doi:10.1038/nature20149

Supplementary Notes

Note S1. A genetic approach to dissect *cis* regulatory functions

Note S2. Defining a window of interest for long-range *cis* regulatory functions

Note S3. A method for high-throughput measurement of allele-specific RNA expression

Note S4. Read-through transcription at the Meg3 and Snhg3 loci

Note S5. Proposed mechanisms for *cis* regulatory effects in Rcc1, linc2025, and Sfmbt2 loci

Note S6. Mechanism for up-regulation of Blustr upon deletion of intron 1

Note S7. Mechanisms for how transcription of Blustr activates Sfmbt2

Note S8. Generality of local regulatory effects by promoters, transcription, and splicing

Note S9. Evolutionary conservation of lncRNAs and their promoters

References

Note S1. A genetic approach to dissect *cis* regulatory functions

We developed a genetic loss-of-function approach to dissect the *cis* regulatory functions of lncRNA and mRNA loci. Interpreting these loss-of-function experiments critically relies on the distinction between (i) changes in neighboring-gene expression resulting from direct, local functions of a gene locus and (ii) changes resulting as downstream consequences of non-local functions of the gene locus (*i.e.*, the lncRNA or mRNA acting elsewhere in the cell). Importantly, homozygous knockout or RNA knockdown approaches cannot distinguish between these two possibilities. For example, RNA knockdown experiments targeting an mRNA often affect the expression of hundreds to thousands of genes as downstream consequences of loss of protein function, including by chance some of the dozens of genes within 1 Mb of the mRNA locus[1]. Similarly, some lncRNAs may function in cellular locations far from their sites of transcription (*e.g.*, in the cytoplasm) and also can have broad effects on gene expression[1]. Thus, to identify and dissect *cis* regulatory functions, we need to distinguish between effects resulting from local and non-local functions.

To distinguish between these possibilities, we generated clonal cell lines carrying heterozygous genetic modifications at a lncRNA locus and compared the expression of nearby genes on the *cis* and *trans* alleles (*i.e.*, on the modified and unmodified homologous chromosomes) in the same cells. Changes in neighboring gene expression that specifically affect the *cis* allele likely result from local functions of the locus, while changes that affect both the *cis* and *trans* alleles likely result as downstream effects of non-local functions. We performed genetic modifications in 129/Castaneus F1 hybrid mouse embryonic stem cells (mESCs), enabling us to distinguish the two alleles using RNA sequencing of polymorphic sites. Similar *cis*-*trans* tests have been extensively used in genetics to study *cis*-acting regulatory elements — both in the context of interpreting phenotypes in mouse knockouts[2-4] as well as in measuring allele-specific expression[5-7].

The high frequency of polymorphic sites in 129/Castaneus hybrid cells (1 single nucleotide polymorphism (SNP) every ~140 bp) helps us to distinguish the *cis* and *trans* alleles, but the frequency of SNPs also raises the possibility that some allele-specific changes in neighboring-gene expression could result from haplotype-specific responses to non-local functions. For example, if we were to knock out a gene encoding a transcription factor, we might observe allele-specific changes in the expression of a non-local gene elsewhere in the genome because this gene has 129/Castaneus genetic variation in a motif for the transcription factor. To account for this possibility, we examined knockouts on each of the 129 and Castaneus alleles and checked that they had consistent effects. Neighboring genes that respond due to local functions of the locus should show changes on the allele linked to the genetic modification (only in *cis*). In contrast, neighboring genes that respond due to haplotype-specific effects of non-local functions should show consistent changes on a single haplotype, regardless of whether or not it is linked to the genetic modification (in *cis* or in *trans*). Thus, by examining knockouts on both haplotypes, we can distinguish effects due to local or non-local functions even in hybrid cells.

