linc-HOXA1 is a noncoding RNA that represses *Hoxa1* transcription in *cis*

Hédia Maamar,[1] Moran N. Cabili,[2,3,4] John Rinn,[2,4] and Arjun Raj[1,5]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; [2]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02142, USA; [3]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA; [4]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA

Recently, researchers have uncovered the presence of many long noncoding RNAs (lncRNAs) in embryonic stem cells and believe they are important regulators of the differentiation process. However, there are only a few examples explicitly linking lncRNA activity to transcriptional regulation. Here, we used transcript counting and spatial localization to characterize a lncRNA (dubbed *linc-HOXA1*) located ~50 kb from the *Hoxa* gene cluster in mouse embryonic stem cells. Single-cell transcript counting revealed that *linc-HOXA1* and *Hoxa1* RNA are highly variable at the single-cell level and that whenever *linc-HOXA1* RNA abundance was high, *Hoxa1* mRNA abundance was low and vice versa. Knockdown analysis revealed that depletion of *linc-HOXA1* RNA at its site of transcription increased transcription of the *Hoxa1* gene *cis* to the chromosome and that exposure of cells to retinoic acid can disrupt this interaction. We further showed that *linc-HOXA1* RNA represses *Hoxa1* by recruiting the protein PURB as a transcriptional cofactor. Our results highlight the power of transcript visualization to characterize lncRNA function and also suggest that PURB can facilitate lncRNA-mediated transcriptional regulation.

[*Keywords*: gene regulation; noncoding RNA; single molecule]

Supplemental material is available for this article.

Received March 6, 2013; revised version accepted May 6, 2013.

With the advent of tiling microarrays and RNA sequencing, researchers have discovered that transcription occurs in many regions of the genome not previously thought to be transcriptionally active (Carninci et al. 2005; Kapranov et al. 2005; The ENCODE Project Consortium et al. 2007; Mercer et al. 2009; Cabili et al. 2011; Rinn and Chang 2012; Derrien et al. 2012). In particular, some of these regions (thought to number in the thousands) (Cabili et al. 2011; Derrien et al. 2012) encode long noncoding RNA molecules (lncRNAs) with little apparent coding potential, despite being spliced and polyadenylated much like messenger RNA (mRNA). While many lncRNAs display conservation between species (Ponting et al. 2009), suggesting that they may play some role in cellular function, the exact role of the vast majority of these molecules remains unclear, often due to technical difficulties associated with the manipulation of lncRNAs. However, associative studies show that lncRNAs appear to be involved in developmental processes (Pauli et al. 2011; Hu et al. 2012) and that they are particularly prevalent in embryonic stem cells and may be involved in the regulation of pluripotency (Dinger et al. 2008; Hawkins and Morris 2010; Guttman et al. 2011).

The four *Hox* gene clusters in mammalian genomes provide an ideal testing ground for studying the function of lncRNAs given that (1) they contain key developmental regulators with broad conservation; (2) their genetic organization is strongly linked to their expression properties, providing candidates for regulation; and (3) they contain a very large number of lncRNAs (Petruk et al. 2006; Mainguy et al. 2007; Rinn et al. 2007; Zhang et al. 2009; Wang et al. 2011). Indeed, several groups have already demonstrated that particular lncRNAs in the cluster appear to regulate some of the Hox genes in the cluster themselves (Petruk et al. 2006; Rinn et al. 2007; Zhang et al. 2009; Wang et al. 2011), part of a broader class of lncRNAs that regulate genetically proximal genes (Ørom et al. 2010; Ørom and Shiekhattar 2011; Wang et al. 2011; Lai et al. 2013). Still, the function (if any) of the majority of lncRNAs in and around the Hox clusters remains unknown.

Here, we elucidated the function of a lncRNA, dubbed *linc-HOXA1*, located ~50 kb away from the *Hoxa1* gene in the Hoxa gene cluster in mice through the use of single-molecule RNA imaging and single-cell analysis. While bulk assays averaging together many cells treated with retinoic acid initially suggested that *linc-HOXA1*

[5]**Corresponding author**
E-mail arjunraj@seas.upenn.edu
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.217018.113.

RNA may activate *Hoxa1* transcription, single-molecule transcript counting revealed that individual cells can have high abundance of *linc-HOXA1* RNA or *Hoxa1* mRNA but never both, suggesting instead that *linc-HOXA1* RNA serves to transcriptionally repress *Hoxa1*. We confirmed this negative regulation by knocking down *linc-HOXA1* RNA at the site of its transcription, demonstrating that the regulation appears to occur in *cis* to the chromosome rather than through a *trans* mechanism. We further demonstrated that this effect is due to a particular region of the RNA that recruits the transcriptional regulator PURB to the site of transcription. We believe our results highlight the potential for single-cell analysis in uncovering the gene regulatory roles of lncRNAs.

Results

Identification and characterization of the linc-HOXA1 *transcript*

Previous studies have found regulatory lncRNAs flanking the *Hoxa* locus in mice—one on the 5′ end of the cluster located beyond *Hoxa13* (Zhang et al. 2009; Wang et al. 2011) and one near the 3′ end of the cluster located between *Hoxa1* and *Hoxa2* (Zhang et al. 2009). However, deep sequencing results in mouse embryonic stem cells (Guttman et al. 2010) have revealed the existence of another lncRNA (*linc-HOXA1*) located ~50 kb 3′ of *Hoxa1* (Fig. 1A). *linc-HOXA1* is transcribed in the opposite direction (5′→3′) of Hoxa1, and the gene is ~12 kb long. Cloning of the *linc-HOXA1* transcripts uncovered the presence of three different isoforms (Fig. 1A), consistent with RNA sequencing studies (Guttman et al. 2010).

We then sought to establish that *linc-HOXA1* encodes a noncoding RNA transcript. While the transcript did not contain any ORFs longer than 151 codons, recent studies have suggested that several lncRNAs may in fact encode short peptides (Ingolia et al. 2011). To rule out this possibility, we analyzed the *linc-HOXA1* coding sequence using phyloCSF (phylogenetic codon substitution frequency) (Lin et al. 2011), a conservation-based method that estimates whether a multispecies nucleotide sequence alignment in a specific locus is more likely to represent a protein coding than a noncoding transcript (Materials and Methods). The low phyloCSF score (9.2) of *linc-HOXA1* is more in line with those of other known noncoding RNAs than with those of known protein coding transcripts (Fig. 1B). To further check for coding potential in *linc-HOXA1*, we overexpressed each isoform in HeLa cells at levels considerably higher than GAPDH and analyzed all of the resulting small peptides via mass spectrometry (Materials and Methods; Fig. 1C; Supple-

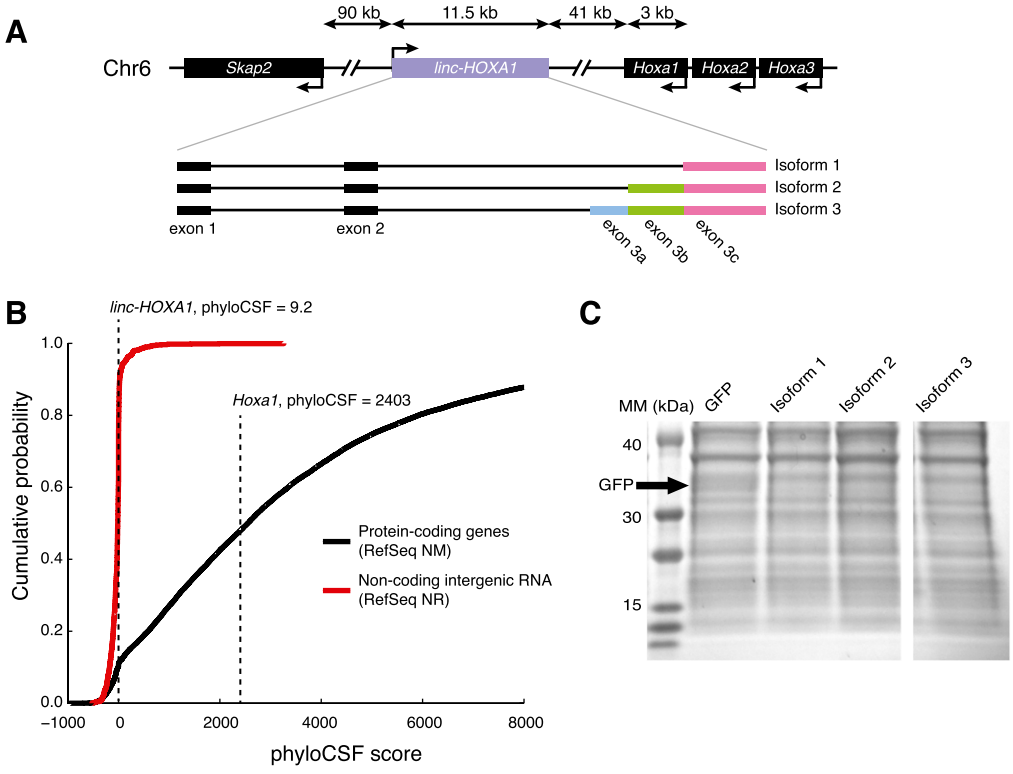

Figure 1. Identification and characterization of *linc-HOXA1* as a noncoding RNA. (*A*) Illustration of the genomic location and transcript structure of the *linc-HOXA1* gene and its three isoforms. (*B*) Cumulative probability distribution of coding potential as measured by phyloCSF score (Materials and Methods) for both noncoding transcripts (RefSeq NR, red) and coding transcripts (RefSeq NM, black). Dashed lines represent the coding potential of *linc-HOXA1* and *Hoxa1*. (*C*) Protein gel stained with Coomassie blue upon overexpression of GFP (control) or *linc-HOXA1* isoforms 1, 2, and 3 in HeLa cells. Markers at the *left* of the gel indicate protein size, and the arrow shows the expected size of the GFP protein.

