Brachury is a well-characterized specific marker of the primitive streak and is critical in the formation and organization of the early stage mesoderm (Wilkinson et al. 1990). We identified seven ncRNAs that exhibited a highly correlated ($R^2 > 0.9$) expression profile with *T* and *Evx1* (Supplemental Fig. S2B), indicating that these ncRNAs may also participate in gastrulation.

Mesoderm differentiation

After 6 d of EB differentiation, various mesodermal tissues begin to develop in parallel, and alternative outcomes can be generated
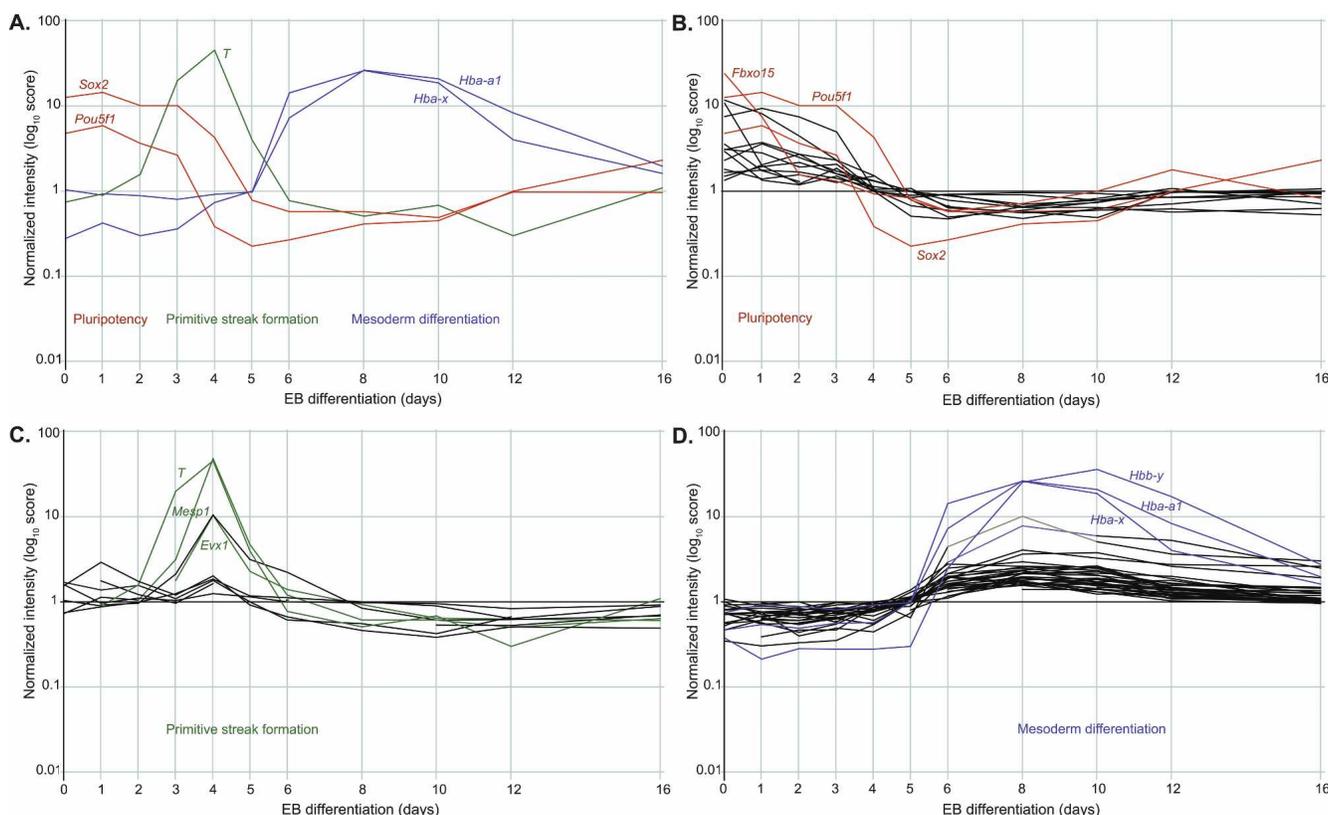

Figure 1. Correlation of expression profiles of ncRNAs with protein-coding gene markers during EB differentiation. Genes with well-characterized roles in EB differentiation (*A*) were used to identify ncRNAs with correlated expression profiles (Pearson's coefficient > 0.9) in pluripotency (*B*; red lines; *Sox2*, *Pou5f1*), primitive streak formation (*C*; green lines; *Evx1*, *T*), and mesoderm differentiation along the hematopoietic lineage (*D*; *Hba-a1*, *Hba-x*). Expression was detected by microarray from 11 RNA samples isolated from differentiating EBs over a 16-d period.

using defined growth factors. In the presence of serum or the growth factor BMP4, ventral mesoderm-derived tissues such as blood, vasculature, and cardiac muscle cells are efficiently generated (Bruce et al. 2007a). Here, we focused on hematopoiesis because it is a very well defined developmental program, which is marked by the expression of hemoglobin genes together with a number of other important regulators (Mikkola and Orkin 2006). Visceral and definitive endoderm is also generated at this time, and endoderm-specific transcription factors such as *Gata6* provide a useful signature of this program (Molkentin 2000; Kapranov et al. 2005). We identified 31 ncRNAs that exhibit a highly correlated ($R^2 > 0.9$) expression profile with hemoglobin genes (Fig. 1D). The majority of these ncRNAs were associated with protein-coding genes, of which 14 have previously defined roles in mesoderm differentiation (Supplemental Fig. S2C). As an alternative means to identify ncRNA with potential roles in mesoderm differentiation, we identified 36 ncRNAs that exhibit strongly negative correlation of expression ($R^2 < -0.9$) with the pluripotency-associated transcription factor *Pou5f1* (Supplemental Fig. S2D).

Genomic association of ncRNAs with protein-coding genes

As ncRNAs often originate from complex transcriptional loci, in which the ncRNAs are coordinately transcribed with their associated protein-coding transcripts (Engstrom et al. 2006), analysis of the genomic context of those ncRNAs that were expressed during ES differentiation could help in predicting their func-

tional role. Therefore, we analyzed the genomic context of ncRNAs that were expressed during ES differentiation to identify putative functional relationships between noncoding and protein-coding transcripts.

We categorized the relationship between ncRNAs and their associated protein-coding genes as *cis*-antisense, intronic, or bidirectional (Fig. 2; see Methods). Of 945 ncRNAs expressed during ES differentiation, we identified 338 intronic, 61 bidirectional, and 36 *cis*-antisense ncRNAs (Supplemental Table S2). Next, we analyzed the correlation of the expression between the protein-coding and associated noncoding transcripts. We found that both intronic and bidirectional ncRNAs showed a significant tendency ($P < 0.0001$ and $P < 0.07$; Mann-Whitney test) to having positively correlated expression profiles with their associated protein-coding gene (Fig. 2A,B), consistent with previous results for these classes of ncRNAs (Engstrom et al. 2006; Nakaya et al. 2007). We identified 73 correlated pairs amongst the intronic-associated ncRNAs and nine amongst the bidirectional-associated ncRNAs (Supplemental Table S2).