We note that this allele-specific approach also increases our ability to rule out artificial changes in gene expression, for example due to off-target effects of the perturbations themselves (CRISPR/Cas9, in our study) or due to biological or technical variation between clonal cell lines. These technical sources of variation would most likely result in changes to the cell that are not local to the targeted locus, and therefore would likely affect the expression of neighboring genes on both the *cis* and *trans* alleles. Thus, by focusing on allele-specific changes in the expression of neighboring genes, we reduce the possibility that the changes we observe might result either from non-local functions or from off-target effects.

Note S2. Defining a window of interest for long-range *cis* regulatory functions

To identify lncRNAs that regulate local gene expression, we examined the allele-specific expression of genes within 1 Mb of the knocked-out promoter. We chose this distance because most experimentally tested enhancer-promoter regulation occurs within this linear distance[8], as do the majority of proximity-based contacts between active promoters and regulatory elements[9]. We note that most known enhancer-promoter connections, including the few that span distances greater than 1 Mb, occur within the confines of topologically associated domains (TADs)[8], as defined by chromosome capture conformation assays. For each of the loci chosen here, the mESC TAD[10] overlapping the knocked out gene is completely contained within the 2-Mb window centered on the promoter, and in many cases this window includes entire neighboring TADs as well.

Several observations suggest that we did not miss *cis* regulatory functions that affect genes outside of this 1-Mb window. First, the maximum distance between a knocked out promoter and a regulated neighboring promoter is 220 kb (Slc30a9 to Bend4). Second, all of the regulatory connections we find involve a gene immediately adjacent to the knocked-out locus. Finally, to experimentally determine whether we might find *cis* regulatory effects beyond our initial 1-Mb window, we re-sequenced the full transcriptomes (before hybrid selection) for the promoter knockouts of 4 lncRNAs (Blustr, linc1405, linc1386, and linc1547), and did not find additional allele-specific *cis* regulatory effects outside the initial window (data not shown).

Note S3. A method for high-throughput measurement of allele-specific RNA expression

To identify lncRNAs that regulate local gene expression, we developed an approach to measure allele-specific expression across many genes. While whole-transcriptome sequencing would be straightforward, the sequencing depth required to measure allele-specific expression across hundreds of samples proved too costly. Accordingly, we developed a hybrid selection approach to enrich for allele-informative RNA fragments in selected genes close to a knocked-out locus. We designed this method to (i) accurately measure the relative expression of each RNA allele, (ii) simultaneously characterize hundreds of genes across hundreds of samples, and (iii) require relative low input RNA corresponding to the contents of a 96-well cell culture plate (~10,000 cells).

In this method, we first generate barcoded RNA sequencing libraries using a multiplexed protocol (see Methods) and combine all of the barcoded libraries into a single pool (**Extended Data Fig. 2a**). To enrich for allele-informative cDNA fragments corresponding to genes of interest, we perform hybrid selection on this combined pool using a library of oligos designed to capture several distinct polymorphic sites in each of ~800 transcripts (see Methods). We include oligos complementary to each allele to minimize sequence-based differences in capture efficiency between the two alleles.

We verified through multiple independent means that this hybrid-selection based approach accurately assesses the relative expression of the two alleles of a given RNA:

(i) We examined the distribution of allelic expression ratios (read counts on 129 allele versus total read counts) across all genes. The distribution was symmetric and centered on 0.5 (corresponding to 50% expression from each allele), indicating that our mapping procedures were not biased toward the reference allele (**Extended Data Fig. 2b**). The substantial fraction of genes showing skewed expression toward one allele or another is consistent with previous reports[11,12].

(ii) We compared allele-specific expression in data from RNA sequencing libraries before and after hybrid selection. The allelic expression ratios correlated well (Pearson's R = 0.88), with most outliers representing transcripts with low absolute expression in mESCs (**Extended Data Fig. 2c**). To mitigate any remaining biases introduced by the hybrid selection in our subsequent knockout experiments, we compared measurements of allele-specific expression in knockouts and controls in the same hybrid selection batch.