mental Fig. 1). We did not find any peptide matches to those encoded by short ORFs in any of the isoforms we cloned, indicating that these transcripts do not encode proteins. We also note here that the subsequent functional studies we outline below provide further evidence that *linc-HOXA1* does not encode functional small peptides.

Single-molecule detection of linc-HOXA1 *isoforms in individual cells*

In order to detect and quantify the different isoforms of *linc-HOXA1* RNA in individual cells, we generated sets of oligonucleotide probes for RNA fluorescence in situ hybridization (RNA FISH) (Raj et al. 2008; Raj and Tyagi 2010) that were specific to each isoform (Supplemental Table 1). By looking for colocalization of signal from probes targeting various parts of the different isoforms, we were able to detect and quantify all the three isoforms at the level of single molecules in individual cells (Fig. 2A). We found that isoform 1 was the dominant isoform at ~41% of the total, with isoforms 2 and 3 making up the remaining 27% and 8%, respectively (Fig. 2B). We also observed other partial isoforms, but the partial isoforms were only ~24% of the total transcripts detected and may represent partial isoforms undergoing degradation (Fig. 2B). At the single-cell level, we found that each of these isoforms correlated very strongly with each other (Fig. 2C); hence, for simplicity, we used a set of probes optimized to target all isoforms for the rest of the analysis in this study.

Differentiation dynamics and single-cell analysis of linc-HOXA1

The close proximity of *linc-HOXA1* to the *Hoxa* gene cluster along with the fact that the *Hox* gene expression is related to their genetic organization suggested to us that the transcription of *linc-HOXA1* may be tied to that of the nearby *Hox* genes and may in fact play some regulatory role therein (Fig. 3A). We first examined this possi-

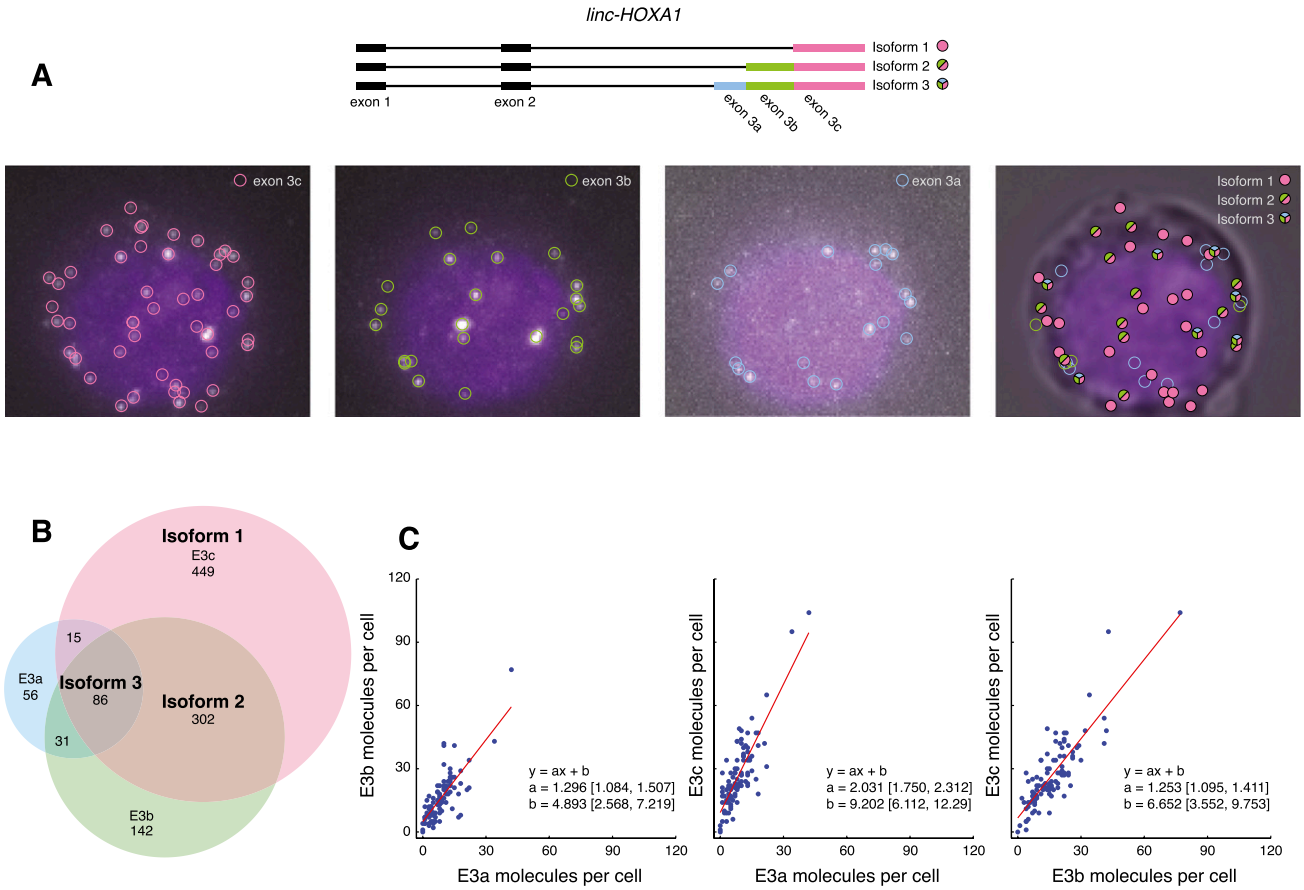

Figure 2. Detection of individual RNA isoforms of *linc-HOXA1* in single cells. (*A*) Raw micrographs of a single cell containing all three isoforms. (Three *left* panels) We designed probes specifically targeting exons 3c, 3b, and 3a, each of which revealed single spots corresponding to individual molecules. The circles represent the computationally identified spot locations (determined in three dimensions). (*Right* panel) We then looked for colocalization of the spots and used that to determine which combinations of spots represented particular isoforms. The unfilled circles in the *right* panel represent spots that we were unable to classify as being isoform 1, 2, or 3. (*B*) Venn diagram showing the relative abundances of the various combinations of exons 3a, 3b, and 3c that we detected through colocalization analysis. Labels indicate the combinations that correspond to the three isoforms identified by sequencing and cloning. (*C*) Pairwise scatter plots showing the numbers of exons 3a, 3b, and 3c versus those same exons in individual cells. Lines represent a least-squares fit with the indicated parameters (including 95% confidence intervals).