Individual inspection showed that many of the ncRNAs were associated with protein-coding genes with well-characterized roles in ES biology. For example, we identified a ncRNA (GenBank accession no. AK017619) within the intron of *Dab2* (Fig. 2D), a gene with important functions in the formation of the primitive endoderm layer during mouse embryogenesis (Yang et al. 2007). The expression profiles of *Dab2* and the intronic ncRNA were correlated ($R^2 = 0.77$), suggesting some relationship in their function and/or regulation. In another example,
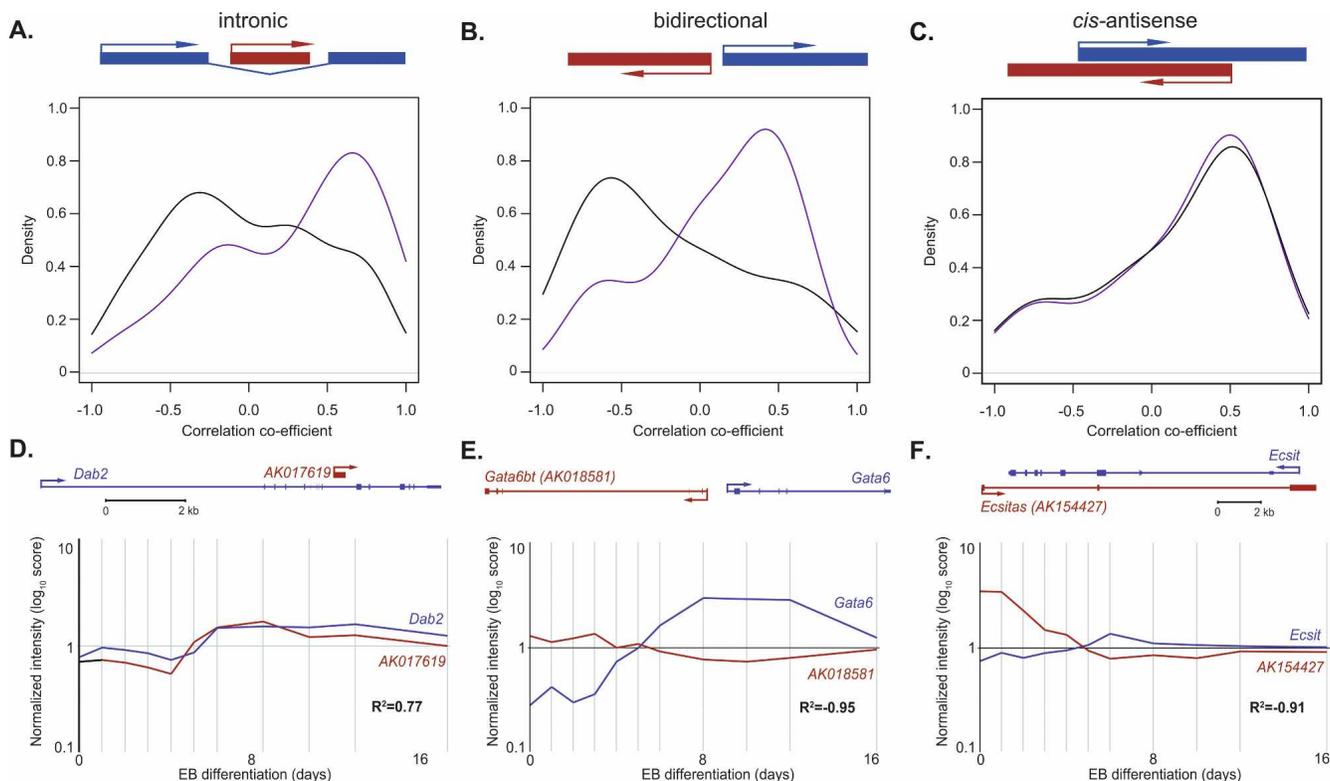

Figure 2. Correlation of expression between ncRNAs and associated protein-coding genes. (*A–C*) Density plots of correlation coefficients between the expression of ncRNAs and their associated intronic (*A*), bidirectional (*B*), or *cis*-antisense (*C*) protein-coding gene (purple line) and randomized pairs (black line). (*D–F*) Examples of correlated (positive or negative) expression between ncRNAs (red) and the associated protein-coding genes (blue). The *upper* panel shows the genomic region of the ncRNA (GenBank accession nos. indicated) and protein-coding gene; the *lower* panel shows the corresponding expression profiles with Pearson's correlation coefficients (R^2) as indicated. Arrows indicate the direction of transcription.

we identified a ncRNA (GenBank Accession AK018581) organized bidirectionally to the gene encoding *Gata6*, which we termed *Gata6bt* (*Gata6* *bidirectional* *transcript*). The expression of *Gata6bt* was negatively correlated with *Gata6*, being up-regulated during pluripotent EB stages and down-regulated during progressive lineage specification (Fig. 2E). *Gata6bt* may have a direct silencing effect on *Gata6*, perhaps via a mechanism similar to that recently described for *CDKN2B* (also known as *p15*) and the associated antisense ncRNA, *p15AS*, which involves epigenetic modifications (Yu et al. 2008).

Although we did not observe any significant general correlation between the expression of *cis*-antisense pairs of protein-coding and noncoding transcripts (Fig. 2C), we did identify nine examples of *cis*-antisense pairs with either positive or negative correlated expression ($R^2 > 0.5$ or $R^2 < -0.5$). For example, a *cis*-antisense ncRNA (GenBank accession no. AK154427) exhibits a negative expression correlation ($R^2 = -0.92$) with its sense protein-coding gene *Ecsit*. This gene has an essential role in epiblast patterning and mesoderm formation, and null mutant mice with homozygous deletions for exons 2 to 8 of *Ecsit* (which include exon 2 of the antisense ncRNA) die at embryonic day 7.5 (E7.5) with abnormal epiblast patterning (Xiao et al. 2003). This stage corresponds to a distinct peak in the expression of *Ecsit* and a corresponding trough in the expression of the antisense ncRNA (Fig. 2F).

If the observed relationships between ncRNAs and the associated protein-coding genes are functionally significant and conserved in mammals, we would expect their genomic organization to be similarly conserved in other species. Therefore, we analyzed all expressed ncRNA transcripts that were positionally conserved with their associated protein-coding gene in the human genome (Supplemental Table S2). In total, 18% (80 out of 435) had positional equivalents in the human transcriptome, which comprised 17% (59 out of 338) intronic, 26% (16 out of 61) bidirectional, and 14% (five out of 36) *cis*-antisense RNAs. Although many ncRNAs evolve quickly and may be lineage-specific (Pang et al. 2006), the observed prevalence of positional equivalents is similar to that seen in previous studies (Trinklein et al. 2004; Engstrom et al. 2006) and supports the significance of the association between some ncRNAs and their adjacent protein-coding genes.

Characterization of ncRNA promoter regions

Complex transcription factor networks and chromatin states regulate gene expression during ES differentiation (Lee et al. 2006; Guenther et al. 2007). The dynamic expression of ncRNAs in this study suggests that their transcription is tightly regulated. This is supported by recent work showing that ncRNA promoters are subject to purifying selection (Ponjavic et al. 2007), are on average more conserved than promoters of protein-coding genes (Carninci et al. 2005), and are associated with pluripotent transcription factors and regulated chromatin marks (Cawley et al. 2004; Boyer et al. 2005). Therefore, we investigated whether the promoters of ncRNAs identified within this study were subject to such modes of regulation.

Mammalian RNA polymerase II (RNAPII)-dependent promoters associated with CpG islands are typically nontissue-specific and regulate housekeeping genes or genes with complex expression patterns, such as the developmental genes expressed during embryonic differentiation (Saxonov et al. 2006). We analyzed ncRNA promoters and identified 311 (30%) ncRNAs expressed during EB differentiation that were associated with high CpG promoters (HCP) (Supplemental Table S3; see Methods). We then examined the epigenetic state of HCP-associated ncRNAs using previously published mouse ES chromatin maps (Mikkelsen et al. 2007), focusing initially on trimethylation of histone 3 at lysine 4 (H3K4me3), which is generally associated with active transcription. We found that 96% (299 out of 311) of HCP-associated ncRNAs had H3K4me3 chromatin modifications in ES cells, which is a similar proportion to that found with HCP-associated mRNAs (99%). In ES cells, promoters with H3K4me3 modifications may be simultaneously associated with the repressive mark of trimethylation at lysine 27 of histone 3 (H3K27me3). These so-called "bivalent" domains often mark key developmental genes whose expression is thought to be "poised" for lineage-specific activation or repression during differentiation (Bernstein et al. 2006). We found that 61 (~20%) HCPs associated with ncRNAs were bivalent, which is a similar proportion to that found with bivalent mRNA promoters (~17%). We found that 17 of these ncRNAs are differentially expressed in our model, all being up-regulated after day 4 of EB differentiation (Supplemental Fig. S3A), suggesting that in undifferentiated ES cells these ncRNAs are "poised" for expression to fulfil roles in lineage differentiation. This is further supported by analysis of previously published chromatin maps (Mikkelsen et al. 2007) that show that 78% of total bivalent ncRNA promoters are resolved to monovalent chromatin domains in two differentiated cell types (Supplemental Table S3).

The potential biological significance of the ncRNAs identified in this analysis is substantiated by the recovery of ncRNAs with previous functional evidence. For example, the ncRNA *Dleu2* (deleted in lymphocytic leukemia 2) was identified as being subject to regulation via a bivalent domain and up-regulated during mesoderm differentiation. *Dleu2* is an antisense transcript that encompasses the *Kcnrg* and *Trim13* genes, as well as two microRNA genes, *Mirn16-1* and *Mirn15a*, which have been previously identified in lymphocytic leukemia (Corcoran et al. 2004). This ncRNA is also conserved in human, where it is transcribed from an area located within a minimal deleted region that is recurrently lost in patients with chronic lymphocytic leukemia (Supplemental Fig. S3B).

