Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription

Michael S Werner[1], Matthew A Sullivan[1,2], Rohan N Shah[1,2], Rangarajan D Nadadur[3], Adrian T Grzybowski[1], Vasiliy Galat[4], Ivan P Moskowitz[3] & Alexander J Ruthenburg[1,2]

We recently described a new class of long noncoding RNAs (lncRNAs) that are distinguished by especially tight chromatin association and whose presence is strongly correlated to expression of nearby genes. Here, we examine the *cis*-enhancer mechanism of this class of chromatin-enriched RNA (cheRNA) across multiple human cell lines. cheRNAs are largely cell type specific and provide the most reliable chromatin signature to predict *cis*-gene transcription in every human cell type examined. Targeted depletion of three cheRNAs decreases expression of their neighboring genes, indicating potential co-activator function, and single-molecule fluorescence *in situ* hybridization (smFISH) of one cheRNA-distal target gene pair suggests a spatial overlap consistent with a role in chromosome looping. Additionally, the cheRNA HIDALGO stimulates the fetal hemoglobin subunit gamma 1 (*HBG1*) gene during erythroid differentiation by promoting contacts to a downstream enhancer. Our results suggest that multiple cheRNAs activate proximal lineage-specific gene transcription.

Noncoding RNAs are thought to promote transcription initiation of coding genes by recruiting histone-modifying complexes[1–4], stabilizing transcription factor or mediator binding[2,5–7], and increasing the strength of promoter–enhancer looping[2,5,8–10]. In light of the recently appreciated promiscuity of lncRNA–protein interfaces[11], how lncRNA interactions can achieve their implicated roles with such limited specificity has become a central question. One possible resolution could be spatially restricted activity due to immobilization of certain lncRNA at the sites of their production, as has been observed in a limited number of cases[5,8–10,12,13]. A more widespread role for this type of mechanism was suggested by our identification of cheRNAs, a new class of several thousand lncRNAs in HEK293 cells defined by high chromatin enrichment as a consequence of their ongoing transcription[14]. Although cheRNAs are molecularly distinct from canonical enhancer RNAs (eRNAs), they exhibit a strong correlation to proximal gene expression. Further support for the idea that biochemical fractionation of chromatin is a powerful approach to identify RNA molecules that act locally was provided by the demonstration that some eRNAs that activate nearby genes in response to epidermal growth factor are also enriched in the chromatin fraction[10]. However, many important questions regarding cheRNAs and their relationship to nearby genes remain. How general are their properties and functions? To what extent are they shared between different cell lineages? Do cheRNA molecules promote neighboring gene transcription, or are they inert byproducts of enhancer transcriptional activity (both cases have been observed for other noncoding RNA (ncRNA) classes[5,8–10,15–18])? Finally, how might these regulatory modules have evolved?

To begin to address these questions, we examined cheRNAs in other cell types and explored the functional consequences of their perturbation. Quantitative chromatin enrichment of nuclear RNA from three distinct cell types shows that the vast majority of cheRNAs are cell type specific. Nevertheless, proximity to a cheRNA is a more effective predictor of *cis*-gene expression than are putative enhancers derived from chromatin mark signatures, previously annotated lncRNAs or eRNAs. Our prior work established that most cheRNAs remain attached to chromatin via RNAP II[14]. We now directly measure the spatial distribution of one cheRNA relative to its site of transcription and the putative target gene and find them to be remarkably colocalized despite a >50-kb spacing along the chromosomal coordinate. Targeted depletion of several candidate cheRNAs produces significant decreases in neighboring gene expression for 75% of the loci examined, establishing cheRNAs as transcriptional activators. Characterizing a more specific example, we find that the cheRNA molecule HIDALGO is required for full stimulation of hemoglobin subunit *HBG1* during erythroid differentiation, and that knockdown of HIDALGO reduces contact between the *HBG1* promoter and a downstream enhancer. Finally, virtually all cheRNAs reside within class I transposable elements, providing a plausible evolutionary path for this form of regulation.
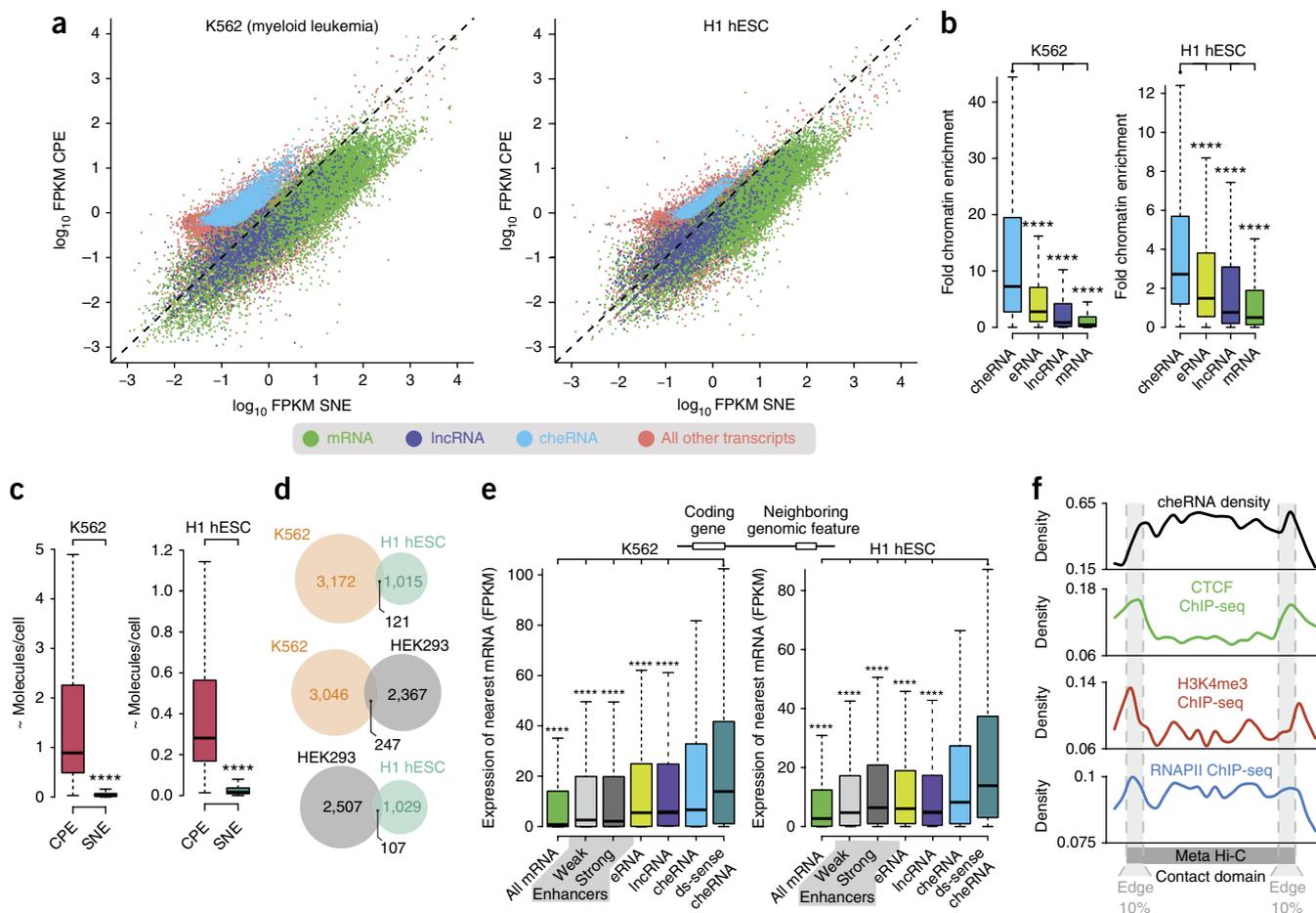
RESULTS
Chromatin-enriched ncRNAs are lineage specific and correlate with proximal gene transcription
To characterize chromatin-enriched RNAs in multiple human cell lines, we performed biochemical fractionation of nuclei, coupled