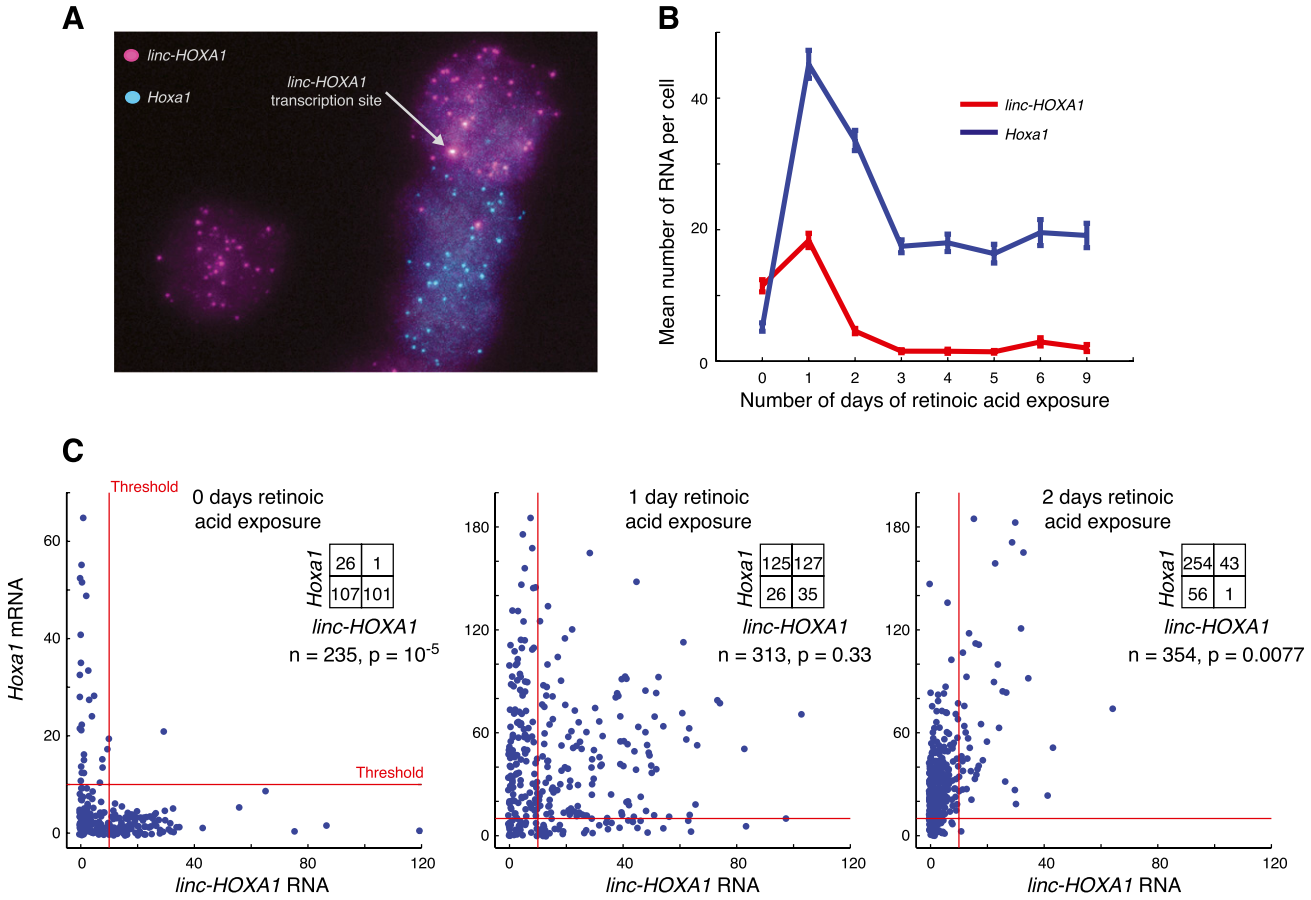

Figure 3. *linc-HOXA1* and *Hoxa1* anti-correlate at the single-cell level. (*A*) Micrographs of three representative mouse embryonic stem cells expressing *linc-HOXA1* (red) and *Hoxa1* (blue). The arrow points to the location of an active transcription site. (*B*) Average numbers of *linc-HOXA1* (red) and *Hoxa1* (blue) RNA molecules per cell in embryonic stem cells exposed to retinoic acid for increasing durations as measured by single-molecule RNA FISH. Error bars represent standard error. The numbers of cells in each condition are 235, 313, 354, 351, 280, 204, 165, and 189 (0–6 and 9 d retinoic acid, respectively). (*C*) Single-cell scatter plots of *linc-HOXA1* and *Hoxa1* RNA abundance at 0, 1, and 2 d of exposure to retinoic acid. The red lines represent thresholds that we chose for *linc-HOXA1* and *Hoxa1* RNA levels, set at 10 RNA molecules for each RNA species. The *inset* tables show the number of cells in the four quadrants delineated by the thresholds. We computed the *P*-value for the distribution of cells in the quadrants via the χ^2 test.

bility through the addition of retinoic acid to our embryonic stem cells, which is an agent known to cause embryonic stem cells to differentiate toward a neural lineage with concomitant activation of *Hoxa* genes (Dupé et al. 1997). We found that exposing the cells to retinoic acid over the course of several days caused a coordinated pulse of expression of *linc-HOXA1*, *Hoxa1* (Fig. 3B), and *Hoxa2* (Supplemental Fig. 2A), peaking at 1 d of exposure. This population-level analysis initially suggested that *linc-HOXA1* and *Hoxa1* might be coherently regulated by either each other or a common upstream factor in the retinoic acid pathway.

We then examined the expression of *linc-HOXA1* and *Hoxa1* RNA in individual cells, reasoning that single-cell correlations or anti-correlations in expression may reveal more information about potential regulatory interactions. We first looked at stem cells that had not been treated with retinoic acid. We found that both *linc-HOXA1* and *Hoxa1* showed large cell-to-cell variability in transcript abundance. Surprisingly, we found that this variability anti-correlated in the sense that cells with a high abundance of *linc-HOXA1* RNA had low numbers of *Hoxa1* mRNA and vice versa; i.e., it was very rare to find individual cells that simultaneously had high levels of *linc-HOXA1* and *Hoxa1* RNA (Fig. 3C). (We also note that the nearby gene *Skap2* did not correlate with *linc-HOXA1* [Supplemental Fig. 2B], showing that correlations and anti-correlations are not generic features of transcription of this gene. In addition, *Oct4* did not correlate with *Hoxa1*, showing that variability in *Hoxa1* was not an artifact of partially differentiating cells [Supplemental Fig. 2C].) This anti-correlation disappeared, however, upon a 1-d exposure to retinoic acid (Fig. 3C), even though the relative abundance of various isoforms were similar (albeit with a modest increase in isoform 3) (Supplemental Fig. 3). Together, these findings demonstrate that while *linc-HOXA1* and *Hoxa1* appear to correlate in bulk population measurements, our single-cell measurements raised the possibility that *linc-HOXA1* RNA may actually negatively regulate *Hoxa1*

transcription or vice versa and that the addition of retinoic acid overrides this negative regulation. Moreover, our single-cell analysis reveals that any putative regulation appears to behave in a sharp threshold-like manner rather than a graded response in that expression of, say, *Hoxa1* drops dramatically once *linc-HOXA1* reaches a particular level (in this case, we chose 10 RNA molecules as a threshold that captures this effect).

Knockdown of linc-HOXA1 *RNA reveals repression of* Hoxa1

In order to determine whether these correlations were manifestations of a true regulatory interaction, we used two different methods to knock down *linc-HOXA1* transcript levels and analyzed the results using RNA FISH (Fig. 4A, top left). The first method used modified antisense oligonucleotides that bind to the RNA and prompt RNA degradation via RNase H activity (Liang et al. 2011). The second method used conventional siRNAs via activation of the RNAi pathway. Both methods resulted in similar decreases in overall *linc-HOXA1* RNA levels (65%, $P = 2 \times 10^{-7}$; 52%, $P = 4 \times 10^{-4}$, respectively). Consistent with *linc-HOXA1* RNA repressing the transcription of *Hoxa1*, we found that knockdown via the antisense oligonucleotides resulted in an increase in *Hoxa1* mRNA abundance by an average of 50% ($P = 2 \times 10^{-4}$) (Fig. 4A). However, we found that knockdown via siRNA did not result in any change in *Hoxa1* mRNA abundance.