Previous studies have identified highly conserved elements within the promoters of important developmental protein-coding genes (Bejerano et al. 2004; Siepel et al. 2005; Woolfe et al. 2005). Similarly, we found that bivalent HCPs of ncRNAs exhibited enrichment for phastCons elements relative to the genome average (4.2-fold) and other promoters (1.9-fold). PhastCons regions associated with promoters also include a number of ultraconserved elements (UCEs), which are genomic regions (>200 nt) essentially unchanged during vertebrate evolution (Bejerano et al. 2004). Many UCEs have been shown to be transcribed as ncRNAs and fulfil enhancer functions (Pennacchio et al. 2006; Visel et al. 2008). Amongst the ncRNAs targeted within this study, we identified six that coincided with such highly conserved elements, two of which had previously been shown to have enhancer function (Pennacchio et al. 2006). Examination of these ncRNAs revealed that they were associated with the *Dlx1/Dlx2* and *Dlx5/Dlx6* loci (Fig. 3A,B). *Dlx* genes encode homeobox transcription factors that regulate a wide range of developmental programs, including hematopoiesis and neurogenesis (Panganiban and Rubenstein 2002). The ncRNA associated with *Dlx5/Dlx6* corresponded to the previously characterized ncRNA *Evf2*, which regulates the binding of the DLX2 transcription factor to
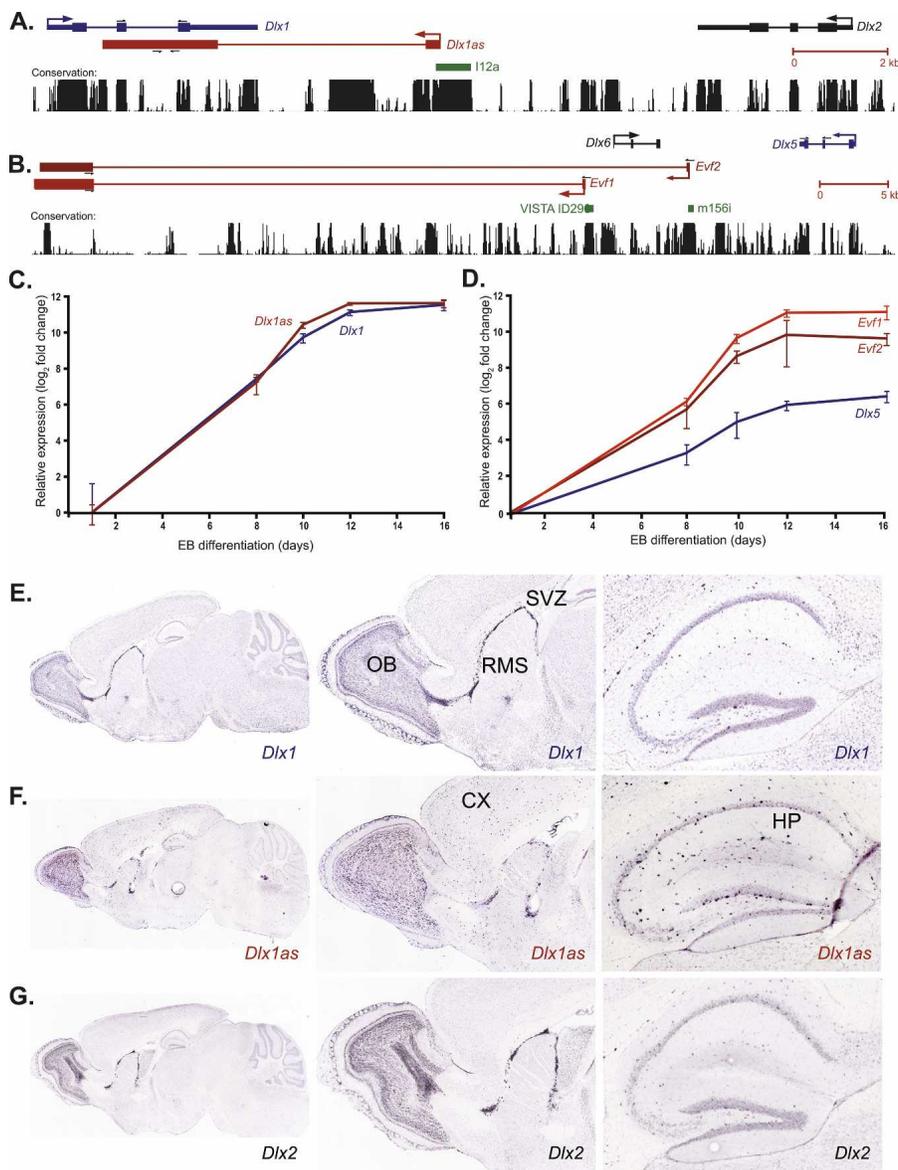
Figure 3. ncRNAs associated with *Dlx1*/*Dlx2* and *Dlx5*/*Dlx6* loci. (*A*,*B*) Genomic context of the *Dlx1*/*Dlx2* (*A*) and *Dlx5*/*Dlx6* (*B*) loci showing the position of the *Dlx* genes (blue), ncRNAs (*Dlx1as* and *Evf1*, light red; *Evf2*, dark red), and the highly conserved enhancers (*I12a* [Park et al. 2004]; *VISTA ID290* [Pennacchio et al. 2006]; *m1561*, [Zerucha et al. 2000]; green). Arrows indicate the direction of transcription. (*C*,*D*) Relative expression profiles of *Dlx1* and *Dlx1as* (*C*) and *Evf1*, *Evf2*, and *Dlx6* (*D*) during EB differentiation as determined by qRT-PCR (relative to day 0 or 1; primer positions indicated in *A*,*B*). Error bars show standard deviation (SD) determined from at least three replicates. (*E*–*G*) ISH of sagittal adult mouse brain sections for *Dlx1* (*E*), *Dlx1as* (*F*), and *Dlx2* (*G*). Whole brain is shown in the *left* panels; subventricular zone (SVZ), rostral migratory stream (RMS), and olfactory bulb (OB) in the *middle* panels; and the hippocampus (HP) in the *right* panels. *Dlx1*, *Dlx1as*, and *Dlx2* show similar expression in the OB, RMS, and SVZ in the brain and in addition *Dlx1as* is strongly expressed in cells dispersed throughout the cortex (CX) and HP.

its originating enhancer element, which in turn regulates the transcriptional activity of the enhancer (Feng et al. 2006).

In addition to the ncRNA associated with *Dlx1*/*Dlx2*, termed *Dlx1as* (GenBank accession no. AK132348; Fig. 3A), we also identified another analogous antisense ncRNA associated with the *Dlx3*/*Dlx4* locus, termed *Dlx4as* (GenBank accession no. AK080562; Supplemental Fig. S5A), although this ncRNA did not correspond to an UCE. *Evf2*, *Dlx1as*, and *Dlx4as* exhibit similar

expression profiles to the associated *Dlx* genes, showing progressively increased expression with EB differentiation (Fig. 3C,D; Supplemental Fig. S5A). Similar expression patterns were also detected in adult mouse tissues, where both *Dlx* genes and associated ncRNAs were detected in the brain (Supplemental Fig. S4). By in situ hybridization (ISH) with adult mouse brain sections, we observe that *Dlx1as* is expressed in the forebrain and in regions associated with neurogenesis (anterior subventricular zone, rostral migratory stream, and olfactory bulb), partially overlapping *Dlx1* and *Dlx2* expression (Fig. 3E–G). An RNA antisense to *Dlx1* has been previously detected in the developing forebrain, most strongly in the subventricular zone, with similar expression to *Dlx1* (McGuinness et al. 1996; Liu et al. 1997). Intriguingly, the highly conserved element corresponding to the *Dlx1as* promoter has been shown to drive reporter gene expression in the mouse embryo, especially in the brachial arches where *Dlx* genes have complex complementary and overlapping patterns of expression (Park et al. 2004). *Dlx4as* is also transcribed from a conserved sequence within *Dlx4* intron 1 and shows specific expression in Purkinje cells in the cerebellum (data not shown). The concordant spatial and temporal expression of the ncRNAs *Dlx1as* and *Dlx4as* and their associated *Dlx* genes further supports the hypothesis that ncRNAs, such as *Evf2*, are transcribed from highly conserved regions to control the expression of adjacent developmental genes (Feng et al. 2006).

Association of ncRNAs with chromatin and chromatin-modifying proteins

Homeotic transcription factors fulfil many important roles in metazoan cell differentiation and development (Kmita and Duboule 2003). Many homeotic genes are associated with ncRNAs and these ncRNAs are often positionally conserved amongst mammalian genomes (Engstrom et al. 2006). Consistent with previous studies, we also identified a number of ncRNAs deriving from homeotic loci, several of which showed concordant expression profiles with the associated homeotic genes.

In mouse and human ES cells, a large number of developmental genes are regulated by Polycomb group (PcG) proteins, which are responsible for establishing H3K27 trimethylation (Boyer et al. 2006; Lee et al. 2006). Homeotic genes are the largest group regulated by PcG proteins, and this regulation is conserved

in metazoans. The repressive role of PcG proteins is counteracted by Trithorax group (TrxG) proteins, which perform H3K4 trimethylation to activate gene expression (Schwartz and Pirrotta 2007). A number of studies have consistently associated the expression of ncRNAs with regulation of *Hox* genes by PcG and TrxG proteins, although the mechanisms are not yet clear and seem to be diverse (Lempradl and Ringrose 2008). Recently, a novel spliced ncRNA, *HOTAIR*, which is transcribed from the human *HOXC* cluster, was shown to repress transcription broadly across the *HOXD* locus (Rinn et al. 2007). *HOTAIR* function is mediated by the interaction with the Polycomb Repressive Complex 2 (PRC2) in *trans* and is required for PRC2 occupancy and H3K27 trimethylation at the *HOXD* locus. Similarly, in *Drosophila* the trithorax protein ASH1 is targeted to *Hox* regulatory elements by ncRNAs, resulting in the activation of *Hox* gene expression (Sanchez-Elsner et al. 2006). However, a parallel with mammalian trithorax proteins has not been established yet. Given the concordant up-regulation of the homeotic genes and their associated ncRNAs seen in our study, we hypothesized that some of these ncRNAs may be involved in regulating chromatin modifications required for the activation of these loci (for example, H3K4 trimethylation).