Figure 1 Tissue specificity and *cis*-gene activity of cheRNAs. (**a**) Scatterplots of *de novo*–assembled transcripts after nuclear fractionation coupled to RNA-seq[14] plotting chromatin pellet extract (CPE) versus soluble nuclear extract (SNE) enrichment for K562 and H1 hESCs cells (*n* = 3 independent cultures fractionated for each). Colors (indicated below) denote Gencode annotation of mRNA and lncRNA[59], along with new cheRNA species and all remaining transcripts. (**b**) Fold chromatin enrichment analysis of the indicated RNA classes (CPE FPKM/SNE FPKM) from the experiment shown in **a**. In **b**, **c**, and **e**, boxes span the lower to upper quartile boundaries; median is indicated with a black line; *P* values are calculated by Mann–Whitney *U* test, ****$P < 2 \times 10^{-16}$. (**c**) CheRNA molecules per fraction, determined by calibration with spiked-in *in vitro*–transcribed standards. *n* = 3 for each fraction analyzed (**Supplementary Fig. 1**). (**d**) Overlap of cheRNAs from K562, H1, and the prior HEK293 (ref. 14) data sets demonstrate they are largely unique to each cell line. (**e**) K562 and H1 expression of nearest genes to indicated genomic features: all mRNA, weak and strong enhancers as annotated by chromatin signatures[26], expressed lncRNA[60], expressed eRNA loci[29], cheRNA loci, and cheRNA downstream and in the same sense as their neighbor (ds-sense cheRNA) from the experiment shown in **a**. More extensive comparisons are available in **Supplementary Figure 2c**. (**f**) Average density of cheRNA from **Fig. 1a** versus published CTCF, RNAPII, and H3K4me3 ChIP-seq[21], contoured over a 'meta' Hi-C contact domain comprising of all TADs[33] that contain cheRNAs in K562 cells, with edge deciles indicated.

to calibrated RNA-seq[14,19,20] from H1 human embryonic stem cells (H1 hESCs) and myeloid leukemia cells (K562), which are the most divergent tier 1 ENCODE cell types[21]. Subnuclear-compartment quantification of *de novo*-assembled transcripts (**Supplementary Fig. 1a–c**) identified 3,293 and 1,136 cheRNAs in K562 cells and H1 hESCs, respectively (**Fig. 1a**, **Supplementary Fig. 2a,b**, and bioinformatics section in **Supplementary Note 1**). This extension of our prior HEK293 results[14] demonstrates the generality of cheRNAs across diverse cell lineages and provides a resource for future exploration of lncRNA mechanisms operating at the chromatin interface (Source Data for **Fig. 1**). Previously annotated lncRNAs and eRNAs also exhibit modest chromatin enrichment, consistent with many of their associated functions[1,2,5–10,22], although they are on average less enriched than cheRNAs (**Fig. 1b**).

Calibrated RNA-seq also provides a rough measure of the RNA copy number and distribution between subnuclear compartments. We measured 120 ± 40 copies of *XIST* RNA in chromatin, as compared to

2.5 ± 0.4 copies in the soluble nuclear extract per human K562 cell, congruent with previous estimates of murine Xist (~50–200 copies per cell)[23]. Given the likelihood of incomplete recovery during nuclear fractionation we estimate that most cheRNAs are present at ~1–10 copies per cell (**Fig. 1c**), consistent with smFISH measurements for other lncRNA in a variety of cell types[1,24].

Comparison of cheRNA species from our previous HEK293 data set with those from K562 and H1 hESCs reveals that the majority of cheRNAs display cell-type-specific expression (**Fig. 1d**), are largely distinct from other annotated ncRNA species in each cell type, and display little coding potential (**Supplementary Fig. 2a,b**). This strongly restricted expression is in contrast to activating RNAs (ncRNA-a), an annotation of *cis*-activating lncRNAs that were largely shared between three disparate cell types[25].

Analogous to observations in HEK293 cells[14], the presence of a proximal cheRNA in K562 and H1 hESCs is highly correlated to nearby gene expression, and substantially more coupled to *cis*-gene

expression than neighboring enhancers annotated by chromatin signatures[26–28] or transcriptionally active eRNA loci[29] (**Fig. 1e**, **Supplementary Fig. 2c**). This correlation is even more pronounced
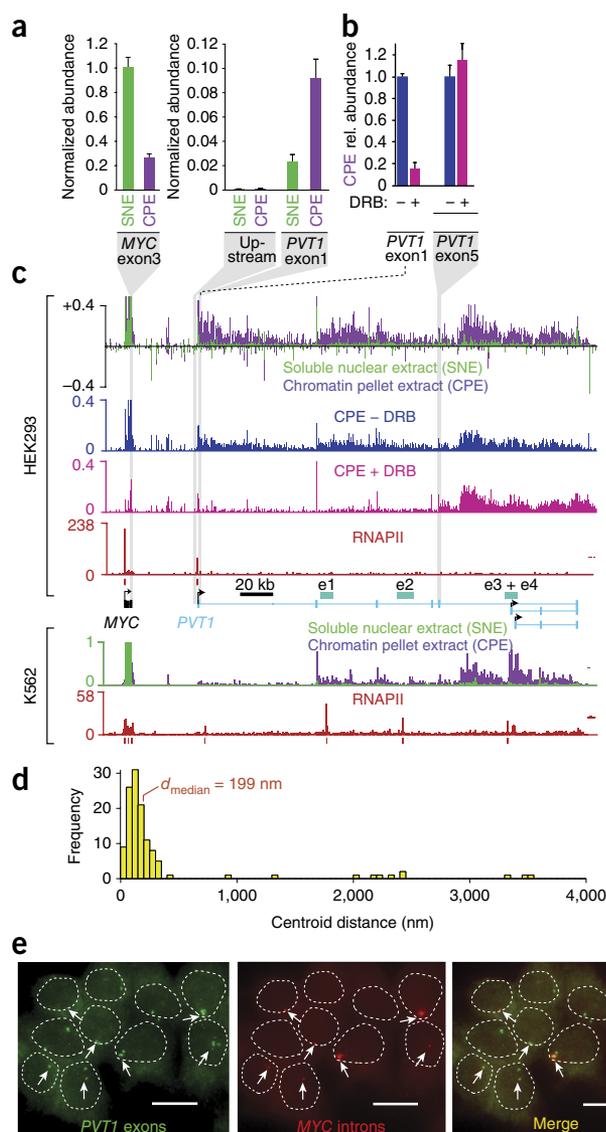
Figure 2 cheRNA *PVT1* bridges the site of its production to its target gene, *MYC*. (**a**) Abundance of *MYC* and *PVT1* transcripts in the soluble nuclear extract (SNE) and chromatin pellet extract (CPE) measured by RT-qPCR, normalized to 18S rRNA ($n = 3$ qPCR technical replicates; error bars represent s.d.). A region upstream of *PVT1* is included to show that this RNA is not the product of transcriptional readthrough from *MYC*. (**b**) Relative abundance of *PVT1* exons 1 and 5 in the CPE with transcriptional inhibition by DRB (100 μM) for 2 h, normalized to the untreated conditions ($n = 3$, as in **a**). (**c**) RNA-seq of chromatin pellet fraction (purple), soluble nuclear fraction (green) from HEK293 cells (top) and K562 (bottom) at the *MYC–PVT1* locus. RNA-seq of the chromatin pellet ± DRB (blue, pink) from HEK293 cells, and RNAPII ChIP-seq (red)[21] from each cell line are depicted with called peaks (red bars), and enhancers of *MYC* determined by a CRISPRi tiling screen (cyan, e1–4)[35]. smFISH of *PVT1* and *MYC* introns. (**d**) Frequency of nuclear puncta exhibiting overlap in both fluorescence channels was determined by measuring centroid distance differences of nearest neighbors, plotted as a histogram. (**e**) Representative field of HEK293 cells from a compressed *z* stack in smFISH experiment with nuclei from differential interference contrast (DIC) outlined (white dashed lines) and *MYC* intron foci (white arrows). Scale bars, 15 μm. A total of 50 cells from eight image fields were examined.