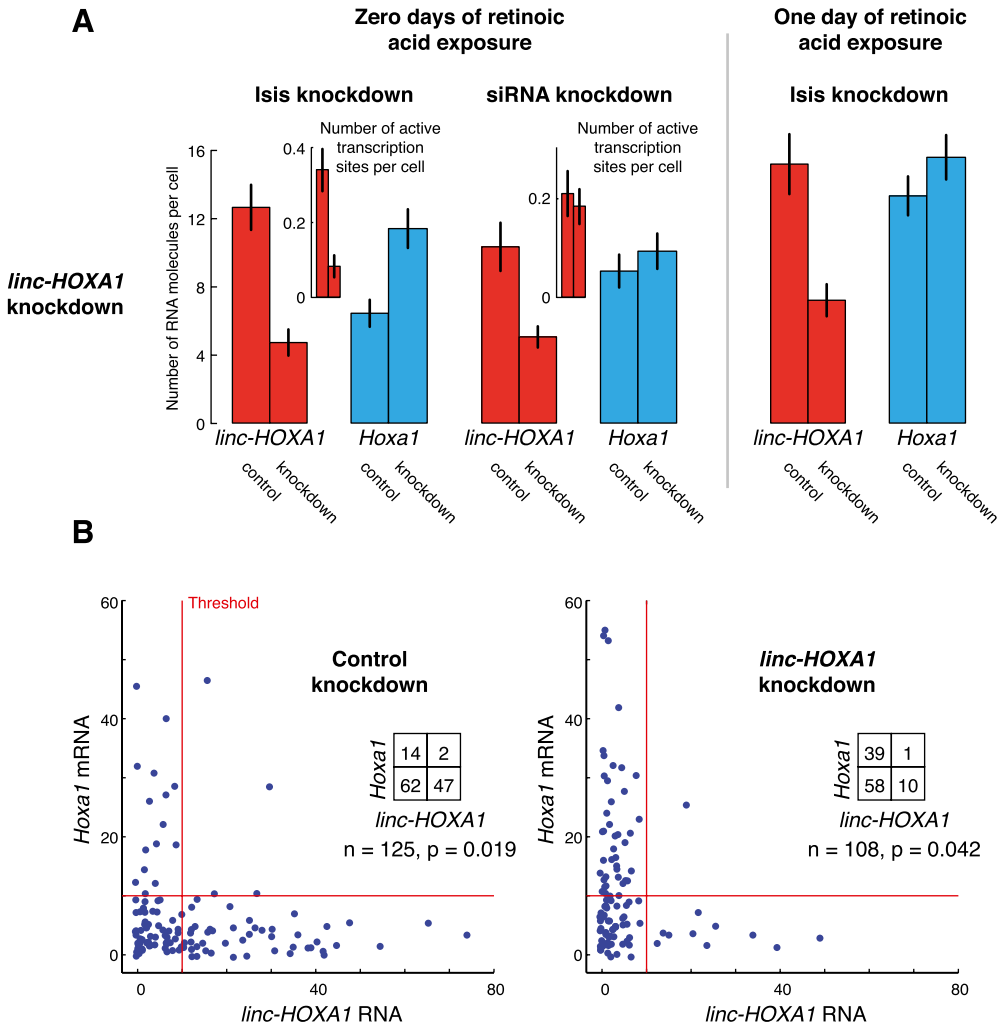

Figure 4. Knockdown of *linc-HOXA1* increases *Hoxa1* levels. (*A*) Bar graphs showing changes in average number of *linc-HOXA1* and *Hoxa1* RNA molecules per cell upon treatment with siRNA or Isis antisense oligonucleotides targeting either *linc-HOXA1* (*top* row) or *Hoxa1* (*bottom* row). In all bar graph pairs, the control condition (targeting with a nonspecific siRNA or Isis antisense oligonucleotides) is on the *left*. The *inset* bar graph shows the average number of active transcription sites in cells treated with control versus *linc-HOXA1* knockdown. All numbers represent counts obtained by single-molecule RNA FISH; all error bars represent standard error of the mean. (*B*) Single-cell scatters showing *linc-HOXA1* versus *Hoxa1* transcript levels upon treatment of cells with control (nonspecific) antisense oligonucleotides or antisense oligonucleotides targeting *linc-HOXA1*. The red lines represent the thresholds of 10 RNA molecules for each RNA species and are the same as those in Figure 3. The *inset* tables show the number of cells in the four quadrants delineated by the thresholds. We computed the *P*-value for the distribution of cells in the quadrants via the χ^2 test.

To find the source of this discrepancy, we analyzed the knockdown of *linc-HOXA1* RNA at the subcellular level. In untreated cells, we found *linc-HOXA1* in both the nuclear and cytoplasmic compartments. The nuclear RNA appeared primarily as bright foci marking the site of transcription (Fig. 3A); these foci represent a pileup of nascent transcripts at the gene locus itself, as verified by colocalization of these foci with probes targeting the introns of *linc-HOXA1* (Supplemental Fig. 4; Fremeau et al. 1986; Xing et al. 1993; Levesque and Raj 2013). Most of the rest of the RNA were in the cytoplasm, as verified by three-dimensional analysis of spot locations (data not shown), although it was difficult to make this assessment rigorously due to the thin layer of cytoplasm in this cell type. We thus checked whether the knockdown was effectively knocking down RNA at the site of *linc-HOXA1* transcription, given that any potential *cis* regulatory effects of *linc-HOXA1* (meaning activity of *linc-HOXA1* on genes located near its site of transcription) would occur at this physical location. We found that the number of sites of active transcription of *linc-HOXA1* per cell decreased markedly upon treatment with antisense oligonucleotides but were relatively invariant to targeting via siRNA (Fig. 4A, inset; see Supplemental Fig. 5 for representative images). These results indicate that repression of *Hoxa1* by *linc-HOXA1* RNA depends primarily on the presence of *linc-HOXA1* RNA molecules at the site of transcription rather than the overall cellular abundance of *linc-HOXA1* RNA, strongly suggesting that *linc-HOXA1* regulates *Hoxa1* through a mechanism mediated by the genetic proximity of these two genes. We note, however, that antisense oligonucleotide knockdown did not affect *Hoxa2* expression (1.13 ± 0.19 mRNA per cell in the control condition to 1.20 ± 0.16 mRNA per cell in the knockdown).

To examine the results of our knockdown experiment at the single-cell level, we grouped our cells into three populations: those with high levels of *linc-HOXA1* (and low levels of *Hoxa1* mRNA), those with high levels of *Hoxa1* (and low levels of *linc-HOXA1*), and those with low levels of both *linc-HOXA1* and *Hoxa1* (Fig. 4B). We found that the apparent effect of the knockdown with antisense oligonucleotides was to shift cells from the high *linc-HOXA1*/low *Hoxa1* group to the high *Hoxa1*/low *linc-HOXA1* group, with few if any cells moving into the low *linc-HOXA1*/low *Hoxa1* group. This result suggests that the fold increase in *Hoxa1* mRNA levels in those cells for which *linc-HOXA1* levels decreased was considerably higher than the population averages would indicate. In this case, single-cell analysis reveals that both cellular heterogeneity and the presence of a sharp threshold for repression serve to diminish the amplitude of the effects of *linc-HOXA1* RNA knockdown at the population level.

Our correlation results also showed that the correlation between *linc-HOXA1* and *Hoxa1* mRNA levels disappeared upon addition of retinoic acid. We tested whether these correlations were a sign of a change in the functional interaction of these genes by knocking down *linc-HOXA1* after 1 d of retinoic acid exposure. We found that knocking down *linc-HOXA1* at this point no longer affected *Hoxa1* levels, showing that activation by retinoic acid overrides the repressive function of *linc-HOXA1* (Fig. 4A). Also, to verify that the anti-correlation between *linc-HOXA1* and *Hoxa1* was not due to the transcription factor activity of HOXA1, we also knocked down *Hoxa1* mRNA both with and without addition of retinoic acid, finding that the knockdown had no effect on *linc-HOXA1* RNA abundance (Supplemental Fig. 6).

Overexpression of linc-HOXA1 *RNA does not change* Hoxa1 *levels*

Our knockdown experiments demonstrate that *linc-HOXA1* RNA negatively affects the transcription of *Hoxa1*, and our transcription site analysis strongly suggests that *linc-HOXA1* activity takes place in *cis*—meaning locally, near its site of transcription. In contrast, a *trans* mechanism of action would consist of *linc-HOXA1* RNA moving within the nucleus to the *Hoxa1* gene regardless of where *linc-HOXA1* was originally transcribed. To confirm the *cis* activity and eliminate the possibility of *trans* activity, we transiently overexpressed each isoform of *linc-HOXA1* in embryonic stem cells to look for changes in *Hoxa1* transcription. Despite a ninefold or greater overexpression of the three *linc-HOXA1* isoforms, we did not observe any changes in *Hoxa1* mRNA abundance (Fig. 5A). At the single-cell level, the anti-correlation between *linc-HOXA1* and *Hoxa1* disappeared (Fig. 5B), directly revealing that cells with high levels of *Hoxa1* RNA (and low initial levels of *linc-HOXA1*) can maintain those high levels of *Hoxa1* RNA even upon addition of exogenous *linc-HOXA1* RNA to those same cells. Adding *linc-HOXA1* to cells that initially had very little *linc-HOXA1* precludes the possibility that the lack of change in *Hoxa1* mRNA levels is due to our merely increasing (to nonphysiologically relevant levels) *linc-HOXA1* RNA numbers in cells that already had high amounts of *linc-HOXA1* to begin with.