We focused on two ncRNAs, which we termed *Evx1as* (GenBank accession no. AK031498) and *Hoxb5/6as* (GenBank accession no. AK002860), that were concordantly up-regulated with their associated homeotic gene(s) during the primitive streak phase of EB differentiation (Fig. 4A,B; Bruce et al. 2007a). *Evx1as* is a spliced 2.9-kb ncRNA that is transcribed antisense to *Evx1* and is up-regulated concomitantly with *Evx1* with an expression peak on day 4. *Hoxb5/6as* is a spliced 585-nt ncRNA that is transcribed antisense to a 15-kb region that encompasses the *Hoxb5* and *Hoxb6* genes, all of which show correlated expression profiles that are specifically induced from day 3 during EB differentiation (Fig. 4B). To further understand the relationship between expression of ncRNAs and their associated homeotic gene, we performed whole-mount ISH in mouse embryos. We observed colocalized expression of *Evx1* and *Evx1as* pairs in the mouse tail bud of E9.5 embryos (Fig. 4C) (no signal was detected using the sense riboprobe for the *Evx1as* transcript). These expression results are consistent with previous observations that show *Evx1* is specifically expressed during early mouse embryogenesis in the visceral endoderm and primitive streak (Dush and Martin 1992), and in the tail bud at the end of gastrulation (Gofflot et al. 1997). We did not detect either *Evx1* or *Evx1as* transcripts in adult tissues by
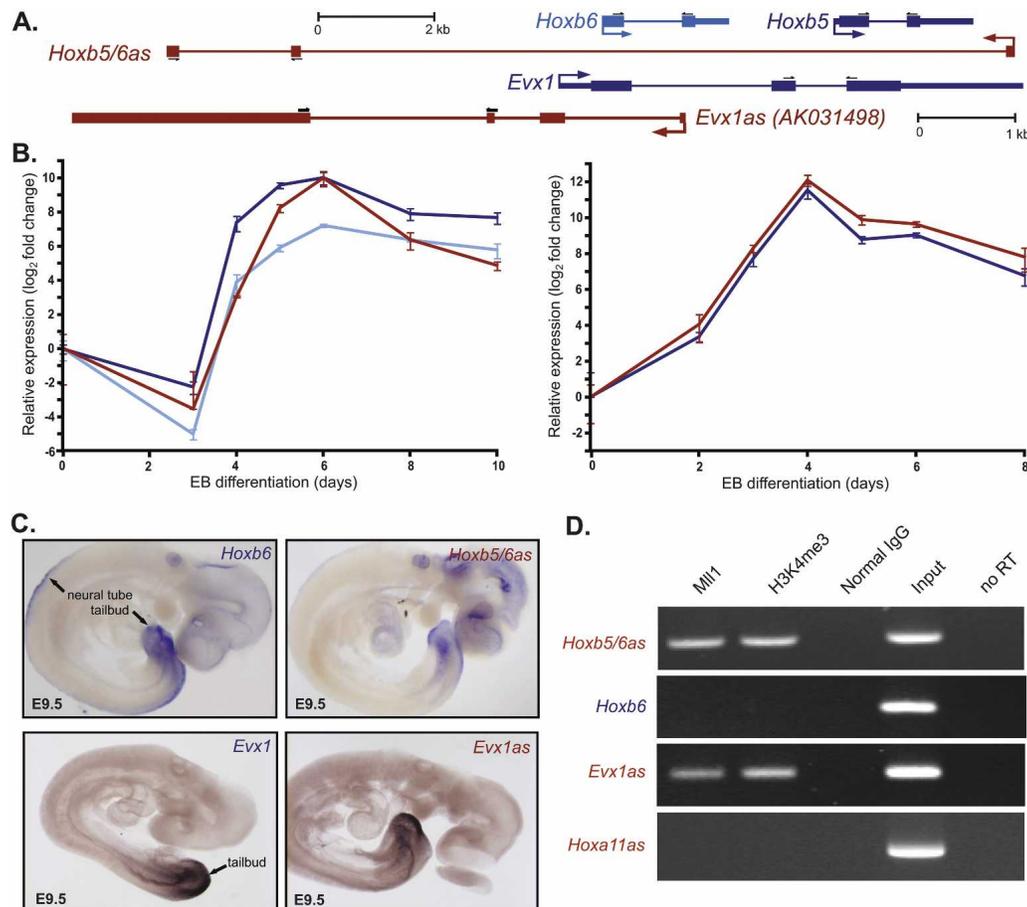

Figure 4. Characterization of ncRNAs associated with *Hoxb5*/*Hoxb6* and *Evx1* loci. (*A*) Genomic context of *Hoxb5*/*Hoxb6* (*top*) and *Evx1* (*bottom*) and their associated ncRNAs, *Hoxb5/6as*, and *Evx1as*. (*B*) Relative expression profiles of *Hoxb5*, *Hoxb6*, and *Hoxb5/6as* (*left*) and *Evx1* and *Evx1as* (*right*) during EB differentiation as determined by qRT-PCR (relative to day 0; primer positions indicated in *A*). Error bars show standard deviation (SD) determined from three replicates. (*C*) Whole-mount ISH showing expression of *Hoxb6* and *Hoxb5/6as* (*upper* panels) and *Evx1* and *Evx1as* (*lower* panels) in the tail bud of E9.5 mouse embryos. (*D*) Association of *Hoxb5/6as* and *Evx1as* RNAs with H3K4me3 chromatin and MLL1 fractions, as detected by ChIP followed by RT-PCR detection (see Methods). Normal IgG was used as a negative control antibody, and input corresponds to RNA present in the samples before ChIP.

RT-PCR (Supplemental Fig. S4) but did detect both RNAs in gastrulating embryos at E6.5 (data not shown). In agreement with previous observations (Medina-Martinez and Ramirez-Solis 2003; Oosterveen et al. 2003), whole-mount ISH of E9.5 embryos showed expression of *Hoxb6* in the posterior embryo (tail) and in the neural tube (Fig. 4C). *Hoxb5/6as* showed concordant expression with *Hoxb6*, although the expression levels were generally weaker (Fig. 4C). Colocalized expression of *Evx1as* and *Hoxb5/6as* and their associated protein-coding genes further suggests a functional connection between these transcripts. In addition, the correlated expression of *Hoxb5/6as* and *Hoxb6* seems to be broadly maintained in mouse adult tissues and cell lines, while *Evx1* and *Evx1as* are similarly detected only in ES/EB day 4 samples (Supplemental Fig. S4).

To determine whether *Evx1as* and *Hoxb5/6as* were associated with chromatin modifications related to transcriptional activation, we used ChIP to isolate H3K4me3-modified chromatin and the associated RNA fraction (RNA-ChIP). Using RT-PCR (see Methods), we found that both *Evx1as* and *Hoxb5/6as* ncRNAs, but not *Hoxb6* mRNA, were present within the precipitated H3K4me3 chromatin fractions (Fig. 4D). A spliced RNA antisense to *Hoxa11* (GenBank accession no. U20367) was also detected in the input sample, but not in the immunoprecipitated chromatin fraction, indicating that the RNA-chromatin association is specific. In addition, the PCR products originated from spliced forms of the ncRNAs, rather than pre-processed primary transcripts, thus excluding contaminating DNA in the chromatin immunoprecipitant. The mammalian trithorax protein MLL1 trimethylates H3K4 and thereby regulates *Hox* loci as well as several other developmental targets in human and mouse cells (Guenther et al. 2005; Milne et al. 2005; Scacheri et al. 2006). We observed that expression of *Mll1* is progressively up-regulated during EB differentiation (Supplemental Fig. S5B). Therefore, we examined the RNA fraction in MLL1 ChIPs to investigate whether *Evx1as* and *Hoxb5/6as* could associate with MLL1 at H3K4me3 loci. Using RT-PCR to analyze the co-immunoprecipitated RNA fraction, we were able to detect both *Hoxb5/6as* and *Evx1as* spliced ncRNAs (Fig. 4D), raising the possibility that these transcripts may be involved in directing the activity of MLL1, in a manner analogous to Ash1 targeting by ncRNAs in *Drosophila* (Sanchez-Elsner et al. 2006).