for cheRNAs that are downstream of and in the same sense as their coding neighbor. In some cases, biogenesis of cheRNAs may be linked to their upstream coding gene[30] or distinct, as defined by a 5′ cap (**Supplementary Fig. 2d**) and canonical promoter-chromatin hallmarks[14]. cheRNA-proximal coding genes also appear to be specifically expressed in their respective cell types, including genes in the ERK1/2 cascade in H1 hESCs[31] and JAK-STAT signaling in K562 (ref. 32) (**Supplementary Fig. 2e**), hinting that cheRNAs may have a role in cell-type-specific gene expression rather than basal function.

To investigate where cheRNAs reside in the 3D genomic architecture, we analyzed cheRNA positions relative to annotated topologically associating domains (TADs) in K562 cells[33]. cheRNA density displays local peaks at TAD boundaries (**Fig. 1f**), congruent with a recent model suggesting that ncRNA transcription can serve as focal points for chromosome domain contacts[17]. Furthermore, the cheRNA correlation to proximal-gene expression applies to all genes within a given TAD (**Supplementary Fig. 2c**).

smRNA-FISH indicates that a cheRNA acts near its site of production

Although the bulk of cheRNA are tightly associated with chromatin through the act of ongoing or stalled RNAPII transcription[14], our prior measurements did not provide spatial information about the site of attachment. We sought to quantify the physical proximity of a cheRNA molecule relative to its site of production and presumptive neighboring target gene by smFISH. The *PVT1* gene, which in patient tumors frequently occurs in tandem with *MYC* amplification[34], encodes a highly chromatin-enriched ncRNA in HEK293 and K562 cells, meriting cheRNA classification (**Fig. 2a,c**). There are multiple enhancers of *MYC* transcription resident within *PVT1* (e1–4)[35] (**Fig. 2c**). Curiously, the latter two enhancers reside in a region in *PVT1* that is resistant to inhibition by the RNAPII elongation inhibitor DRB (**Fig. 2b,c**).

We simultaneously targeted *MYC* introns and *PVT1* exons with specific probes in two-color smFISH (**Supplementary Table 1**) to query the location of all *PVT1* forms as compared to the nascent pool of *MYC* transcripts still resident at the *MYC* locus (**Fig. 2d,e**)[19]. We observe that *PVT1* exon staining is largely resident in the nucleus, distributed into only a few discrete puncta per cell (mean = 1.6 ± 0.4, **Fig. 2e**). Similarly, *MYC* intronic RNA, indicative of local transcription at the *MYC* locus, is largely restricted to approximately one nucleus-resident body per cell, and many cells did not display any focal staining. Analysis of nuclei that contain at least one of each color focus shows that *PVT1* RNA is strikingly colocalized with ongoing transcription from the *MYC* gene (**Fig. 2d,e**), predominantly overlapping within the optical diffraction limit for these dyes. Specifically, the median distance between the nearest *PVT1* and *MYC* foci for a given nucleus, 199 nm, is far closer than the minimum spacing between the sites of RNA biogenesis in extended conformation (the distance range from a notional 30-nm to 10-nm fiber would be ~420–7,700 nm for 55 kb)[36]. Our results provide new and orthogonal single-cell evidence that the *PVT1* enhancers are in close proximity to the *MYC* locus, consistent with physical contact observed by RNAPII ChIA-PET[37], while arguing that the *PVT1* cheRNA stays largely resident at the site of its production.

Function of cheRNA transcription on neighboring gene expression

The high correlation of active gene expression neighboring cheRNA loci, and other examples of ncRNA acting *in cis*[1,2,5,8–10,25], prompted us to test whether cheRNAs promote local gene expression. We used CRISPRi[38] in K562 cells to inhibit transcription of two cheRNAs located 67 kb and 71 kb downstream of their nearest coding genes, *B3GNT2* and *PDCD6IP*, and one 19 kb upstream of its nearest neighbor,
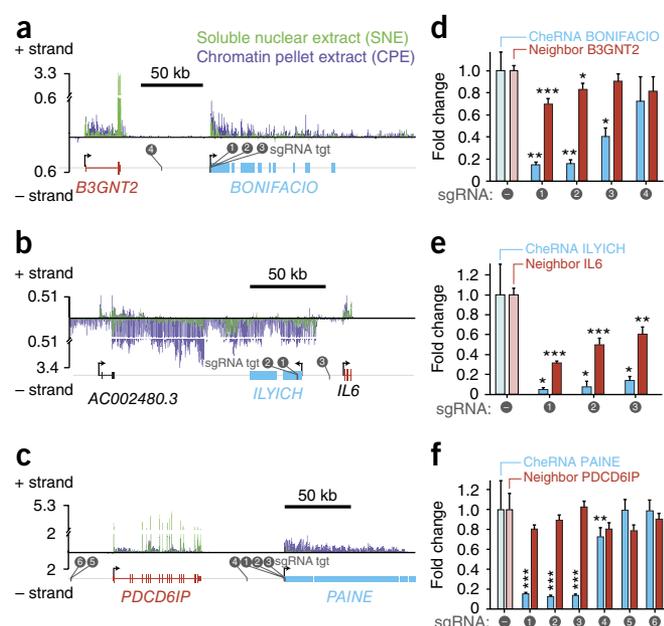
Figure 3 Examination of *cis*-enhancer activity of cheRNA–gene pairs. (**a**–**c**) Density of RNA-seq reads from K562 SNE (green) and CPE (purple) contoured over the indicated chromosomal region of origin, encompassing a cheRNA (cyan) and nearby gene (red) pair. (**d**–**f**) RT-qPCR of the cheRNA and its neighboring gene corresponding to **a**–**c** after CRISPRi-mediated knockdown of individual cheRNAs. A minimum of three distinct sgRNAs were used to target individual cheRNA loci at the indicated positions. Fold change values represent mean $2^{\Delta\Delta Ct}$ relative to a nontargeting negative control sgRNA[38] (–), with normalization to the *PPIB* reference gene and adjustment for primer-specific amplification factors; error bars represent s.e.m. from $n = 3$ (**d**,**e**) or $n = 4$ (**f**) independent targeting transfections and $n = 6$ (**d**,**e**) or $n = 7$ (**f**) independent negative control sorted transfections (Online Methods). *$P < 0.05$, **$P < 0.005$, ***$P << 0.001$ relative to negative control (Welch's two-tailed t test).