Combining these results with our *linc-HOXA1* knockdown results, we conclude that the regulation of *Hoxa1* by *linc-HOXA1* occurs largely due to the local activity of *linc-HOXA1* near its site of transcription. We also note that these results provide further evidence that the *linc-HOXA1* transcripts do not encode a protein (including any short peptides) that mediates its transcriptional effects: In such a scenario, overexpression would result in an increase in proteins and thus would have resulted in a *trans* repression of the *Hoxa1* mRNA production.

linc-HOXA1 *exerts its repressive effects through interaction with PURB*

There are several examples of lncRNAs that alter gene expression via the recruitment of proteins involved in transcriptional regulation (Rinn et al. 2007; Lee 2012; Hawkins and Morris 2010; Huarte et al. 2010). To check what proteins *linc-HOXA1* RNA may be associated with, we biotinylated in vitro synthesized RNA corresponding to RNA containing a variety of different exons from *linc-HOXA1* (Fig. 6A) and used them to pull out proteins

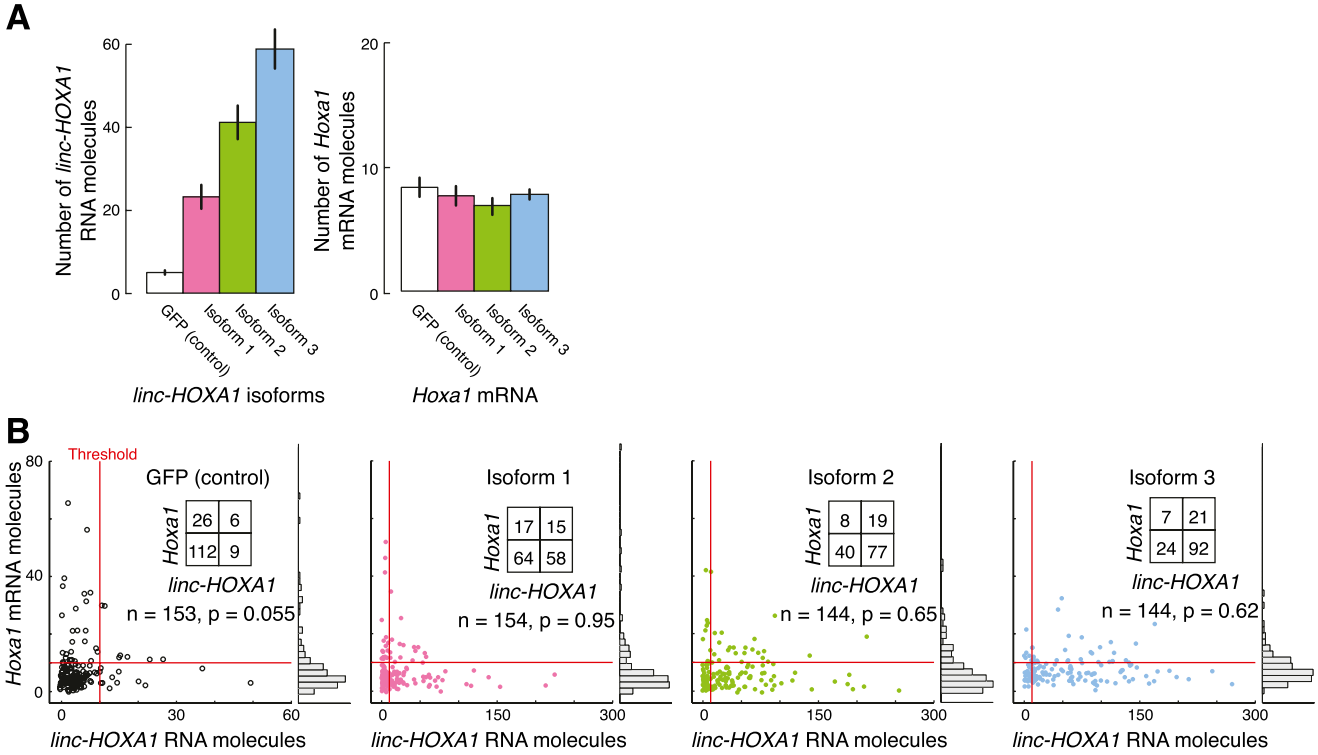

Figure 5. Overexpression of *linc-HOXA1* isoforms does not alter *Hoxa1* abundance. (*A*) Bar plots showing the number of the various *linc-HOXA1* RNA isoforms overexpressed in embryonic stem cells (and a vector containing GFP as a control) (*left*) and the resultant number on *Hoxa1* transcripts (*right*). Overexpression is likely underestimated due to high levels of expression, resulting in some spot undercounting. All error bars represent standard error of the mean. (*B*) Single-cell scatter plots showing *linc-HOXA1* versus *Hoxa1* transcript levels upon overexpression of GFP or each of the three *linc-HOXA1* isoforms. The red lines represent the thresholds of 10 RNA molecules for each RNA species and are the same as those in Figures 3 and 4. The *inset* tables show the number of cells in the four quadrants delineated by the thresholds. We computed the *P*-value for the distribution of cells in the quadrants via the χ^2 test.

from embryonic stem cell lysate (Fig. 6B). We separated these proteins by electrophoresis and looked for bands with differential intensity between the antisense RNA controls and the various synthetic RNAs used. We found a prominent band (Fig. 6B) and used mass spectrometry to identify the proteins that bound specifically to various isoforms of *linc-HOXA1*. Of particular interest was the protein encoded by *Purb*, a known modulator of transcription (Ramsey and Kelm 2009); indeed, experiments show that a related protein, PURA, which often binds together with PURB, can mediate transcriptional regulation via lncRNA (Hawkins and Morris 2010). PURB binds to purine-rich tracts of single-stranded nucleic acids (Bergemann and Johnson 1992; Gallia et al. 1999), and mass spectrometry revealed that PURB was bound only to RNA that contained the long stretch of guanines and adenosines in the middle of exon 3b (Fig. 6B).

In order to verify whether this interaction occurred inside of embryonic stem cells, we wanted to use an antibody against PURB to pull down all RNA associated with PURB. Lacking an effective antibody that targeted PURB directly, we instead overexpressed a version of *Purb* in which we incorporated a Myc Flag tag into the protein coding sequence. We then pulled down the RNA and used RT–PCR to check for the presence of *linc-HOXA1* RNA, finding that *linc-HOXA1* RNA was indeed bound to PURB in vivo (Fig. 6C). Together, these results strongly suggest that *linc-HOXA1* RNA interacts with PURB in mouse embryonic stem cells.

To check whether the interaction of *linc-HOXA1* and PURB is important for *linc-HOXA1* to repress the transcription of *Hoxa1*, we knocked down *Purb* mRNA and looked for changes in the relationship between *linc-HOXA1* and *Hoxa1*. We found that *Hoxa1* mRNA levels increased upon *Purb* mRNA knockdown (Fig. 6D) and that the anti-correlation between *linc-HOXA1* RNA and *Hoxa1* mRNA at the single-cell level disappeared (Fig. 6E), showing that PURB is critical in mediating *linc-HOXA1* RNA's repressive effects.

Discussion

Here, we showed that the lncRNA *linc-HOXA1* acts to repress the transcription of *Hoxa1* and that this repression occurs in *cis* to the chromosome. We discovered this regulation through the use of single-cell multiplex transcript counting, which enabled us to see effects that otherwise may be difficult to observe using bulk measurements.