In light of the highly concordant expression profiles between *Evx1as* and *Hoxb5/6as* and their associated protein-coding genes, we hypothesized that they may be similarly regulated at transcriptional and/or post-transcriptional levels. Therefore, we treated EBs at day 4 of differentiation with the RNAPII inhibitor α-amanitin and quantified the transcript levels after 6, 12, and 24 h of treatment. *Evx1* mRNA and *Evx1as* abundance was similarly affected post α-amanitin treatment, suggesting they have similar biogenesis and turnover rates in differentiating EBs (Supplemental Fig. S6A). In contrast, expression of *Hoxb5/6as* was dissimilar to that of either *Hoxb5* or *Hoxb6* after α-amanitin treatment (Supplemental Fig. S6B). While *Hoxb5* and *Hoxb6* mRNA levels were reduced after 6 h of treatment with α-amanitin, *Hoxb5/6as* levels were not significantly reduced even after 24 h of treatment, although the observed up-regulation during differentiation was impaired. This result suggests that *Hoxb5/6as* is transcribed by RNAPII but is more stable than the associated protein-coding genes. In addition, the greater temporal resolution of this experiment reveals that *Hoxb5/6as* exhibits a distinct expression profile to the *Hoxb5* and *Hoxb6* genes, suggesting independent regulation, although the greater stability can contribute to its accumulation.

Discussion

Cellular identity ultimately arises from changes in gene expression. Two fundamental and interrelated mechanisms that coordinate these changes during lineage commitment are specific alterations to chromatin and transcription factor activity. Several studies indicate an involvement of long ncRNAs in chromatin remodelling (Rougeulle and Heard 2002; Rinn et al. 2007) and co-activation (Feng et al. 2006) during development. Furthermore, long ncRNAs often arise from regions of the genome that are shared with developmental genes, and this positioning is frequently conserved (Engstrom et al. 2006). Therefore, we aimed to determine whether long ncRNAs are broadly involved in development using an ES cell model.

One of the persistent challenges in the investigation of long ncRNAs is that there is no unifying model that can explain their function or mechanism of action, although such models are expected to emerge over the next few years. The relatively few long ncRNAs that have been characterized to date appear to function by diverse mechanisms (Prasanth and Spector 2007). Consequently, large-scale long ncRNA characterization has been resource-intensive and has been met with a low success rate so far (Willingham et al. 2005). As an alternative approach to address this issue, we employed a combination of genome-wide techniques to identify candidates for further functional study. Initially, using a custom microarray, we examined the expression profiles of 3659 ncRNAs over a 16-d EB differentiation time course. We found that 954 were expressed above background of which 174 were significantly differentially expressed. Next, we classified these according to their correlation (positive or negative) to established markers for pluripotency, primitive streak formation, and mesoderm differentiation. This resulted in the identification of ncRNAs coordinately expressed with each of these developmental stages. To further resolve the potential functions of these transcripts, we then examined their genomic context relative to nearby protein-coding genes and the chromatin marks associated with their promoters. These analyses identified several ncRNAs transcribed from regions close to protein-coding genes with developmental roles, and in most cases this transcriptional organization was conserved in other mammals. The conservation and dynamic chromatin modification of the ncRNA promoters further substantiate the roles for ncRNAs in development.

In combination, these analyses have allowed us to identify candidate ncRNAs for further characterization. In analyzing ncRNAs associated with homeotic loci, we observed expression of ncRNAs associated with all three *Dlx* clusters and a similar expression profile between the noncoding and protein-coding genes. This observation raises the possibility that the ncRNAs could participate in the regulatory network that involves *Dlx* genes and coordinate the expression amongst genes of this family (Panganiban and Rubenstein 2002). Furthermore, given that genomic organization often is reflected in coregulation of associated genes in complex developmental processes such as hematopoiesis (Kosak et al. 2007), it is likely that the expression of ncRNAs play a role in this coordination. We also examined two selected ncRNAs, *Evx1as* and *Hoxb5/6as*, which are coexpressed with the associated homeotic genes. A similar observation of concordant expression has been reported for *HOXA* genes and associated ncRNAs in a human tarotcarcinoma cell line during retinoic acid-stimulated differentiation, and it was accompanied by the loss of the Polycomb binding and H3K27me3 mark (Sessa et al. 2007). These observed coordinated expressions may reflect the

sharing of regulatory elements controlling their transcription, but also that these RNAs may have related regulatory functions over the associated genes or in similar processes. In addition, we found using RNA-ChIP that *Evx1as* and *Hoxb5/6as* ncRNAs are associated with H3K4me3 histones and MLL1, indicating that they may have an epigenetic role in *cis*, although a regulatory function in *trans* is also possible, such as the one found for *HOTAIR* from the *HOXC* cluster. Given that homeotic genes are usually associated with both trimethylated H3K4 and H3K27 chromatin regions in undifferentiated ES cells (Bernstein et al. 2006), another attractive possibility is that the chromatin-associated ncRNAs have a role in resolving bivalent domains to activate gene expression by counteracting the silencing mark established by Polycomb proteins. Functional experiments are underway to explore the different possibilities. Nevertheless, our data reinforce the emerging picture that a substantial subset of long ncRNAs, including *Evx1as* and *Hoxb5/6as*, has a chromatin-related function. In fact, eukaryotic chromatin is comprised of a large mass of associated RNAs that is essential for its general structural organization (Nickerson et al. 1989; Rodriguez-Campos and Azorin 2007) and likely mediates a number of chromatin-regulation processes.

From a broader perspective, our results build further support for the notion that long ncRNAs are intrinsically functional. The dynamic expression profiles of the ncRNAs in ES cell differentiation, which are both concordant and discordant to nearby protein-coding genes, suggest their expression and breakdown are specifically regulated. Similar regulated expression of ncRNAs was observed by microarray during myoblast, T-cell, and neuronal cell differentiation (M.E. Dinger, T.R. Mercer, K.C. Pang, W. Chen, G.E. Muscat, M.F. Mehler, and J.S. Mattick, unpubl.), indicating that different ncRNAs may have functions in a variety of developmental processes. This is consistent with a recent study of long ncRNAs expressed in the adult mouse brain, which also revealed extraordinary cell- and tissue-specific expression profiles of hundreds of ncRNAs (Mercer et al. 2008). Although it remains possible that the act of transcription of noncoding regions confers function (Chakalova et al. 2005; Pauler et al. 2007), the purifying selection of long ncRNA splice sites and primary sequence (Ponjavic et al. 2007) as well as the growing number of independently characterized long ncRNAs (Prasanth and Spector 2007) argue that many of these transcripts may be intrinsically functional, especially in view of the findings that at least some of these RNAs may act in *trans* (Rinn et al. 2007; Yu et al. 2008) and our present observations that some of these transcripts are associated with active chromatin. Moreover, it has been recently shown that the majority of the chromatin is open in undifferentiated ES cells and that there is widespread low-level transcription of both protein-coding and noncoding sequences, which is progressively restricted as the cells undergo differentiation (Efroni et al. 2008), although our results clearly indicate that a subset of ncRNAs (as well as mRNAs, which has been known for some time) is induced upon differentiation. With the number of distinct long ncRNAs being of similar order to mRNAs in mammals (Carninci et al. 2005), it is likely that their influence in ES cell biology and early embryonic development and, more broadly, the molecular functioning of complex eukaryotes are considerable. Therefore, the inclusion of long ncRNAs in genome-wide screens, which are becoming prevalent in many areas of biology, will be essential in order to tackle the functional aspect of this profuse component of the transcriptional output of the genome.

Methods

Cell culturing and mouse tissue samples

Low passage number (P18) W9.5 ES cells were maintained in 15% fetal calf serum (FCS) on mitotically inactive mouse embryonic fibroblasts (MEFs) with 1000 U/mL LIF as described (Bruce et al. 2007b). Differentiation was performed in Dulbecco's modified Eagle's medium (DMEM) containing 10% FCS in 1% methylcellulose (GIBCO 10912-012). Feeder-depleted ES cells were seeded at densities ranging from 1×10^5/mL (D1 harvest) to 2×10^3/mL (D16 harvest) to reduce EB aggregation during prolonged differentiation time periods. For RNAPII inhibition experiments, 2×10^4 ES cells were seeded in 6-cm plates in 10% FCS DMEM. At day 4 of differentiation, EB cultures were treated with 20 µg/mL α-amanitin (Sigma) and harvested at 0, 6, 12, and 24 h of treatment. For ChIP experiments, 1×10^6 ES cells were plated in 150-cm² dishes and differentiated in 10% FCS DMEM in the absence of LIF for 6 d. Mouse mammary epithelial HC11 cells (Ball et al. 1988) were cultured in RPMI 1640 medium supplemented with 10% FCS, 5 µg/mL insulin, and 10 ng/mL EGF. Mouse testis-derived cell lines TM3 (ATCC Number CRL-1714) and TM4 (ATCC Number CRL-1715) were cultured in high-glucose DMEM medium containing 10% FCS, 2.5 mM L-glutamine, and 0.5 mM sodium pyruvate. Adult CD1 or C57BL mice were dissected and tissues were collected, frozen in liquid nitrogen, and stored at −80°C until later use for RNA extraction.