IL6. These pairs were chosen from the most highly expressed cheR-NAs in K562 cells. Consistent with our metagene analysis (**Fig. 1f**), each of these cheRNA–gene pairs fell on the edge of a chromosome-contact domain[33], although they were not selected on this basis (**Supplementary Fig. 3d–f**). Several guide RNAs (sgRNAs) targeting each cheRNA (**Supplementary Table 1**) reduced cheRNA levels by 60–95%, leading to proportional reduction of expression of the most proximal gene in two of three cases ($P < 0.05$, Welch's two-tailed t test, **Fig. 3**). Despite effective targeting of PAINE, none of the sgRNAs produced substantial changes in *PDCD6IP* expression, revealing that not all cheRNAs act on their nearest-neighbor gene.

We conclude that the CRISPRi effects are on target, as distinct sgRNAs display similar perturbations, sgRNA targeting sites between *B3GNT2* and its neighboring cheRNA BONIFACIO did not alter *B3GNT2* expression, and no consistent perturbations were observed for the housekeeping gene *GAPDH* (**Supplementary Fig. 3a**). Moreover, the same sgRNAs transfected in HEK293 CRISPRi cells, which modestly express *PDCD6IP* and *B3GNT2* but not the corresponding cheRNAs, did not lead to knockdown of these genes (**Supplementary Fig. 3b,c**). Collectively, these data indicate that cheRNA loci can act as transcriptional activators *in cis*, although they do not distinguish whether an act of transcription or the cheRNA molecule itself is responsible for the effect. We subsequently explored this distinction in the context of a developmentally induced gene–cheRNA pair.
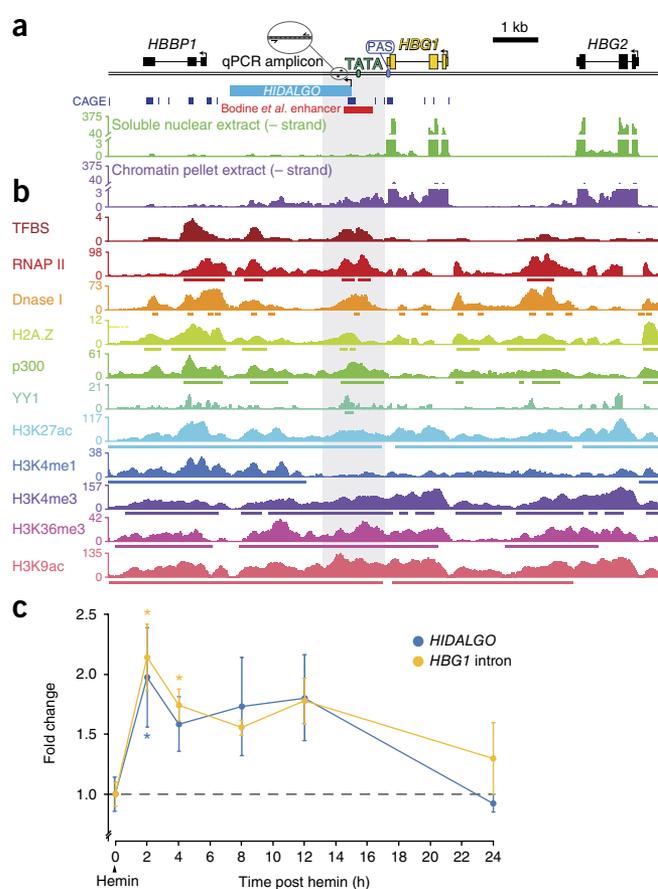

Figure 4 *HIDALGO* exhibits hallmarks of a promoter and is induced with *HBG1* during erythrogenesis. (**a**) RNA-seq of K562 chromatin (purple) and soluble nuclear extract (green) contoured over the *HBG1* locus and flanking regions, CAGE peaks[60] arising from 5′-capped transcripts, and a previously identified enhancer element[40,41] are indicated. (**b**) Chromatin signatures from ChIP-seq and DNase I hypersensitivity measurements in K562 (ref. 21) indicate a regulatory region downstream of *HBG1* where *HIDALGO* is transcribed. Called peaks are depicted as colored bars beneath the coverage track. (**c**) Time course measuring the levels of fetal *HBG1* with intron-specific primers and *HIDALGO* cheRNA by RT-qPCR following addition of 50 μM hemin to induce erythroid differentiation. y axis represents mean fold change ($2^{\Delta\Delta Ct}$) by RT-qPCR relative to time $t = 0$ and 18S rRNA. Error bars represent s.e.m. from $n = 4$ (2, 4 h) or $n = 3$ (8, 12, 24 h) independent experiments; for $t = 0$, error bars represent s.e.m. from $n = 4$ qPCR technical replicates, summed in quadrature across $n = 4$ independent experiments (*$P < 0.05$ versus $t = 0$, Welch's two-tailed t test).

cheRNA HIDALGO couples an enhancer and promoter of *HBG1 to* activate *HBG1* transcription

To determine whether cheRNAs play a role in differentiation, we induced K562 cells toward the erythroid lineage by treatment with the small molecule hemin for 48 h[39], and then performed nuclear fractionation and sequencing. In contrast to our cell-line comparisons, 75% of cheRNA were shared with uninduced K562 cells (**Supplementary Fig. 4a**). Of the 172 upregulated coding genes, 27 were flanked by a cheRNA within 100 kb, a slight overrepresentation over chance expectation ($P < 0.02$, two-tailed Fisher's exact test).