(iii) To validate these measurements of allelic expression ratios through an independent technique, we used allele-specific quantitative PCR assays that internal fluorescent probes to distinguish between two alleles of a cDNA amplicon (see Methods). We measured the allele-specific expression of two genes (Blustr and Sfmbt2) across 15 different clones containing genetic modifications in the Blustr locus, and found that the allelic expression ratio as measured by hybrid selection RNA-sequencing correlated

closely with the ratio as measured by allele-specific qPCR (Pearson's $R = 0.99$ for Blustr and 0.98 for Sfmbt2, **Extended Data Fig. 2d,e**).

Thus, this hybrid-selection based approach accurately measures allele-specific RNA expression and allows comparisons of knockout and control clones.

Note S4. Read-through transcription at the Meg3 and Snhg3 loci

Based on the initial set of 12 lncRNA loci, we dissected the functions of 5 that showed *cis* effects. We note that we identified 2 additional promoter knockouts (Meg3 and Snhg3) that apparently reduced the expression of a neighboring gene. However, inspection of these loci revealed that the effects in these loci resulted from read-through transcription that continued past the annotated 3' end of the knocked-out gene into a second gene downstream on the same strand. In the Meg3 locus, for example, promoter knockout eliminated the expression of not only Meg3 but also two downstream lncRNAs (Rian and Mirg). Although currently annotated as separate genes[13-16], these three lncRNAs appear to originate from a single promoter and transcript (see Methods, **Extended Data Fig. 3a**)[17]. In the Snhg3 locus, we found an alternative RNA isoform that continued past the annotated 3' end of Snhg3 and sometimes spliced into the downstream Rcc1 mRNA (see Methods, **Extended Data Fig. 3b**), likely explaining the reduction in Rcc1 expression upon Snhg3 promoter knockout. Indeed, RefSeq annotates a joint SNHG3-RCC1 RNA isoform in the corresponding locus in human[18]. We did not consider Meg3 and Snhg3 as potential *cis* regulators in subsequent analysis, although this phenomenon represents another potential mechanism for coordinating gene expression within local neighborhoods.

Note S5. Proposed mechanisms for *cis* regulatory effects in Rcc1, linc2025, and Sfmbt2 loci

For each gene locus that affected a neighbor upon deleting the promoter, we attempted to insert a pAS downstream of the promoter. While we successfully obtained functional pAS insertions for most loci, we did not obtain functional pAS insertions for Rcc1, linc2025, and Sfmbt2. For Rcc1 and Sfmbt2, we did not obtain clones containing insertions, likely due to inefficiencies in the specific sgRNAs and/or homology arms used. In the linc2025 locus, we did obtain clones containing pAS insertions, but the pAS insertions did not fully halt transcription of the lncRNA, making the results uninformative. For the Rcc1 locus, we have additional information that helps to distinguish potential mechanisms. We discuss each of these 3 loci below.

Rcc1 locus. When we deleted the promoter of Rcc1, we observed a ~18% *increase* in the expression of Trnau1ap, whose promoter is located ~2.4 kb downstream of the endogenous pAS of Rcc1. Thus, in contrast to the enhancer-like functions we describe in other loci, the promoter of Rcc1 appears to have a repressive function on its neighboring gene. Although we did not obtain clones with pAS insertions at the Rcc1 locus, we did obtain and characterize a promoter inversion clone in which the promoter-proximal sequences were inverted on the 129 allele (see Methods), thereby eliminating Rcc1 expression on that allele while maintaining the presence of the promoter sequences. This inversion clone provided an alternative approach to distinguish whether the *cis* effect on Trnau1ap expression requires transcription of Rcc1 or whether the *cis* effect depends only on the promoter-proximal sequences. While in promoter deletion clones the expression of Trnau1ap increased in an allele-specific manner, in the promoter inversion clone Trnau1ap expression was unaffected. This suggests that transcription through the Rcc1 locus and the Rcc1 mRNA are dispensable for the effect on Trnau1ap. We expect that the most likely explanation for these observations is that the repressive effect is mediated by the Rcc1 promoter, similar to previous reports of "promoter competition" in which two promoters compete for shared local activating signals such as distal enhancers[19,20].