RNA preparation

RNA from mouse tissues and cell cultures was purified using TRIzol (Invitrogen) or RNeasy Mini Kit (Qiagen) and treated with DNase I (Invitrogen), according to the protocols provided by the manufacturers. The quality of purified total RNA samples was assessed with an RNA 6000 Nano assay kit using the Agilent 2100 Bioanalyzer (Agilent Technologies) according to the manufacturer's instructions. For microarray experiments, RNA was amplified and labeled using the Amino Allyl Message Amp II kit (Ambion) following the instructions provided by the manufacturer. Amplified aRNA from each time point as well as a reference pooled sample comprising a mixture of RNA from all time points were labeled with either Cy3 or Cy5 monoreactive dyes (Amersham Biosciences) according to the MessageAmp II protocol (Ambion). The quality and quantity of amplified RNA samples were assessed using the Agilent 2100 Bioanalyzer as described above.

Microarray expression analysis

The microarrays contained 22,038 65-mer oligonucleotide probes from the Mouse OligoLibrary (Compugen) and 2118 70-mer oligonucleotide probes that were designed to target ncRNAs, including known mouse pre-miRNAs from miRBase (Griffiths-Jones et al. 2006), longer mouse ncRNAs from RNAdb (Pang et al. 2005), and "high confidence" ncRNAs identified from the FANTOM3 project (Carninci et al. 2005). The custom 70-mer probes were printed alongside Mouse OligoLibrary probes on Power Matrix slides (Full Moon BioSystems) at the SRC Microarray Facility (University of Queensland, Brisbane, Australia). The quality of the print run was verified by hybridizing random 10-mer oligonucleotides to the fisrt and last slides of the run. The array design is available from the ArrayExpress Data Warehouse (EMBL-EBI; ArrayExpress Accession: A-MEXP-1070).

Labeled RNA from each of the 11 time points was hybridized with the common pooled sample to individual microarrays. Two technical replicates were performed for each time point. Blocking, hybridization, and washing were performed according to the manufacturer's instructions (Full Moon BioSystems). Slides were

scanned at 5 µm resolution using a DNA microarray scanner (Agilent Technologies). Feature extraction was performed using ImaGene software (BioDiscovery), with manual grid adjustment and auto-spot finding and segmentation. Data were exported from ImaGene as text files, then uploaded and analyzed using the Linear Models for Microarray Data (LIMMA) software package via the R Project for Statistical Computing (www.r-project.org). Data were background-corrected, normalized both within and between arrays (Smyth and Speed 2003), and differential expression analysis was performed by fitting a linear model of the data to the experimental design matrix and then calculating Bayesian statistics (B-statistics; posterior log odds) adjusted for multiple testing using Benjamini-Hochberg analysis (Smyth 2004). Raw and processed microarray data are available at the ArrayExpress Data Warehouse (EMBL-EBI; ArrayExpress accession no. E-TABM-433).

Classification of probes as protein-coding or nonprotein-coding

Although the Mouse OligoLibrary probe set was predominantly designed to recognize known or putative protein-coding transcripts, several thousand probes targeted miscellaneous cDNAs and ESTs whose coding status was not well-characterized at the time this commercial probe set was first produced. To update the annotation of these probes and to clarify whether they targeted protein-coding or noncoding regions, a computational pipeline was designed to reannotate the entire probe set. Sequences for all probes were mapped to the February 2006 (NCBI Build 36) assembly of the mouse genome using BLAT (Kent 2002) (parameters: minScore = 50, minIdentity = 99, stepSize = 5, tileSize = 11, ooc = 11.ooc). Probes that could not be reliably mapped were excluded from the study. Targeted transcripts were then defined as protein-coding and noncoding as described previously (Mercer et al. 2008).

Determination of genomic context of probe targets

The genomic context of ncRNAs (relative to protein-coding genes) was determined as described previously (Engstrom et al. 2006; Mercer et al. 2008). Briefly, *cis*-antisense probes were defined where the probe mapped to the opposite strand of a 5′ untranslated region (UTR), coding sequence, or 3′ UTR; intronic probes were defined where the probe mapped within the intron of a protein-coding gene; and bidirectional probes were defined as noncoding probes that targeted transcripts that were oriented head-head to a protein-coding gene within 1000 bp.

Conservation and secondary structure predictions of ncRNAs

Enrichment for conservation of ncRNAs was determined from the proportion of transcript bases annotated as phastCons elements by Siepel et al. (2005). The secondary structural composition of expressed ncRNAs was determined by intersecting their chromosomal positions with those of the RNAz structural predictions made across the entire mouse genome as previously described (Washietl et al. 2005a; Mercer et al. 2008). RNAz uses multiple genome alignments to predict regions that contain thermodynamically stable and conserved RNA secondary structures (Washietl et al. 2005b). The significance of the classification is quantified as "RNA-class probability", P. Conserved RNA secondary structures were considered significant at confidence threshold levels of $P > 0.5$ or $P > 0.9$.

Quantitative real-time PCR (qRT-PCR)

cDNA preparation and qRT-PCR analysis were performed as described (Bruce et al. 2007a). Primers were designed spanning

splice sites in most cases (Supplemental Table S4) and PCR products were sequenced to confirm the identity of the fragments. In α-amanitin inhibition experiments, cDNA was produced using random primers and quantified relative to 18S RNA expression. In all qRT-PCR experiments, a minimum of three replicates were performed. For tissue expression analysis, cDNA was used in PCR for 35 cycles and amplification products were visualized after electrophoresis in 2%–3% agarose gels. For in situ hybridization (ISH) probe preparation, cDNA from ES cells was amplified (see primers in Supplemental Table S5) and PCR products were cloned into pGEM-T Easy Vectors (Promega), sequenced, and used in PCR with T7 and SP6 primers to generate PCR templates for in vitro transcription reactions (see below).

In situ hybridization

Adult mouse brain section ISH was performed as previously described (Lein et al. 2007). Whole-mount ISH was performed according to the protocol described previously (Christiansen et al. 1995). Briefly, embryos were dissected from pregnant C57BL mice at 9.5 dpc, fixed in 4% paraformaldehyde in PBS overnight at 4°C, and subsequently washed twice in PBTX (PBS containing 0.1% Triton X-100) for 10 min at 4°C. Embryos were then dehydrated and rehydrated through a methanol series and washed in PBTX twice for 10 min at room temperature. Embryos were then treated with 10 µg/mL proteinase K in PBTX at 37°C for 10 min, washed, and refixed accordingly. Next, embryos were incubated overnight at 65°C in prehybridization buffer containing 50% formamide, 5× SSC, 2% blocking powder, 0.1% Triton X-100, 0.5% CHAPS, 1 mg/mL yeast tRNA, 5 mM EDTA, and 50 µg/mL heparin. Digoxygenin (DIG)-labeled riboprobes were transcribed from the PCR templates using T7 or SP6 polymerase added to prehybridized embryos at a concentration of 1–2 µg/mL and incubated overnight at 65°C. The embryos were then washed, blocked, and incubated overnight at 4°C with anti-DIG antibody (Roche). Subsequently, embryos were washed and incubated with color reagent (NBT/BCIP; Roche) until the color had developed to the desired extent, washed several times in PBTX to remove the background color, and photographed.

Classification of promoters based on CpG content

Promoter regions were classified as described previously (Mikkelsen et al. 2007). Transcripts with a 500 bp interval within −0.5 kb to +2 kb of the transcription start site (TSS) with a GC fraction ≥0.55 and an observed to expected ratio (O/E) ≥ 0.6 were classified as high CpG promoters (HCPs). Promoters where all 500 bp intervals within −0.5 kb to +2 kb of TSS have CpG O/E ≤ 0.4 were classified as low CpG promoters (LCPs). All remaining promoters were classified as intermediate CpG promoters (ICPs). The CpG O/E ratio was calculated as described previously (Gardiner-Garden and Frommer 1987).

RNA-chromatin immunoprecipitation (RNA-ChIP)

ChIP was performed according to a previously described procedure (Boyer et al. 2006) with modifications to preserve the RNA-associated fraction. RNaseOUT (100 U/mL; Invitrogen) was added to all lysis buffers, and sonication of formaldehyde fixed cells was performed with a Vibra Cell Sonicator (Sonicas & Materials Inc.) for 10 pulses of 30 sec at Set 5, including a 60-sec incubation on ice bath between each pulse. Five percent of sonicated lysate volumes were separated as "input" material and stored at −80°C. Immunoprecipitation was carried out overnight at 4°C with 5 µg of polyclonal antibodies raised in rabbit against histone H3 trimethyl K4 (Abcam), MLL1 (Bethyl Laboratories), or control normal rabbit IgG (Abcam), using 50 µL of Dynabeads

M-280 sheep anti-rabbit IgG (Invitrogen). After immunoprecipitation, the magnetic beads were washed four times in 1 mL of RIPA wash buffer and one time in 1 mL of TE containing 50 mM NaCl. Immunoprecipitants were eluted in 200 μL of elution buffer containing RNaseOUT (100 U/mL) and, together with input samples, incubated at 65°C for 4 h for cross-linking reversal. Samples were then treated with 80 μg of proteinase K and nucleic acids were phenol-chloroform-extracted, ethanol-precipitated using 50 μg of Glycoblue (Invitrogen), and resuspended in 20 μL of DEPC-H_2O. Samples were treated with 2 U of DNase I (Invitrogen) and reverse-transcribed using SuperScriptIII and oligo(dT) or random hexamers (Invitrogen) in 40 μL reactions, following the manufacturer's instructions. RNA-ChIP analysis was performed as described (Peritz et al. 2006), in which 2 μL of cDNA were used in the first-round PCR (50 μL) for 28 cycles using an external primer, and 1 μL of PCR product used in the second-round PCR (50 μL) for 35 cycles using nested primers spanning splice junctions (primer sequences are listed in Supplemental Table S4). Amplification products were visualized by running 10 μL of the PCR samples in 3% agarose gels.