To better understand cheRNA biogenesis and putative enhancer mechanisms in differentiation, we analyzed an erythroid cheRNA–gene pair. A hallmark of erythroid commitment is upregulation of the *HBG1* and *HBG2* chains of fetal hemoglobin (γ-globin), for which hemin induction of K562 cells is an effective model system[39]. We observed chromatin-enriched transcription extending 3.7 kb beyond
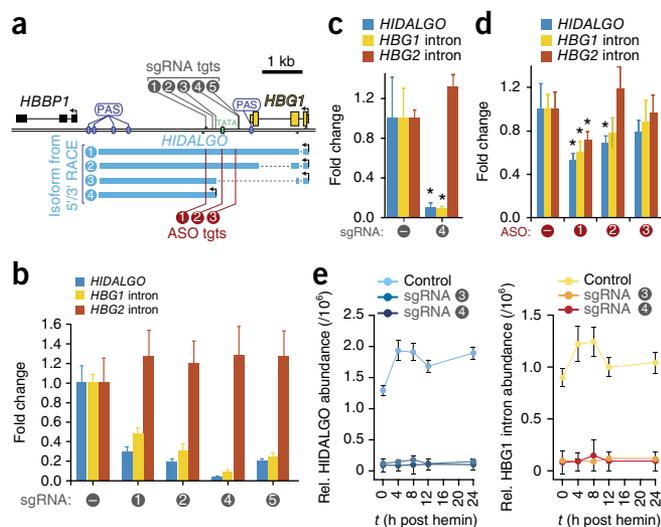
Figure 5 *HIDALGO* promotes *HBG1* expression. (**a**) Diagram of *HIDALGO* 5′ and 3′ RACE products and location of CRISPRi sgRNAs and ASO target sites (denoted tgts). (**b**) CRISPRi-mediated knockdown of *HIDALGO* with four distinct gRNAs decreases *HBG1* transcription proportionally. Fold change is calculated relative to a nontargeting negative control sgRNA (–) and to 18S RNA by RT-qPCR ($n = 1$ sorted transfection; mean $2^{\Delta\Delta Ct}$ and error bars (s.d.) are from $n = 3$ qPCR technical replicates). (**c**) CRISPRi with sgRNA4 ($n = 3$ independent sorted transfections; error bars represent s.e.m., *$P < 0.05$, Welch's two-tailed *t* test). (**d**) Knockdown of *HIDALGO* RNA with three different ASOs decreases *HBG1* expression ($n = 4$ independent experiments; error bars represent s.e.m., *$P < 0.05$, Welch's two-tailed *t* test). (**e**) Time courses of *HIDALGO* (left) and *HBG1* induction (right) upon erythrogenesis in dCas9-Krab K562 cells or polyclonal stable integrants of sgRNA3 or sgRNA4 in this background ($n = 1$ sorted transfection; values are mean expression relative to 18S rRNA measured by RT-qPCR; error bars represent s.d. from $n = 4$ qPCR technical replicates).

HBG1 in both uninduced and induced states in a region previously shown to have enhancer activity in reporter assays[40,41] (**Fig. 4a**), whereas no transcription was observed at this locus in H1 hESC cells (**Supplementary Fig. 4b**). ChIP-seq data[21] in this region reveals chromatin features characteristic of an unannotated promoter downstream of *HBG1*, with overlapping peaks for transcription factor binding sites, H3K4me3, H3K27ac, RNAP II, and DNase I hypersensitivity coupled to evidence of 5′-capped transcripts (**Fig. 4b**). Intriguingly, transcription of this cheRNA, hereafter called HIDALGO for 'hemin-induced cheRNA downstream of fetal hemoglobin', is induced early in erythroid differentiation within 2–4 h after hemin addition, and then returns to basal levels within 2 d (**Fig. 4c**).

We examined HIDALGO RNA biogenesis by 5′ rapid amplification of cDNA ends (5′ RACE), which revealed a complex set of transcripts emanating from the TSS of *HBG1* and a location downstream near our predicted *HIDALGO* TSS (**Fig. 5a**). Whereas one transcript that originates from the *HBG1* TSS represents readthrough that escapes polyadenylation (isoform #2), two others are out of frame and riddled with stop codons, seemingly due to errant or alternative splicing (**Fig. 5a, Supplementary Fig. 4c**). Owing to incomplete processing and chromatin tethering, all of these transcripts are *de facto* cheRNAs, and we refer to them as HIDALGO isoforms herein. To assess the proportion of *HBG1* TSS transcripts that escape polyadenylation, we performed 3′ RACE on *HBG1*, which revealed that >83% of transcripts are processed at the normal polyadenylation site (PAS) to become mature mRNA (**Supplementary Fig. 4d**). Readthrough from the *HBG1* promoter, particularly isoforms #1 and 2, composes

the majority (90–95%) of basal HIDALGO transcript levels (**Supplementary Fig. 4e,f**). Hemin induces all four RACE transcripts, although the transcript emanating from the cryptic TATA box (#4) represents the greatest fold change, comprising ~15% of HIDALGO RNA 2 h after induction (**Supplementary Fig. 4e,f**).

We used CRISPRi to inhibit readthrough transcription from the *HBG1* gene and initiation from the downstream TATA box (**Fig. 5a**). Because the two fetal hemoglobin genes are only 3.5 kb apart on chromosome 11 and are >99% identical at the mature RNA level, *HBG2* transcripts serve as an excellent control for *HBG1*-specific effects. To this end, we deployed primer sets that target unique intronic or 3′-UTR sequences to distinguish these RNA species (**Supplementary Fig. 5b,f** and **Supplementary Fig. 6**). Each of the sgRNAs led to a decrease in transcription of *HBG1*, but not *HBG2*, proportional to the level of cheRNA knockdown (**Fig. 5b,c** and **Supplementary Fig. 4g**). As a control for the spatial distribution of dCas9-KRAB to the 3′ end of *HBG1*, we confirmed that a gene without a nearby cheRNA was not suppressed when using an sgRNA at the same relative location (**Supplementary Fig. 5a**). Moreover, 3′ RACE of *HBG1* demonstrates that the majority of transcripts are processed immediately following the PAS (**Supplementary Fig. 4d**), so any effect on this pool is restricted to the fewer than 17% of transcripts that escape 3′ processing, and thus could not account for the observed 88% decrease in *HBG1* (**Fig. 5b,c**).

While our CRISPRi experiments demonstrate that the *HIDALGO* locus is an activator of *HBG1*, they do not distinguish whether the act of transcription through *HIDALGO* or the RNA molecule itself is functionally relevant[17,18]. To test the latter mechanism, we used antisense oligonucleotides (ASOs) to specifically degrade complementary RNA through nuclear RNase H–initiated cleavage[42]. We observed decreases in *HBG1* transcription commensurate to the degree of *HIDALGO* knockdown ($P < 0.05$, *t* test), demonstrating that the RNA molecule plays a functional *cis*-regulatory role (**Fig. 5d**). Finally, inhibiting *HIDALGO* during hemin-induced erythroid differentiation prevents *HBG1* induction (**Fig. 5e**), suggesting a role for this cheRNA in developmental transcriptional plasticity.