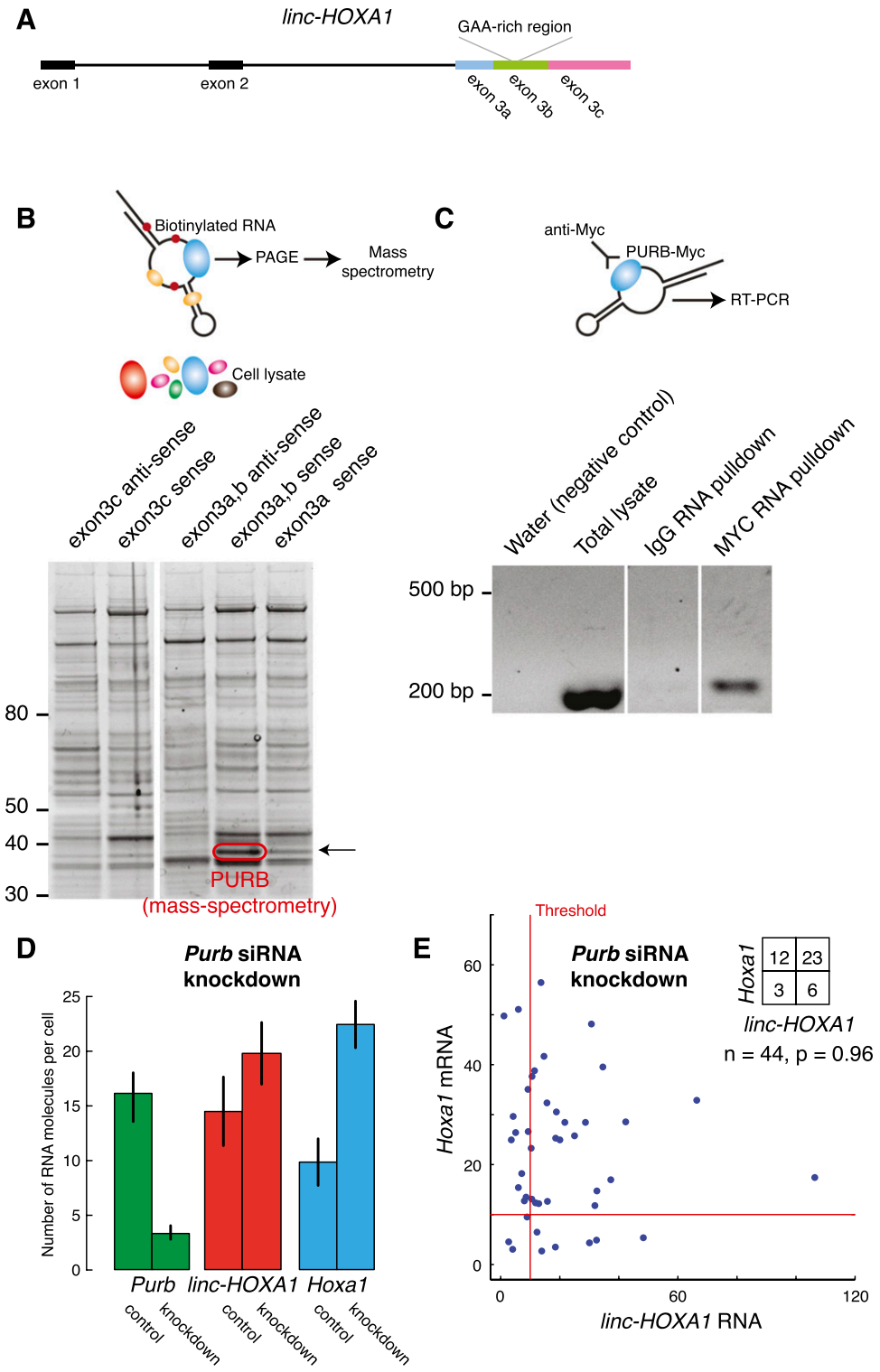

Figure 6. PURB binds to *linc-HOXA1* in vitro and in vivo and mediates repression of *Hoxa1* transcription. (*A*) Illustration of *linc-HOXA1* gene structure showing the location of the GAA purine-rich region. (*B*) Protein products visualized by gel electrophoresis that we pulled down using different biotinylated *linc-HOXA1* deletion and antisense transcripts (lanes are from the same gel, reordered for clarity). The arrow points to the length of bands containing PURB, and the red circle indicates the band that contained PURB, as determined by mass spectrometry. (*C*) We transfected Myc-tagged *Purb* into mouse embryonic stem cells, immunoprecipitated them, performed RT–PCR, and ran the products on an agarose gel to detect the presence of the *linc-HOXA1* transcript. The "water" lane (shown at *left*) served as a negative control for the PCR reaction, "total lysate" omits the immunoprecipitation step, "IgG" used an IgG antibody for the immunoprecipitation, and "MYC" used a MYC antibody to pull down Myc-tagged PURB. (*D*) *Purb* siRNA knockdown and the effects on *linc-HOXA1* and *Hoxa1* transcript levels. Error bars report the standard error of the mean. (*E*) Scatter plots showing number of *linc-HOXA1* and *Hoxa1* transcript levels in single cells in *Purb* knockdown conditions. The *inset* table shows the number of cells in the four quadrants delineated by the thresholds. We computed the *P*-value for the distribution of cells in the quadrants via the χ^2 test.

We found that the repression of *Hoxa1* by *linc-HOXA1* RNA disappeared upon the addition of retinoic acid, showing that *linc-HOXA1* RNA exerts its regulatory influence on *Hoxa1* only in the absence of retinoic acid. These results suggest that the mechanism of action of *linc-HOXA1* RNA is presumably easy to override by the binding of the retinoic acid receptor to its cognate binding sites in the *Hoxa1* promoter (Dupé et al. 1997). Also, *linc-HOXA1* is relatively distal from the *hoxa1* gene locus (~50 kb). One possibility is that in the absence of retinoic acid, the *Hoxa* cluster adopts a conformation in which *linc-HOXA1* is physically proximal to the *Hoxa1* locus, but upon addition of retinoic acid, the binding of the retinoic acid receptor to its binding site (and subsequent activation) induces a conformational change that pulls *linc-HOXA1* away from the *Hoxa1* locus, thereby making that regulatory interaction impossible. Indeed, several other results have suggested a relationship between *cis* regulation by lncRNAs and chromosome structure (Zhang et al. 2009; Ørom and Shiekhattar 2011; Wang et al. 2011; Lee 2012; Tan-Wong et al. 2012; Lai et al. 2013), and we further note the excellent work of Petruk et al. (2006) in which the investigators demonstrate regulation in *Drosophila Hox* genes by a lncRNA over a distance scale similar to that which we observed here. Such an interpretation is certainly consistent with our results showing that upon addition of retinoic acid, the correlation between *linc-HOXA1* and *Hoxa1* disappears, and *linc-HOXA1* RNA knockdown produces no effect. We further note that others have shown that the lncRNA HOTAIR-M1, located between *Hoxa1* and *Hoxa2*, serves to activate *Hoxa1* only upon addition of retinoic acid (Zhang et al. 2009). Taken together, these results point to regulation of *Hoxa1* by a pair of lncRNAs—one that represses it before retinoic acid induction, and one that activates it upon induction.

Another striking feature of *linc-HOXA1* repression of *Hoxa1* was that it appeared to involve a very sharp threshold, meaning that *linc-HOXA1* had to be below ~10 molecules per cell in order for us to observe significant *Hoxa1* transcript levels. It is possible that the manner by which lncRNAs regulate gene expression lends itself to the generation of such thresholds. On a technical note, the existence of sharp thresholds can often obscure the effects of RNA knockdown. For instance, here we knocked down *linc-HOXA1*, resulting in an increase in *Hoxa1* mRNA abundance, but the effects on the overall population are seemingly moderate due to the fact that while many (although not all) cells have some partial reduction in *linc-HOXA1* levels, only a few cells are reduced to subthreshold levels, and thus those are the only ones with an increase in *Hoxa1* mRNA abundance. It is possible that such thresholds may obscure the potential regulatory roles of many lncRNAs as measured by conventional assays. Such thresholds may play an important role in developmental switches (Maamar et al. 2007; Choi et al. 2008; Raj et al. 2010).

Interestingly, we also observed that in cells with high levels of *linc-HOXA1*, knockdown of *linc-HOXA1* resulted in those same cells then having relatively high levels of *Hoxa1*, rather than a mixture of high and low *Hoxa1* transcript abundances, as in the population overall. These results suggest that the cells fluctuate between two general states: one in which the *linc-HOXA1*/*Hoxa1* genomic region is transcriptionally inert, and one in which that region is transcriptionally active. For the cells that are in the transcriptionally active state, they have a choice to transcript either *linc-HOXA1* or *Hoxa1* but not both. In that case, a knockdown would result in those cells with high *linc-HOXA1* expression becoming cells with high *Hoxa1* expression, consistent with our data. It is possible that the transitions between the region being transcriptionally active and transcriptionally inert reflect the transcriptional changes associated with widespread phenomenon of transcriptional "bursts," which are short but intense periods of transcription interspersed between periods of transcriptional silence (Golding et al. 2005; Chubb et al. 2006; Raj et al. 2006; Ingolia et al. 2011; Suter et al. 2011). In that case, it may be that at the onset of a transcriptional burst in the region, either *linc-HOXA1* or *Hoxa1* may begin transcribing, but if *linc-HOXA1* starts transcribing first, it will repress *Hoxa1* and prevent it from transcribing during that burst. It is possible that this population that is transcriptionally competent at the 3′ end of the *Hoxa* gene cluster represents embryonic stem cells that are transiently primed for differentiation, and that *linc-HOXA1* provides some regulation of the differentiation process.

Together, our findings provide strong evidence that *linc-HOXA1* represses *Hoxa1* in *cis* through a sharp threshold. How might such a threshold arise? One possibility is that as soon as nascent *linc-HOXA1* RNA is transcribed, PURB binds to it and then represses the transcription of *Hoxa1* mRNA. In such a scenario, the production of even just a few transcripts during a burst of *linc-HOXA1* transcription would be enough to repress *Hoxa1*, thereby resulting in the sharp repression threshold that we observed. Such thresholds may serve important roles in digitizing transcriptional output, and if such mechanisms end up being ubiquitous features of lncRNA function, it may be that lncRNAs are a class of molecule particularly well suited for producing such responses. Such responses may also depend on the nature of the protein cofactor involved in the transcriptional control, and so it may be worth searching for other lncRNAs that may act similarly through interactions with PURB (or its close relative, PURA) (Hawkins and Morris 2010).