Linc2025 locus. When we deleted the promoter of linc2025, we observed a ~47% increase in the expression of Chd2, whose promoter is located ~2 kb downstream of the endogenous pAS of linc2025. To attempt to determine whether this *cis* effect requires transcription in the linc2025 locus, we inserted a pAS signal into the first intron using the same strategy as with other lncRNA loci. However, in the linc2025 locus, even a triple pAS cassette (see Methods) did not fully halt transcription and production of the mature lncRNA sequence. Specifically, the level of linc2025 RNA upstream of the pAS was comparable to that of wild-type cells and the level of linc2025 RNA downstream of the pAS was reduced by at most 60% (data not shown). In these linc2025 pAS insertion clones, Chd2 expression was unaffected. Because we did not fully eliminate transcription in the linc2025 locus or expression of the linc2025 RNA, we cannot definitively assess the mechanism by which deleting the linc2025 promoter affects Chd2 expression. Given our results in the Rcc1 locus (see above), we speculate that this repressive effect results from competition between the linc2025 and Chd2 promoters for a shared enhancer.

Sfmbt2 locus. Sfmbt2 has two alternative promoters in mESCs, located 5 kb and 7 kb from the Blustr promoter. We deleted sequence including the 5'-most TSS of Sfmbt2. Deletion of this TSS led to a 70% reduction in overall Sfmbt2 expression (as measured in the downstream exons) and also reduced Blustr expression by 18%. We were unable to obtain pAS insertions downstream of this promoter, likely due to inefficiencies in the specific sgRNAs or homology arms used for these experiments. We suspect that the most likely explanation for this *cis* effect mediated by the Sfmbt2 locus would involve promoter-proximal DNA elements, similar to the other promoters examined.

Note S6. Mechanism for up-regulation of Blustr upon deletion of intron 1

Upon deleting 19.2 kb from the first intron of Blustr, we unexpectedly observed a ~5-fold up-regulation of Blustr (**Fig. 3b**). We suspect that the increase in Blustr RNA expression (**Fig. 3b**) and transcription (**Fig. 3c**) results from an increase in splicing efficiency in the modified locus. Previous studies have shown that splicing can have a dramatic (up to 100-fold) activating effect on transcriptional activity[21-24], and we show that deleting the 5' splice site in this locus indeed results in a strong decrease in Blustr transcription (**Fig. 3c**). The deletion in the first intron of Blustr leads to a closer juxtaposition of the first splice donor and acceptor (separated by 0.5 kb after the deletion compared to 19.7 kb before the deletion), and may thereby increase splicing-mediated feedback. We note that an alternative possibility is that the Blustr intron contains a repressive DNA element whose removal leads to an up-regulation of both Blustr and Sfmbt2. Regardless of the precise mechanism for the increase in Blustr transcription, subsequent pAS insertion experiments demonstrate that changing the length/amount of Blustr transcriptional activity affects Sfmbt2 transcription (**Fig. 3b**).

Note S7. Mechanisms for how transcription of Blustr activates Sfmbt2

Our genetic manipulations in the Blustr locus indicate that Blustr activates Sfmbt2 through a mechanism that requires transcription and splicing but does not depend on specific sequences in the mature RNA transcript. This contrasts with previous examples of loci in which noncoding transcription or noncoding transcripts have been shown to affect gene expression. For example, in the Airn locus, transcription of the lncRNA regulates an overlapping mRNA locus through a "transcriptional interference" mechanism in which transcription directly overlaps the targeted promoter[25]; in contrast, Blustr transcription activates Sfmbt2 but does not overlap it. LncRNAs like Xist[26] encode specific sequences that are required for their functions in silencing gene expression; in contrast, Blustr does not require specific sequences in the mature transcript to activate its neighbor.