As several lncRNAs and eRNAs facilitate contact between promoter and enhancer elements through chromatin looping[2,5,8–10], we tested whether a similar model operates at the *HIDALGO-HBG1* locus. Chromatin confirmation capture (3C)[43] demonstrates that the *HBG1* promoter contacts *HBG1* exon 2 and the *HIDALGO* #4 TSS, both of which are diminished by ASO (**Fig. 6a**) or CRISPRi depletion of *HIDALGO* (**Supplementary Fig. 5c**). Although each of these perturbations acts through distinct mechanisms, as reflected by distinct changes in the histone modification patterns (**Fig. 6b**), the consequences in regard to contact frequency are similar. CRISPRi targeting of promoters is thought to act by recruiting the Set1DB methyltransferase to install the H3K9me3 mark[44]. Remarkably, our ICeChIP quantification[45] demonstrates that H3K9me3 approaches saturation (100%) proximal to the site of dCAS9-KRAB-sgRNA3 binding near the TSS of *HIDALGO* #4, with concomitant slight increases at the *HBG1* promoter. In contrast, antisense oligonucleotide targeting of the HIDALGO molecule does not substantially alter the pattern of H3K4me3 and H3K9me3 at the two sites queried. Yet there is a slight increase in H3K27me3 at the *HBG1* promoter, perhaps indicating spreading of this mark[46] as a consequence of altered chromatin architecture. Crucially, the TSS of *HIDALGO* #4 near the 3C contact is a potent transcriptional activator in luciferase assays, consistent with a potential enhancer role modulated by the HIDALGO cheRNA and supported by prior reports of enhancer elements within this region[40,41] (**Fig. 6c**). Taken together, our results
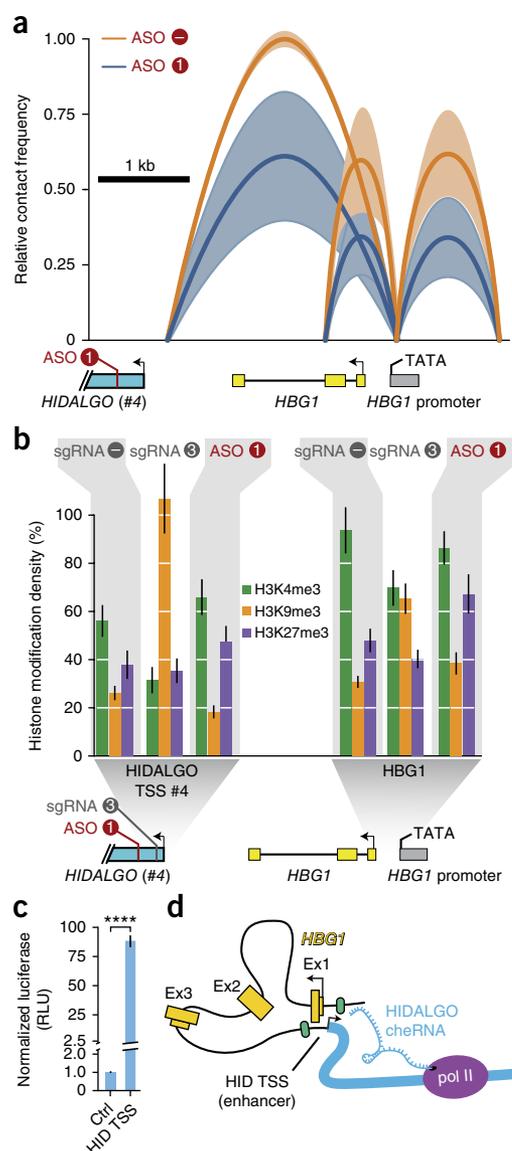
Figure 6 *HIDALGO* couples an enhancer at its TSS to the *HBG1* promoter. (**a**) 3C in ASO-transfected K562 cells after sorting reveals a *HIDALGO*-sensitive contact between the *HBG1* promoter and a region near the *HIDALGO* transcript #4 TSS (*n* = 4 independent experiments, Online Methods). Line arcs depict average relative contact frequency of the two vertices, with s.d. indicated in lighter color. An analogous experiment using CRISPRi is presented in **Supplementary Figure 5c**. (**b**) ICeChIP-qPCR quantification of the levels of the three indicated histone lysine methyl marks at *HIDALGO* TSS #4 and the *HBG*– promoter, comparing sgRNA3- and ASO1- mediated *HIDALGO* perturbations to control ('–' sgRNA). The ASO '–' control is shown in **Supplementary Figure 5e**. Error bars represent s.e.m. from three qPCR technical replicates; each ICeChIP was performed on material from a single cell culture. (**c**) Enhancer assay by relative luciferase activity of the *HIDALGO* TSS#4 (HID TSS) in K562 cells. HID TSS firefly luciferase activity normalized to co-transfected *Renilla* luciferase vector. Luciferase (*n* = 6 independent experiments; error bars represent s.e.m., ****$P < 10^{-4}$, Welch's two-tailed *t* test). (**d**) Model of *HIDALGO* activation of *HBG1* by bridging the enhancer (HID TSS) to the promoter of *HBG1*.

indicate that the HIDALGO RNA molecule confers *cis* activation of *HBG1* by mediating contacts with a downstream enhancer element (**Fig. 6d**) to promote *HBG1* induction during the early stages of erythroid differentiation.

DISCUSSION

CheRNAs are operationally defined by statistically significant enrichment in chromatin upon biochemical fraction of nuclei. Here, we find that cheRNAs are largely cell type specific and that their presence is more highly correlated with *cis*-gene expression than other metrics of enhancer annotation. In human cells, the majority of genes that cheRNAs abut are tissue restricted, suggesting potential roles in lineage differentiation or maintenance. Beyond this correlation, we have demonstrated a functional role of several cheRNAs in promoting proximal-gene expression.

Despite the modest overlap between cheRNA, eRNA and lncRNA transcripts (**Supplementary Fig. 2a**), our approach may also capture the *cis*-acting subpopulations of the latter two classes of molecules. Several lines of evidence support the concept that *cis*-regulatory-element transcription mediates enhancer activity[5,9,10,14,47–50]. Whether apparent ncRNA distinctions such as length or bidirectional transcription are functionally consequential remains a crucial question for the field. Given the strong correlation of ncRNA biochemically isolated from chromatin to *cis*-gene transcription[10,14] and data presented herein, a classification based on chromatin enrichment may prove to be a more faithful metric of enhancer function and could be a powerful adjunct to the use of other chromatin signatures[26–28] in *de novo* enhancer prediction.

Chromatin looping from cheRNAs to tether enhancers to target promoters

The high correlation of gene activation with downstream sense cheRNAs suggests a model in which pioneering rounds of transcription that bypass normal termination could potentiate the transcription of a downstream enhancer. The cheRNA product could facilitate looping from the newly activated enhancer to the gene promoter (**Fig. 6d**), setting up a feed-forward loop for stable expression analogous to the gene loops described in yeast[51]. Including prior experiments with *PVT1* (ref. 35), knockdown of three out of four cheRNAs in the downstream sense orientation using CRISPRi led to a decrease in expression of their upstream neighbors. However, activation of *IL6* transcription by the upstream divergent *ILYICH* cheRNA indicates that this orientation is not an absolute requirement. Our more detailed analysis of the *HBG1-HIDALGO* locus supports the model of pioneering read-through transcription of the coding gene to potentiate downstream enhancer transcription. The granular kinetics of transcriptional activation through the *HBG1-HIDALGO* locus upon erythroid differentiation, where both transcripts increase seemingly in lockstep, is consistent with this model. In particular, ASO depletion of the cheRNA HIDALGO, some of which represent readthrough transcripts from the upstream *HBG1* promoter, led to a decrease of *HBG1* transcripts far greater than can be accounted for by depletion of only the readthrough pool. By targeting the cheRNA for cleavage without altering its transcription or changing the underlying DNA sequence, we demonstrate that, at least in this scenario, the RNA molecule itself is also important in promoting *cis*-enhancer activity. Knockdown of *HIDALGO* by either ASO or CRISPRi led to decreased chromatin contacts between the enhancer at the TSS of one of the HIDALGO isoforms with the promoter of *HBG1*, supporting a role for the RNA in bridging these two elements to facilitate successive rounds of transcription.