Acknowledgments

References

Ball, R.K., Friis, R.R., Schoenenberger, C.A., Doppler, W., and Groner, B. 1988. Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *EMBO J.* **7:** 2089–2095.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. 2004. Ultraconserved elements in the human genome. *Science* **304:** 1321–1325.

Bernstein, E. and Allis, C.D. 2005. RNA meets chromatin. *Genes & Dev.* **19:** 1635–1655.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315–326.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122:** 947–956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441:** 349–353.

Bruce, S.J., Gardiner, B.B., Burke, L.J., Gongora, M.M., Grimmond, S.M., and Perkins, A.C. 2007a. Dynamic transcription programs during ES cell differentiation towards mesoderm in serum versus serum-free[BMP4] culture. *BMC Genomics* **8:** 365. doi: 10.1186/1471-2164-8-365.

Bruce, S.J., Rea, R.W., Steptoe, A.L., Busslinger, M., Bertram, J.F., and Perkins, A.C. 2007b. In vitro differentiation of murine embryonic stem cells toward a renal lineage. *Differentiation* **75:** 337–349.

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 mammalian genome. *Science* **309:** 1559–1563.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116:** 499–509.

Chakalova, L., Debrand, E., Mitchell, J.A., Osborne, C.S., and Fraser, P. 2005. Replication and transcription: Shaping the landscape of the genome. *Nat. Rev. Genet.* **6:** 669–677.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113:** 643–655.

Christiansen, J.H., Dennis, C.L., Wicking, C.A., Monkley, S.J., Wilkinson, D.G., and Wainwright, B.J. 1995. Murine Wnt-11 and Wnt-12 have temporally and spatially restricted expression patterns during embryonic development. *Mech. Dev.* **51:** 341–350.

Corcoran, M.M., Hammarsund, M., Zhu, C., Lerner, M., Kapanadze, B., Wilson, B., Larsson, C., Forsberg, L., Ibbotson, R.E., Einhorn, S., et al. 2004. *DLEU2* encodes an antisense RNA for the putative bicistronic *RFP2/LEU5* gene in humans and mouse. *Genes Chromosomes Cancer* **40:** 285–297.

Dush, M.K. and Martin, G.R. 1992. Analysis of mouse *Evx* genes: *Evx-1* displays graded expression in the primitive streak. *Dev. Biol.* **151:** 273–287.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., et al. 2008. Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2:** 437–447.

Engstrom, P.G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Luzi, L., Tan, S.L., Yang, L., et al. 2006. Complex loci in human and mouse genomes. *PLoS Genet.* **2:** e47. doi: 10.1371/journal.pgen.0020047.

Evans, M.J. and Kaufman, M.H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292:** 154–156.

Feng, J., Bi, C., Clark, B.S., Mady, R., Shah, P., and Kohtz, J.D. 2006. The *Evf-2* noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes & Dev.* **20:** 1470–1484.

Gardiner-Garden, M. and Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196:** 261–282.

Garzon, R., Pichiorri, F., Palumbo, T., Iuliano, R., Cimmino, A., Aqeilan, R., Volinia, S., Bhatt, D., Alder, H., Marcucci, G., et al. 2006. MicroRNA fingerprints during human megakaryocytopoiesis. *Proc. Natl. Acad. Sci.* **103:** 5078–5083.

Ginger, M.R., Shore, A.N., Contreras, A., Rijnkels, M., Miller, J., Gonzalez-Rimbau, M.F., and Rosen, J.M. 2006. A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc. Natl. Acad. Sci.* **103:** 5781–5786.

Gofflot, F., Hall, M., and Morriss-Kay, G.M. 1997. Genetic patterning of the developing mouse tail at the time of posterior neuropore closure. *Dev. Dyn.* **210:** 431–445.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. 2006. miRBase: MicroRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34:** D140–D144.

Guenther, M.G., Jenner, R.G., Chevalier, B., Nakamura, T., Croce, C.M., Canaani, E., and Young, R.A. 2005. Global and Hox-specific roles for the MLL1 methyltransferase. *Proc. Natl. Acad. Sci.* **102:** 8603–8608.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130:** 77–88.

Hirst, C.E., Ng, E.S., Azzola, L., Voss, A.K., Thomas, T., Stanley, E.G., and Elefanty, A.G. 2006. Transcriptional profiling of mouse and human ES cells identifies SLAIN1, a novel stem cell gene. *Dev. Biol.* **293:** 90–103.

Houbaviy, H.B., Murray, M.F., and Sharp, P.A. 2003. Embryonic stem cell-specific MicroRNAs. *Dev. Cell* **5:** 351–358.

Ivanova, N.B., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., and Lemischka, I.R. 2002. A stem cell molecular signature. *Science* **298:** 601–604.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442:** 533–538.

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.

Keller, G.M. 1995. In vitro differentiation of embryonic stem cells. *Curr. Opin. Cell Biol.* **7:** 862–869.

Kent, W.J. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* **12:** 656–664.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132:** 1049–1061.

Kmita, M. and Duboule, D. 2003. Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301:** 331–333.

Kosak, S.T., Scalzo, D., Alworth, S.V., Li, F., Palmer, S., Enver, T., Lee, J.S., and Groudine, M. 2007. Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol.* **5:** e309. doi: 10.1371/journal.pbio.0050309.

Lakshmipathy, U., Love, B., Goff, L.A., Jornsten, R., Graichen, R., Hart, R.P., and Chesnut, J.D. 2007. MicroRNA expression pattern of undifferentiated and differentiated human embryonic stem cells. *Stem Cells Dev.* **16:** 1003–1016.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125:** 301–313.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445:** 168–176.

Lempradl, A. and Ringrose, L. 2008. How does noncoding transcription regulate Hox genes? *Bioessays* **30:** 110–121.

Liu, J.K., Ghattas, I., Liu, S., Chen, S., and Rubenstein, J.L. 1997. Dlx genes encode DNA-binding proteins that are expressed in an overlapping and sequential pattern during basal ganglia differentiation. *Dev. Dyn.* **210:** 498–512.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38:** 431–440.

Martin, G.R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci.* **78:** 7634–7638.

Mattick, J.S. 2007. A new paradigm for developmental biology. *J. Exp. Biol.* **210:** 1526–1547.

Mattick, J.S. and Makunin, I.V. 2005. Small regulatory RNAs in mammals. *Hum. Mol. Genet.* **14:** R121–R132.

McGuinness, T., Porteus, M.H., Smiga, S., Bulfone, A., Kingsley, C., Qiu, M., Liu, J.K., Long, J.E., Xu, D., and Rubenstein, J.L. 1996. Sequence, organization, and transcription of the *Dlx-1* and *Dlx-2* locus. *Genomics* **35:** 473–485.

Medina-Martinez, O. and Ramirez-Solis, R. 2003. In vivo mutagenesis of the Hoxb8 hexapeptide domain leads to dominant homeotic transformations that mimic the loss-of-function mutations in genes of the Hoxb cluster. *Dev. Biol.* **264:** 77–90.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. 2008. Specific expression of long noncoding RNAs in the adult mouse brain. *Proc. Natl. Acad. Sci.* **105:** 716–721.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–560.

Mikkola, H.K. and Orkin, S.H. 2006. The journey of developing hematopoietic stem cells. *Development* **133:** 3733–3744.

Milne, T.A., Dou, Y., Martin, M.E., Brock, H.W., Roeder, R.G., and Hess, J.L. 2005. MLL associates specifically with a subset of transcriptionally active target genes. *Proc. Natl. Acad. Sci.* **102:** 14765–14770.

Mineno, J., Okamoto, S., Ando, T., Sato, M., Chono, H., Izu, H., Takayama, M., Asada, K., Mirochnitchenko, O., Inouye, M., et al. 2006. The expression profile of microRNAs in mouse embryos. *Nucleic Acids Res.* **34:** 1765–1771.

Molkentin, J.D. 2000. The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J. Biol. Chem.* **275:** 38949–38952.

Nakaya, H.I., Amaral, P.P., Louro, R., Lopes, A., Fachel, A.A., Moreira, Y.B., El-Jundi, T.A., da Silva, A.M., Reis, E.M., and Verjovski-Almeida, S. 2007. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol.* **8:** R43. doi: 10.1186/gb-2007-8-3-r43.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95:** 379–391.