We conclude that Blustr-mediated activation of Sfmbt2 involves "the process of transcription", by which we mean one or more co-transcriptional processes or associated factors that are recruited irrespective of the precise sequence of the RNA — in other words, processes that we expect to find at essentially all transcribed genes. These processes or factors potentially include chromatin regulators, splicing factors, pause release factors, RNA polymerase itself, or other RNA binding proteins that interact with many nascent transcripts through degenerate recognition sites. Increasing the local concentration of one or more of these factors is likely required to activate the Sfmbt2 promoter (**Fig. 3c**), similar to previously proposed mechanisms[27-30]. We note that this is plausible inasmuch as Blustr is ~5-fold more highly transcribed than Sfmbt2 in mESCs (**Fig. 3a**). Indeed, we found that altering transcription or splicing in the Blustr locus led to changes in chromatin state and reduced occupancy of engaged RNA polymerase in the paused position just downstream of the Sfmbt2 TSS (**Extended Data Fig. 8e,f**). Interestingly, the Sfmbt2 promoter is bivalently marked by H3K4me3 and H3K27me3 in mESCs, and deletion of the Blustr promoter or splice site led to the spreading of the repression-associated H3K27me3 modification across ~30 kb in this locus (**Extended Data Fig. 8f**). Thus, changes in Blustr transcription and splicing may affect Sfmbt2 expression in part by altering chromatin state and RNA polymerase occupancy at the Sfmbt2 promoter (**Fig. 3d**). We note that, for both the promoter deletion and 5' splice site knockouts, Sfmbt2 p(A)+ RNA levels dropped by 85-90% (**Fig. 3b**) whereas GRO-Seq measurements showed a more modest 60% decrease (**Extended Data Fig. 8b**). Because we examined only one clonal cell line for each of these GRO-seq experiments, it is possible that this difference simply results from clonal variation in Sfmbt2 expression levels; however, it is also possible that Blustr could regulate Sfmbt2 at both transcriptional and post-transcriptional steps.

We note that while each exon and intron of the Blustr is individually dispensable for activation of Sfmbt2, we do not exclude redundant or otherwise sequence non-specific functions of the RNA transcript itself. Indeed, the process of splicing, which we show is required for Sfmbt2 activation, involves interactions between the spliceosome and the nascent Blustr RNA, implicating the RNA transcript in this regulatory activity. Beyond the spliceosome, many RNA-

binding proteins accompany the transcriptional machinery and interact promiscuously with nascent transcripts, and some of these proteins are thought to have dual roles in RNA processing and transcriptional regulation[31-34]. Because these ubiquitous factors bind to many nascent transcripts with minimal sequence specificity, we regard these as components of the "process of transcription" referred to in the main text. As demonstrated by the splicing of Blustr, these general factors present at many gene loci may play important roles in local gene regulation.

Note S8. Generality of local regulatory effects by promoters, transcription, and splicing

In total, we find that 9 of 18 loci — including both lncRNAs and mRNAs — affect the expression of a neighboring gene through mechanisms involving their promoters, transcription, and/or splicing. Although these 18 loci were chosen subject to some constraints, the selection criteria were broad enough that there are at least thousands of other loci that display similar properties (**Extended Data Fig. 1**). Here we discuss the selection criteria and how they might affect the generalizability of our findings.

(i) We selected lncRNAs previously defined by a chromatin signature of H3K4me3 at promoters and H3K36me3 throughout gene bodies[35] — this chromatin signature is found at thousands of other lncRNAs[36]. Other reported sets of lncRNAs include transcripts that do not display these chromatin properties, but in many cases these transcripts have promoters displaying the marks of enhancers (*e.g.*, H3K4me1) and thus we expect that similar mechanisms are present in some of these loci.

(ii) Although we initially selected lncRNAs and mRNAs that span a range of abundance levels, we found that loci that affect the expression of a neighboring gene produce transcripts spanning a range of abundance levels (**Extended Data Fig. 1a**).