Among mechanisms previously described in the literature, that involving, estrogen-inducible eRNA molecules tethered near distal enhancers that promote transcriptional activation of gene targets[9] is most similar to the HIDALGO-*HBG1* mechanism. As with HIDALGO, changes in locus architecture occur in response to both small-molecule activation and ASO-mediated depletion of eRNA. Looping is an implied function of several lncRNA-coding-gene paradigms as well,

but direct evidence of the transcript acting *in cis* has remained elusive. The class of molecules termed 'ncRNA-a' play important roles in chromatin looping through the transcriptional co-activator complex mediator and RNA-processing complex integrator[5,10], but we observed no requirement for these factors in HIDALGO function (**Supplementary Fig. 7**). Moreover the susceptibility of neighboring transcriptional effects of ncRNA-a and related lncRNA to RNAi[1,5,6,10,25] suggests that they may operate *in trans*[42,52], consistent with the intermediate levels of chromatin enrichment compared to cheRNAs. Rather than altering the local chromatin loop structure, other lncRNAs may promote neighboring gene transcription by recruiting methyltransferase complexes to install the transcriptionally activating histone modification H3K4me3 (ref. 1). In the case of the transcriptional and architectural perturbations of HIDALGO by ASOs, the levels of H3K4me3 do not change appreciably at the *HBG1* promoter (**Fig. 6b**), arguing that similar mechanisms are not functionally relevant in this case.

Our data are consonant with the model that promoters of lncRNA may act as enhancer elements, as observed with a recent, elegant report of allele-specific engineering of five lncRNA loci that act *in cis* to enhance proximal gene expression[13]. However, unlike the *HIDALGO–HBG1* gene pair, the functional mechanisms are apparently independent of the RNA molecule itself. We infer *cis* activity of cheRNA as they are predominantly attached to chromatin through the act of their transcription[14], and we observed one cheRNA still linked to its site of production (**Fig. 2**). Furthermore, perturbing cheRNA transcription often negatively impacts neighboring gene transcription and, in one example, the chromatin architecture coupling an enhancer to the promoter of the neighboring gene is altered when the cheRNA molecule is cleaved (**Figs. 3, 5** and **6**). Definitive proof of *cis* activity of HIDALGO and other cheRNAs requires allele-specific engineering and testing. Further investigation is also needed to precisely define the mechanisms by which cheRNAs promote neighboring gene activation and to explore potential repressive functions as reported for other lncRNAs[4,52–54].

Despite their overall correlation with gene activation, different cheRNAs are unlikely to function by identical mechanisms. One of the cheRNAs we examined, PAINE, does not significantly affect transcription of its nearest neighbor (**Fig. 3c,f**), although the present data do rule out a role for PAINE in activating more distal loci or the possibility that PAINE plays no role in transcriptional activation. Nevertheless, four out of five cheRNAs were observed to potentiate transcription of their neighboring gene, and this observation, together with the earlier observation of CRISPRi depletion of PVT1 (ref. 38), argues for a more general function of cheRNA.

A possible evolutionary origin for cheRNA transcription

Class I transposable elements (TEs) carry their own promoters and might provide an evolutionary origin of cheRNAs similar to other lncRNAs[55,56]. Indeed, 96% of K562 and 98% of H1 CAGE-supported cheRNA overlap with class I TEs. While this enrichment is similar to Gencode lncRNAs bearing CAGE peaks (**Supplementary Fig. 8b**), there is only modest correspondence between class I TEs and enhancers annotated by either chromatin signatures (7–38%) or eRNAs (9–15%) (**Supplementary Fig. 8b**). Intriguingly, we also identified an ~800-bp region in HIDALGO that corresponds to the insertion of three primate-specific class I TEs (L1PA11, MER41A, and L1P3) during the split between simians and prosimians (**Supplementary Fig. 8a,c**) ~35–55 million years ago. It is possible that insertion of these endogenous retroviruses introduced regulatory elements controlling the transition from hemoglobin γ to β, which occurs only during simian primate development[57]. In support of this hypothesis, a reporter construct containing the *HIDALGO* promoter supported a >80-fold induction of luciferase (**Fig. 6c**), whereas a longer

promoter fragment containing these TEs displayed a 4.4-fold decrease in luciferase expression (**Supplementary Fig. 8d**). Future experiments will address whether these elements contain repressors that contributed to fetal hemoglobin switching during primate evolution, similar to a recently described contribution of TEs to innate immune response[58].

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
A.J.R. and M.S.W. designed the study and wrote the paper with valuable input from all of the other authors. M.S.W. performed the experiments for **Figures 1** and **2** and the initial experiments for **Figures 3, 4**, and **5**. M.A.S. repeated most of the **Figures 3, 4**, and **5** experiments in higher replicate with assistance from M.S.W., and performed 3C with ASO-treated cells. R.N.S. and A.T.G. performed ICeChIP experiments. R.D.N. performed the luciferase assay to examine enhancer activity of the HIDALGO TSS with oversight from I.P.M. V.G. cultured H1 ESCs used.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Wang, K.C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
2. Yang, L. *et al.* lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* **500**, 598–602 (2013).
3. Dinger, M.E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18**, 1433–1445 (2008).
4. Rinn, J.L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
5. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
6. Trimarchi, T. *et al.* Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell* **158**, 593–606 (2014).
7. Feng, J. *et al.* 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 (2006).
8. Xiang, J.-F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **24**, 513–531 (2014).
9. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
10. Lai, F., Gardini, A., Zhang, A. & Shiekhattar, R. Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**, 399–403 (2015).
11. Davidovich, C. *et al.* Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol. Cell* **57**, 552–558 (2015).
12. Mao, Y.S., Sunwoo, H., Zhang, B. & Spector, D.L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95–101 (2011).
13. Engreitz, J.M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
14. Werner, M.S. & Ruthenburg, A.J. Nuclear fractionation reveals thousands of chromatin-tethered noncoding RNAs adjacent to active genes. *Cell Rep.* **12**, 1089–1098 (2015).