More generally, we believe that our approach of looking for regulatory interactions by analyzing single-cell correlations in transcript abundance may prove generally useful. In particular, we think such tools may be of particular utility in the analysis of lncRNAs, which are often hard to knock down effectively due to nuclear localization and for which overexpression may not produce effects if the lncRNA acts in *cis*. We anticipate that the use of correlations in combination with genome-wide techniques will provide many new insights into the functional roles of lncRNAs.

Materials and methods

Cell culture

We used the embryonic mouse cell line V6.5 (ThermoFisher Scientific), which we grew and maintained according to manufacturer's instructions. Briefly, cells were cultured at 37°C in a 5% (v/v) CO_2 incubator in KnockOut D-MEM medium (Invitrogen) supplemented with 10% FBS (GlobalStem), 2 mM L-glutamine, 1% (v/v) nonessential amino acids (Invitrogen), 10^3 U/mL LIF (Millipore), and 1 mM 2-mercaptoethanol. We coated the dishes with gelatin (0.2% [v/v]) and grew the cells on top of a layer of feeder cells (C57BL/6 mouse embryonic fibroblasts, GlobalStem). Cells were passaged as needed, and the growth medium was changed every day. To induce differentiation, cells were grown with no feeder layer in culture medium devoid of LIF supplemented with 5 μM retinoic acid. HeLa cells (gift from the laboratory of Phillip Sharp, Massachusetts Institute of Technology) were grown in Dulbecco's modified Eagle's medium (DMEM) with Glutamax (Life Technologies) supplemented with penicillin/streptomycin and 10% fetal bovine serum.

RNA FISH

For RNA FISH, embryonic stem cells were recovered from the culture dish by trypsin EDTA (0.25%). Feeder cells were removed from this mixture by incubating all of the cells for 45 min on gelatin-coated plates, in which time the feeder cells adhered to the bottom, thereby separating them from the embryonic stem cells. Nonadherent undifferentiated cells or retinoic acid-differentiated cells were washed with 1× PBS and resuspended in 1× PBS containing 1% BSA and 2 mM EDTA, to which we added formaldehyde for a final concentration of 4% for fixation. After 10 min of fixation, cells were washed twice with 1× PBS and stored in 70% ethanol at 4°C until we proceeded with RNA FISH.

RNA FISH staining followed Raj et al. (2008) with slight modifications. Fixed cells in suspension were washed with 1× PBS with 0.1% Triton and were allowed to adhere to poly-L-lysine-coated coverslip chambers before hybridization. Samples were imaged on a standard inverted epifluorescence microscope (Nikon Ti-E) using a 100× 1.43 NA oil immersion objective and a Princeton Instruments Pixis 1024BR cooled CCD camera. Thirty-five Z sections with a 0.35-μm spacing were taken for each field of view. Image analysis was performed using custom scripts in Matlab. RNA spot counting was performed as previously described (Raj et al. 2008), and colocalization of mRNA spots from different channels was performed as described by Levesque and Raj (2013). The oligonucleotide sequences that we used in our RNA FISH probes are in Supplemental Table 1, and the binding locations for *linc-HOXA1* exons 3a, 3b, and 3c; the entire *linc-HOXA1* transcript; and the *linc-HOXA1* intron are provided in Supplemental File 1. We believe our probes provide specificity based on previous experiments from our laboratory and others (Vargas et al. 2005; Raj et al. 2008) and the fact that we observed colocalization of differently labeled probes targeting different areas of the same transcript.

RNA knockdown experiments

The DNA oligonucleotides used to knock down *linc-HOXA1* were manufactured by Isis Pharmaceuticals and are thought to target nuclear as well as cytoplasmic RNA for degradation through the activity of RNase H. The following oligonucleotides were used (5′ to 3′): *linc-HOXA1*, TGCTGCAAGGCTTTACCCGA (Isis, no. 474983) and CCCACTGAAGATAGATCGGA (Isis, no. 474990); and control (no specific target), CCTTCCCTGAAGGTTCCTCC (Isis, no. 141923).

Transfections using a 40-nM pool of oligonucleotides were carried out in 12-well plates with using Lipofectamine as a transfection reagent.

We also targeted both linc-HOXA1 and Hoxa1 via conventional siRNAs (Ambion) using pools of oligonucleotides at 50 nM and Lipofectamine 2000 as a transfection reagent.

siRNAs targeting *linc-HOXA1* with the same sequences as Isis oligos and the siRNAs targeting *Hoxa1* in the embryonic stem cells had the following sequences (5′–3′): GCAGCGAU GAGAAAACGGATT and GACCUUUGACUGGAUGAAATT.

Additionally, three 27-mer siRNA duplexes targeting *Purb* were purchased from OriGene (catalog no. SR410793) and used at a final concentration of 10 nM with Lipofectamine 2000 (Invitrogen) as a transfection reagent.

In all cases, cells were harvested after 24 h of transfection, fixed as described above, and then analyzed by RNA FISH.

Cloning

We cloned the *linc-HOXA1* transcripts by amplifying cDNA from RNA isolated from undifferentiated embryonic stem cells. Primers to clone the *linc-HOXA1* were designed based on the sequence deduced from RNA sequencing data and bioinformatic analysis (Guttman et al. 2010). Using the primers in Supplemental Table 2, we obtained the three different isoforms of *linc-HOXA1* described in the text. We cloned both sense and antisense constructs into a TOPO vector for the generation of in vitro RNA from the T7 promoter.

We also cloned the three isoforms and GFP into pcDNA#.1(−) plasmid (Invitrogen) for overexpression via the CMV promoter by transfection into both the HeLa and mouse embryonic stem cells. The GFP sequence was amplified by PCR using the primers GFPFor and GFPRev (Supplemental Table 2) from pGEMT-GFP (Clontech).

Coding potential via phyloCSF

We used phyloCSF (Lin et al. 2011) to estimate the degree of evolutionary pressure on sequence substitutions acting to preserve an ORF in each of the *linc-HOXA1* isoforms that we cloned as well as all coding transcripts and intergenic noncoding transcripts annotated in RefSeq (Pruitt et al. 2003). Briefly, phyloCSF determines whether a multispecies nucleotide sequence alignment in a specific locus is more likely to represent a protein coding than a noncoding transcript. It does so by applying a probabilistic model that examines the overrepresentation of evolutionary signatures characteristic of alignments of conserved coding regions, such as the high frequencies of synonymous codon substitutions and conservative amino acid substitutions.

Coding potential from protein overexpression

The pcDNA vectors expressing the three isoforms and GFP were transfected into HeLa cells using 20 μg of plasmid and Lipofectamine 2000 (Invitrogen) as a transfection reagent. Cells were released from the culture dish by trypsin EDTA (0.05%), washed with cold 1× PBS, and then lysed using CellLytic M (Sigma). Lysate was cleared of cell debris by centrifugation at

13,000 rpm for 20 min. Protein concentration in the lysate was determined using the BCA protein assay kit (Pierce) with BSA as a standard. Proteins from each cell lysate were separated by a 10%–20% gradient Tricine SDS–polyacrylamide gel (Invitrogen) in SDS Tricine buffer. The proteins on the gel were visualized using the Simply Blue safe stain (Invitrogen). The protein content of the bottom 1 cm of the gel (corresponding to proteins <18 kDa) in each sample was analyzed by targeted mass spectrometry. The samples were digested with trypsin and analyzed with nano-liquid chromatography/tandem mass spectrometry (nanoLC/MS/MS) at the University of Pennsylvania Proteomics Core. The data were analyzed with Sequest and Scaffold software packages. A database of all possible peptide masses encoded by the *linc-HOXA1* isoforms (Supplemental Table 3) was included along with the UniProt Human database in the analysis of the raw data.

RNA pull-downs

Biotin-labeled RNA of the three isoforms of *linc-HOXA1*, along with deletions of these isoforms and antisense versions, were prepared using the Biotin RNA labeling mix (Roche) and T7 RNA polymerase (Roche). Biotinylated RNAs were treated with RNase-free DNase I (Roche) and purified on prepacked spin columns with Bio-Gel P-30 in 1× saline sodium citrate (Bio-Rad).