(iii) We selected lncRNAs from the 50% with subcellular localization most biased toward the nucleus (**Extended Data Fig. 1b**). However, we find that even mRNA loci (whose RNA localization is biased toward the cytoplasm) can affect the expression of neighboring genes, and so we do not expect this subcellular localization criterion to limit the generality of these mechanisms.

(iv) We chose our lncRNA set to include some that are conserved across mammalian evolution (Snhg3, Snhg17, Meg3, and linc2025), but we found that both lncRNAs that are conserved and those that are not conserved can affect the expression of a neighboring gene (*e.g.*, the spliced Blustr transcript is found only in mouse, while Snhg17 is deeply conserved).

(v) We selected lncRNAs and mRNAs whose TSSs are located at least 5 kb from another gene. This applies to ~60% of all H3K4me3-marked lncRNAs in mESCs. We anticipate that for the remaining lncRNAs originating closer to mRNA genes — for example, those originating divergently from the same promoter as an mRNA[37], or for the subset of lncRNAs that overlap another gene — crosstalk as a result of promoters, transcription, or splicing may be even more prevalent than for the set we investigated.

Accordingly, we anticipate that similar *cis* regulatory effects will be found at many other gene loci.

Note S9. Evolutionary conservation of lncRNAs and their promoters

We used evolutionary conservation analysis to explore whether other lncRNA loci might similarly encode *cis* regulatory functions independent of their RNA transcripts. To identify candidates, we searched for loci where the lncRNA promoter is conserved across mammalian evolution even when the lncRNA transcript itself is not. When we examined the 209 mESC lncRNA loci whose transcripts are "mouse-specific" (no syntenic transcript detected in rat, chimp or human pluripotent stem cells, see Methods)[38], we found that their promoters, as a whole, showed signatures of sequence constraint (**Extended Data Fig. 10a**). Most of this signal is driven by the 71 lncRNAs that emerge from bidirectional mRNA promoters (**Extended Data Fig. 10a**); in these loci, the promoter clearly has a conserved function in regulating the mRNA that is independent of the presence of the lncRNA transcript. While some of these divergent lncRNAs may have evolved neutrally, transcription or splicing in these loci could feed back to regulate the promoter and thus the expression of the mRNA. Of the remaining 136 loci, 11 appear to have evolved from ancestral enhancers: the sequence serving as a promoter in mouse is adjacent to the same genes in mouse and human and corresponds to a conserved DNA element marked in human embryonic stem cells by a chromatin signature associated with enhancers (**Fig. 4b-d**, **Extended Data Fig. 10b**, see Methods). These results suggest that these sequences have deeply conserved roles as *cis* regulatory elements, although not as lncRNA promoters. Finally, 59 mouse-specific lncRNAs emerge from lineage-specific endogenous retroviruses (**Extended Data Fig. 10a**), which contain their own promoters and 5' splice sites that drive transcription of sequences downstream of the insertion to produce noncoding transcripts[39]. These classes of retroelements are known to have the capacity to act as enhancers in mESCs[40], suggesting the potential for *cis* regulatory function despite their recent introduction in the mouse lineage. Together, these observations highlight many lncRNA promoters that are likely to have *cis* regulatory functions independent of the lncRNA transcripts themselves, and also suggest that lncRNAs may frequently evolve from pre-existing regulatory elements including promoters and enhancers (**Extended Data Fig. 10c**).