Nickerson, J.A., Krochmalnic, G., Wan, K.M., and Penman, S. 1989. Chromatin architecture and nuclear RNA. *Proc. Natl. Acad. Sci.* **86:** 177–181.

Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T.G., and Landweber, L.F. 2007. RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* **451:** 153–158.

Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A.P., Disteche, C.M., Avner, P., and Heard, E. 2005. Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature* **438:** 369–373.

Oosterveen, T., Niederreither, K., Dolle, P., Chambon, P., Meijlink, F., and Deschamps, J. 2003. Retinoids regulate the anterior expression boundaries of 5′ Hoxb genes in posterior hindbrain. *EMBO J.* **22:** 262–269.

Pang, K.C., Stephen, S., Engstrom, P.G., Tajul-Arifin, K., Chen, W., Wahlestedt, C., Lenhard, B., Hayashizaki, Y., and Mattick, J.S. 2005. RNAdb—A comprehensive mammalian noncoding RNA database. *Nucleic Acids Res.* **33:** D125–D130.

Pang, K.C., Frith, M.C., and Mattick, J.S. 2006. Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. *Trends Genet.* **22:** 1–5.

Panganiban, G. and Rubenstein, J.L. 2002. Developmental functions of the *Distal-less*/Dlx homeobox genes. *Development* **129:** 4371–4386.

Park, B.K., Sperber, S.M., Choudhury, A., Ghanem, N., Hatch, G.T., Sharpe, P.T., Thomas, B.L., and Ekker, M. 2004. Intergenic enhancers with distinct activities regulate *Dlx* gene expression in the mesenchyme of the branchial arches. *Dev. Biol.* **268:** 532–545.

Pauler, F.M., Koerner, M.V., and Barlow, D.P. 2007. Silencing by imprinted noncoding RNAs: Is transcription the answer? *Trends Genet.* **23:** 284–292.

Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444:** 499–502.

Peritz, T., Zeng, F., Kannanayakal, T.J., Kilk, K., Eiriksdottir, E., Langel, U., and Eberwine, J. 2006. Immunoprecipitation of mRNA-protein complexes. *Nat. Protoc.* **1:** 577–580.

Ponjavic, J., Ponting, C.P., and Lunter, G. 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17:** 556– 565.

Prasanth, K.V. and Spector, D.L. 2007. Eukaryotic regulatory RNAs: An answer to the 'genome complexity' conundrum. *Genes & Dev.* **21:** 11–42.

Prelle, K., Zink, N., and Wolf, E. 2002. Pluripotent stem cells Model of embryonic development, tool for gene targeting, and basis of cell therapy. *Anat. Histol. Embryol.* **31:** 169–186.

Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., and Melton, D.A. 2002. "Stemness": Transcriptional profiling of embryonic and adult stem cells. *Science* **298:** 597–600.

Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M., et al. 2006. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.* **16:** 11–19.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129:** 1311–1323.

Robertson, S.M., Kennedy, M., Shannon, J.M., and Keller, G. 2000. A transitional stage in the commitment of mesoderm to hematopoiesis requiring the transcription factor SCL/tal-1. *Development* **127:** 2447–2459.

Rodriguez-Campos, A. and Azorin, F. 2007. RNA is an integral component of chromatin that contributes to its structural organization. *PLoS ONE* **2:** e1182. doi: 10.1371/journal.pone.0001182.

Rougeulle, C. and Heard, E. 2002. Antisense RNA in imprinting: Spreading silence through Air. *Trends Genet.* **18:** 434–437.

Sanchez-Elsner, T., Gou, D., Kremmer, E., and Sauer, F. 2006. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science* **311:** 1118–1123.

Saxonov, S., Berg, P., and Brutlag, D.L. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci.* **103:** 1412–1417.

Scacheri, P.C., Davis, S., Odom, D.T., Crawford, G.E., Perkins, S., Halawi, M.J., Agarwal, S.K., Marx, S.J., Spiegel, A.M., Meltzer, P.S., et al. 2006. Genome-wide analysis of menin binding provides insights into MEN1 tumorigenesis. *PLoS Genet.* **2:** e51. doi: 10.1371/journal.pgen.0020051.

Schratt, G.M., Tuebing, F., Nigh, E.A., Kane, C.G., Sabatini, M.E., Kiebler, M., and Greenberg, M.E. 2006. A brain-specific microRNA regulates dendritic spine development. *Nature* **439:** 283–289.

Schulz, W.A. and Hoffmann, M.J. 2007. Transcription factor networks in embryonic stem cells and testicular cancer and the definition of epigenetics. *Epigenetics* **2:** 37–42.

Schwartz, Y.B. and Pirrotta, V. 2007. Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet.* **8:** 9–22.

Sessa, L., Breiling, A., Lavorgna, G., Silvestri, L., Casari, G., and Orlando, V. 2007. Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human *HOXA* cluster. *RNA* **13:** 223–239.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15:** 1034–1050.

Sleutels, F., Zwart, R., and Barlow, D.P. 2002. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415:** 810–813.

Smith, A.G. 2001. Embryo-derived stem cells: Of mice and men. *Annu. Rev. Cell Dev. Biol.* **17:** 435–462.

Smyth, G.K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3:** Article3.

Smyth, G.K. and Speed, T. 2003. Normalization of cDNA microarray data. *Methods* **31:** 265–273.

Thakur, N., Tiwari, V.K., Thomassin, H., Pandey, R.R., Kanduri, M., Gondor, A., Grange, T., Ohlsson, R., and Kanduri, C. 2004. An antisense RNA regulates the bidirectional silencing property of the *Kcnq1* imprinting control region. *Mol. Cell. Biol.* **24:** 7855–7862.

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. 2004. An abundance of bidirectional promoters in the human genome. *Genome Res.* **14:** 62–66.

Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E.M., and Pennacchio, L.A. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat. Genet.* **40:** 158–160.

Wang, H., Iacoangeli, A., Lin, D., Williams, K., Denman, R.B., Hellen, C.U., and Tiedge, H. 2005. Dendritic BC1 RNA in translational control mechanisms. *J. Cell Biol.* **171:** 811–821.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. 2006. A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444:** 364–368.

Washietl, S., Hofacker, I.L., Lukasser, M., Huttenhofer, A., and Stadler, P.F. 2005a. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.* **23:** 1383–1390.

Washietl, S., Hofacker, I.L., and Stadler, P.F. 2005b. Fast and reliable

prediction of noncoding RNAs. *Proc. Natl. Acad. Sci.* **102:** 2454–2459.

Wienholds, E., Kloosterman, W.P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., Horvitz, H.R., Kauppinen, S., and Plasterk, R.H. 2005. MicroRNA expression in zebrafish embryonic development. *Science* **309:** 310–311.

Wilkinson, D.G., Bhatt, S., and Herrmann, B.G. 1990. Expression pattern of the mouse *T* gene and its role in mesoderm formation. *Nature* **343:** 657–659.

Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B., and Schultz, P.G. 2005. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **309:** 1570–1573.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., et al. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3:** e7. doi: 10.1371/journal.pbio.0030007.

Xiao, C., Shim, J.H., Kluppel, M., Zhang, S.S., Dong, C., Flavell, R.A., Fu, X.Y., Wrana, J.L., Hogan, B.L., and Ghosh, S. 2003. Ecsit is required for Bmp signaling and mesoderm formation during mouse embryogenesis. *Genes & Dev.* **17:** 2933–2949.

Yan, M.D., Hong, C.C., Lai, G.M., Cheng, A.L., Lin, Y.W., and Chuang, S.E. 2005. Identification and characterization of a novel gene *Saf* transcribed from the opposite strand of *Fas. Hum. Mol. Genet.* **14:** 1465–1474.

Yang, D.H., Cai, K.Q., Roland, I.H., Smith, E.R., and Xu, X.X. 2007. Disabled-2 is an epithelial surface positioning gene. *J. Biol. Chem.* **282:** 13114–13122.

Young, T.L., Matsuda, T., and Cepko, C.L. 2005. The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr. Biol.* **15:** 501–512.

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P., and Cui, H. 2008. Epigenetic silencing of tumour suppressor gene *p15* by its antisense RNA. *Nature* **451:** 202–206.

Zambrowicz, B.P., Friedrich, G.A., Buxton, E.C., Lilleberg, S.L., Person, C., and Sands, A.T. 1998. Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392:** 608–611.

Zerucha, T., Stuhmer, T., Hatch, G., Park, B.K., Long, Q., Yu, G., Gambarotta, A., Schultz, J.R., Rubenstein, J.L., and Ekker, M. 2000. A highly conserved enhancer in the *Dlx5/Dlx6* intergenic region is the site of cross-regulatory interactions between *Dlx* genes in the embryonic forebrain. *J. Neurosci.* **20:** 709–721.

Zhao, Y., Samal, E., and Srivastava, D. 2005. Serum response factor regulates a muscle-specific microRNA that targets *Hand2* during cardiogenesis. *Nature* **436:** 214–220.