Mouse embryonic stem cell lysate was prepared from 5×10^7 cells that were collected by trypsinization and washed with cold 1× PBS. Cell were suspended in 4 mL of RNA immunoprecipitation (RIP) buffer (150 mM KCl, 25 mM Tris at pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, 1× protease inhibitor cocktail [Sigma]) (Rinn et al. 2007). Cells were incubated in the buffer for 30 min at 4°C with gentle agitation, after which the lysate was sonicated for 10 min. Lysate was cleared from the cell debris by centrifugation at 13,000 rpm for 20 min. Protein concentration in the lysate was determined using the BCA protein assay kit (Pierce) with BSA as a standard.

Ten picomoles of biotinylated RNA was heated for 10 min to 60°C and slow-cooled over the course of 40 min to 4°C. RNA was mixed with 1 mg of precleared embryonic stem cell lysate in RIP buffer supplemented with tRNA (0.1 μg/μL), 5 mM $MgCl_2$, and 80 U/mL RNaseOut (Invitrogen) and incubated for 2 h at 4°C with gentle rotation. Twenty-five microliters of washed T1 Streptavidin magnetic Dynabeads (Invitrogen) was added to each binding reaction and further incubated for 2 h at 4°C. Beads were washed three times with RIP-supplemented buffer (each wash carried out for 5 min at 4°C) and then boiled in SDS buffer loading buffer. The retrieved proteins were separated by a 4%–12% gradient Bis-Tris SDS–polyacrylamide gel (Invitrogen) in MOPS buffer. The proteins on the gel were visualized using a SYPRO Ruby stain (Invitrogen). The sample was digested with trypsin and analyzed with nanoLC/MS/MS at the University of Pennsylvania Proteomics Core. The data were analyzed with Sequest and Scaffold software packages.

Cell lysate from 10^7 embryonic stem cells transfected with the plasmid MR225874 (Origene) expressing C-terminal Myc-DDK-tagged *Purb* under the control of the CMV promoter was prepared as described above using 1 mL of RIP buffer supplement with 80 U/mL RNaseOut. Antibodies targeting the Myc epitope (Cell Signaling, no. 2278P) or targeting IgG (Cell Signaling, no. 2729S) were added to 0.4 mg of the cell lysate and incubated for 2 h at 4°C with gentle rotation. Twenty-five microliters of protein G magnetic Dynabeads (Invitrogen) was added to the mix and incubated for 1 h at 4°C with gentle rotation. Beads were collected and washed three times with 800 mL of RIP buffer supplemented with RNaseOut, with each wash being carried out for 5 min at 4°C. Beads were resuspended in 700 μL of Qiazol, and RNA was extracted from the beads using the miRNeasy minikit (Qiagen) and treated with RNase-free DNase I (Qiagen) according to the manufacturer's instructions. RNA from 50 μL of the total lysate was extracted simultaneously and used as positive control for RNA extraction, reverse transcription, and the PCR reaction. We generated cDNA from the RNA using both poly-dT oligonucleotide and E5 rev as primers for the reverse transcription reaction. The presence of the noncoding RNA *linc-HOXA1* in the cDNA samples was detected using PCR using the primers lncHoxa1E3bFor and lncHoxa1E3cRev (Supplemental Table 2).

Acknowledgments

We thank Biosearch Technologies for providing many reagents crucial to these experiments. We also thank Gautham Nair for help with statistical analysis, Marshall Levesque for help with image analysis, and Ezgi Hacisuleyman for help with the RNA pull-down experiments. We thank other members of the Raj laboratory for useful discussions. H.M. and A.R. acknowledge support from a Burroughs-Wellcome Fund Career Award at the Scientific Interface. A.R. also acknowledges support from an NIH Director's New Innovator Award (1DP2OD008514). M.N.C. is supported by a Howard Hughes Medical Institute International Student Research Fellowship. J.R. acknowledges support from Director's New Innovator Awards 1DP2OD00667, P01GM099117, and P50HG006193-01. J.R. is a Damon Runyon-Rachleff, Searle, Smith Family, and Merkin Fellow.

References

Bergemann AD, Johnson EM. 1992. The HeLa Pur factor binds single-stranded DNA at a specific element conserved in gene flanking regions and origins of DNA replication. *Mol Cell Biol* **12:** 1257–1265.

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

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

Choi PJ, Cai L, Frieda K, Xie XS. 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322:** 442–446.

Chubb JR, Trcek T, Shenoy SM, Singer RH. 2006. Transcriptional pulsing of a developmental gene. *Curr Biol* **16:** 1018–1025.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* **22:** 1775–1789.

Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Soldà G, Simons C, et al. 2008. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* **18:** 1433–1445.

Dupé V, Davenne M, Brocard J, Dollé P, Mark M, Dierich A, Chambon P, Rijli FM. 1997. In vivo functional analysis of the Hoxa-1 3′ retinoic acid response element (3′RARE). *Development* **124:** 399–410.

The ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816.

Fremeau RT, Lundblad JR, Pritchett DB, Wilcox JN, Roberts JL. 1986. Regulation of pro-opiomelanocortin gene transcription in individual cell nuclei. *Science* **234:** 1265–1269.

Gallia GL, Darbinian N, Johnson EM, Khalili K. 1999. Self-association of Purα is mediated by RNA. *J Cell Biochem* **74:** 334–348.

Golding I, Paulsson J, Zawilski SM, Cox EC. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell* **123:** 1025–1036.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, et al. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* **28:** 503–510.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477:** 295–300.

Hawkins PG, Morris KV. 2010. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *transcription* **1:** 165–175.

Hu W, Alvarez-Dominguez JR, Lodish HF. 2012. Regulation of mammalian cell differentiation by long non-coding RNAs. *EMBO Rep* **13:** 971–983.

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, et al. 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142:** 409–419.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147:** 789–802.

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

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494:** 497–501.

Lee JT. 2012. Epigenetic regulation by long noncoding RNAs. *Science* **338:** 1435–1439.

Levesque MJ, Raj A. 2013. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat Methods* **10:** 246–248.

Liang X-H, Vickers TA, Guo S, Crooke ST. 2011. Efficient and specific knockdown of small non-coding RNAs in mammalian cells and in mice. *Nucleic Acids Res* **39:** e13.

Lin MF, Jungreis I, Kellis M. 2011. PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27:** i275–i282.

Maamar H, Raj A, Dubnau D. 2007. Noise in gene expression determines cell fate in Bacillus subtilis. *Science* **317:** 526–529.

Mainguy G, Koster J, Woltering J, Jansen H, Durston A. 2007. Extensive polycistronism and antisense transcription in the Mammalian Hox clusters. *PLoS ONE* **2:** e356.

Mercer T, Dinger M, Mattick J. 2009. Long non-coding RNAs: Insights into functions. *Nat Rev Genet* **10:** 155–159.

Ørom UA, Shiekhattar R. 2011. Noncoding RNAs and enhancers: Complications of a long-distance relationship. *Trends Genet* **27:** 433–439.

Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, et al. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143:** 46–58.

Pauli A, Rinn JL, Schier AF. 2011. Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* **12:** 136–149.

Petruk S, Sedkov Y, Riley KM, Hodgson J, Schweisguth F, Hirose S, Jaynes JB, Brock HW, Mazo A. 2006. Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* **127:** 1209–1221.

Ponting CP, Oliver PL, Reik W. 2009. Evolution and functions of long noncoding RNAs. *Cell* **136:** 629–641.

Pruitt KD, Tatusova T, Maglott DR. 2003. NCBI Reference Sequence project: Update and current status. *Nucleic Acids Res* **31:** 34–37.

Raj A, Tyagi S. 2010. Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. *Methods Enzymol* **472:** 365–386.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4:** e309.

Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. 2008. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5:** 877–879.

Raj A, Rifkin SA, Andersen EC, van Oudenaarden A. 2010. Variability in gene expression underlies incomplete penetrance. *Nature* **463:** 913–918.

Ramsey JE, Kelm RJ Jr. 2009. Mechanism of strand-specific smooth muscle α-actin enhancer interaction by purine-rich element binding protein B (Purβ). *J Phys Chem B* **48:** 6348–6360.

Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81:** 145–166.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, et al. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129:** 1311–1323.

Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332:** 472–474.

Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM, Proudfoot NJ. 2012. Gene loops enhance transcriptional directionality. *Science* **338:** 671–675.

Vargas DY, Raj A, Marras SAE, Kramer FR, Tyagi S. 2005. Mechanism of mRNA transport in the nucleus. *Proc Natl Acad Sci* **102:** 17008–17013.

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472:** 120–124.

Xing Y, Johnson CV, Dobner PR, Lawrence JB. 1993. Higher level organization of individual gene transcription and RNA splicing. *Science* **259:** 1326–1330.

Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M, Gingeras TR, Kapranov P, Weissman SM, Newburger PE. 2009. A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* **113:** 2526–2534.