References:

1. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477,** 295–300 (2011).
2. Yoon, J. K., Olson, E. N., Arnold, H. H. & Wold, B. J. Different MRF4 knockout alleles differentially disrupt Myf-5 expression: cis-regulatory interactions at the MRF4/Myf-5 locus. *Dev. Biol.* **188,** 349–362 (1997).
3. Zuniga, A. *et al.* Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev* **18,** 1553–1564 (2004).
4. Dimitrova, N. *et al.* LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint. *Mol Cell* **54,** 777–790 (2014).
5. Singer-Sam, J., Chapman, V., LeBon, J. M. & Riggs, A. D. Parental imprinting studied by allele-specific primer extension after PCR: paternal X chromosome-linked genes are transcribed prior to preferential paternal X chromosome inactivation. *Proc Natl Acad Sci U S A* **89,** 10469–10473 (1992).
6. Cowles, C. R., Hirschhorn, J. N., Altshuler, D. & Lander, E. S. Detection of regulatory variation in mouse genes. *Nat Genet* **32,** 432–437 (2002).
7. Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. Allelic variation in human gene expression. *Science* **297,** 1143–1143 (2002).
8. Nora, E. P., Dekker, J. & Heard, E. Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? *Bioessays* **35,** 818–828 (2013).
9. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489,** 109–113 (2012).
10. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
11. Eckersley-Maslin, M. A. *et al.* Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev Cell* **28,** 351–365 (2014).
12. Gendrel, A.-V. *et al.* Developmental dynamics and disease potential of random monoallelic gene expression. *Dev Cell* **28,** 366–380 (2014).
13. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420,** 563–573 (2002).
14. Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296,** 916–919 (2002).
15. Bertone, P. *et al.* Global identification of human transcribed sequences with genome tiling arrays. *Science* **306,** 2242–2246 (2004).
16. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309,** 1559–1563 (2005).
17. Tierling, S. *et al.* High-resolution map and imprinting analysis of the Gtl2-Dnchc1 domain on mouse chromosome 12. *Genomics* **87,** 225–235 (2006).
18. Pruitt, K. D. *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42,** D756–63 (2014).
19. Choi, O. R. & Engel, J. D. Developmental regulation of beta-globin gene switching. *Cell* **55,** 17–26 (1988).
20. Ohtsuki, S., Levine, M. & Cai, H. N. Different core promoters possess distinct regulatory activities in the Drosophila embryo. *Genes Dev* **12,** 547–556 (1998).
21. Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E. & Palmiter, R. D. Introns increase transcriptional efficiency in transgenic mice. *Proc Natl Acad Sci U S A* **85,** 836–840 (1988).
22. Fong, Y. W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414,** 929–933 (2001).
23. Furger, A., O'Sullivan, J. M., Binnie, A., Lee, B. A. & Proudfoot, N. J. Promoter proximal splice sites enhance transcription. *Genes Dev* **16,** 2792–2799 (2002).

24. Damgaard, C. K. *et al.* A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell* **29,** 271–278 (2008).
25. Latos, P. A. *et al.* Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338,** 1469–1472 (2012).
26. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30,** 167–174 (2002).
27. Travers, A. Chromatin modification by DNA tracking. *Proc Natl Acad Sci U S A* **96,** 13634–13637 (1999).
28. Yoo, E. J., Cooke, N. E. & Liebhaber, S. A. An RNA-independent linkage of noncoding transcription to long-range enhancer function. *Mol Cell Biol* **32,** 2020–2029 (2012).
29. Hirota, K. *et al.* Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* **456,** 130–134 (2008).
30. Fromm, G. *et al.* An embryonic stage-specific enhancer within the murine β-globin locus mediates domain-wide histone hyperacetylation. *Blood* **117,** 5207–5214 (2011).
31. Änkö, M.-L. *et al.* The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol* **13,** R17 (2012).
32. Miyagawa, R. *et al.* Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* **18,** 738–751 (2012).
33. Sanford, J. R. *et al.* Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* **19,** 381–394 (2009).
34. Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39,** 925–938 (2010).
35. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).
36. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25,** 1915–1927 (2011).
37. Sigova, A. A. *et al.* Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A* **110,** 2876–2881 (2013).
38. Chen, J. *et al.* Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol* **17,** 19 (2016).
39. Kelley, D. & Rinn, J. L. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* **13,** R107 (2012).
40. Rowe, H. M. *et al.* TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res* **23,** 452–461 (2013).