15. Petruk, S. *et al.* Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* **127**, 1209–1221 (2006).
16. Schmitt, S., Prestel, M. & Paro, R. Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev.* **19**, 697–708 (2005).
17. Melé, M. & Rinn, J.L. "Cat's cradling" the 3D genome by the act of lncRNA transcription. *Mol. Cell* **62**, 657–664 (2016).
18. Paralkar, V.R. *et al.* Unlinking an lncRNA from its associated cis element. *Mol. Cell* **62**, 104–110 (2016).
19. Bhatt, D.M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
20. Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* **14**, 7219–7225 (1994).
21. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
22. Postepska-Igielska, A. *et al.* LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. *Mol. Cell* **60**, 626–636 (2015).
23. Sunwoo, H., Wu, J.Y. & Lee, J.T. The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. *Proc. Natl. Acad. Sci. USA* **112**, E4216–E4225 (2015).
24. Cabili, M.N. *et al.* Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20 (2015).
25. Ørom, U.A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
26. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
27. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
28. Heintzman, N.D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
29. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
30. Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T. & Steitz, J.A. Widespread inducible transcription downstream of human genes. *Mol. Cell* **59**, 449–461 (2015).
31. Na, J., Furue, M.K. & Andrews, P.W. Inhibition of ERK1/2 prevents neural and mesendodermal differentiation and promotes human embryonic stem cell self-renewal. *Stem Cell Res.* **5**, 157–169 (2010).
32. de Groot, R.P., Raaijmakers, J.A., Lammers, J.W., Jove, R. & Koenderman, L. STAT5 activation by BCR-Abl contributes to transformation of K562 leukemia cells. *Blood* **94**, 1108–1112 (1999).
33. Rao, S.S.P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
34. Tseng, Y.-Y. *et al.* PVT1 dependence in cancer with MYC copy-number increase. *Nature* **512**, 82–86 (2014).
35. Fulco, C.P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
36. Robinson, P.J. & Rhodes, D. Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr. Opin. Struct. Biol.* **16**, 336–343 (2006).
37. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
38. Gilbert, L.A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
39. Addya, S. *et al.* Erythroid-induced commitment of K562 cells results in clusters of differentially expressed genes enriched for specific transcription regulatory elements. *Physiol. Genomics* **19**, 117–130 (2004).
40. Bodine, D.M. & Ley, T.J. An enhancer element lies 3′ to the human A gamma globin gene. *EMBO J.* **6**, 2997–3004 (1987).
41. Purucker, M., Bodine, D., Lin, H., McDonagh, K. & Nienhuis, A.W. Structure and function of the enhancer 3′ to the human A γ globin gene. *Nucleic Acids Res.* **18**, 7407–7415 (1990).
42. Maamar, H., Cabili, M.N., Rinn, J. & Raj, A. linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes Dev.* **27**, 1260–1271 (2013).
43. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
44. Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G. & Rauscher, F.J. III. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932 (2002).
45. Grzybowski, A.T., Chen, Z. & Ruthenburg, A.J. Calibrating ChIP-Seq with nucleosomal internal standards to measure histone modification density genome wide. *Mol. Cell* **58**, 886–899 (2015).
46. Hansen, K.H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**, 1291–1300 (2008).
47. Andersson, R., Sandelin, A. & Danko, C.G. A unified architecture of transcriptional regulatory elements. *Trends Genet.* **31**, 426–433 (2015).
48. Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
49. Lam, M.T.Y. *et al.* Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**, 511–515 (2013).
50. Arner, E. *et al.* Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010–1014 (2015).
51. O'Sullivan, J.M. *et al.* Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* **36**, 1014–1018 (2004).
52. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011).
53. Pandey, R.R. *et al.* Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
54. Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
55. Kapusta, A. *et al.* Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).
56. Lynch, V.J. *et al.* Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Rep.* **10**, 551–561 (2015).
57. Johnson, R.M. *et al.* Fetal globin expression in New World monkeys. *J. Biol. Chem.* **271**, 14684–14691 (1996).
58. Chuong, E.B., Elde, N.C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083–1087 (2016).
59. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
60. Forrest, A.R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).

ONLINE METHODS

Cell culture and fractionation. H1 hESCs were grown feeder free on Matrigel (BD Bioscience) in StemPro media (Invitrogen). K562 cells were maintained at ~0.1–1 × 10^6 cells/ml in RPMI 1640 (Gibco), 2 mM Glutamine, 10% FBS, 1% penicillin/streptomycin. 'Plus hemin' cells were treated with freshly prepared 50 µM hemin (Chem IMPEX International) at indicated time points. H1 cells were provided by V. Galat (Northwestern University), and K562 cells were provided by J. Weissman (UCSF). Cell lysis, nuclear fractionation, and RNA isolation were performed as previously described on three independent cultures of 10^7 K562 cells or H1 hESCs[14]. Briefly, purified nuclei were extracted with 0.5 M Urea and 0.5% NP-40 substitute to solubilize loosely bound factors from chromatin and fractionated by centrifugation. RNA from both the chromatin pellet (CP) and soluble nuclear extract (SN) were obtained by Trizol extraction (Life Technologies) and further purified by RNA-Clean and Concentrator columns (Zymo Research) with in-column DNase I digestion as described in the manufacturer's protocol. *In vitro* transcribed RNA standards (below) were added to purified chromatin pellet and soluble nuclear extract RNA isolates, ribosomal RNA was depleted using Ribo-Zero Gold (Illumina), and stranded cDNA libraries were made using NEBNext Ultra Directional DNA Library Prep Kit for Illumina and sequenced on an Illumina HiSeq2000. K562 and H1 hESC libraries were sequenced by single-end 100-bp reads, and two replicates of hemin-treated K562 cell libraries were sequenced with single-end 50-bp reads.

Calibrated RNA-seq. Spike-in standards were *in vitro* transcribed with recombinant T7 polymerase[61] and were selected based on lack of homology to human genes and length similarity within the set (777–1,290 nucleotides, **Supplementary Fig. 1a**). RNA was purified with Zymo RNA-Clean and Concentrator columns, serially diluted in a buffer containing 50 mM NaCl, 0.01% NP-40 substitute, 100 ng/µl pUC19, 10 mM Tris-HCl pH 7.5, and 1 mM EDTA, and added to CP and SN RNA before rRNA depletion with Ribo-Zero Gold (Illumina). The four RNA standards were added at $2.7 × 10^6$, $9 × 10^5$, $3 × 10^5$, and $1 × 10^5$ copies per K562 library, and $9 × 10^5$, $3 × 10^5$, $1 × 10^5$, and $3 × 10^4$ copies per H1 library to create calibration curves. We performed linear regression of the absolute read counts from RNA-seq versus the number of molecules of RNA standard added per cell number equivalent to each library (calculated from the number of cells that each extract was derived from, **Supplementary Fig. 1b**). The resulting linear fit equation was used to compute the approximate molecules per cell for cheRNAs based on absolute read counts for each pool (soluble, chromatin pellet) for each biological triplicate (**Supplementary Fig. 1c**) and to confirm chromatin versus soluble nuclear extract enrichment. Details of bioinformatics analysis is presented in **Supplementary Note 1**.

Reverse transcription and RT-qPCR. Reverse transcription of isolated total RNA was performed in 20 µL reactions using 0.5 µg (LNA ASO, *HIDALGO* knockdown experiments) or 1 µg (all other experiments) total RNA with 100 ng random hexamers (IDT) and 100 U MMLV-HP Reverse Transcriptase (Epicentre) according to manufacturer's instructions. RNA was degraded by 100 mM KOH + 13.3 mM Tris base (final concentration) and incubated at 95 °C for 10 min. Afterwards, the pH was adjusted to ~8.0 using 150 mM HCl, and samples were diluted with 50–200 µL TE buffer (10 mM Tris-HCl + 1 mM EDTA-NaOH, pH 8.0). Real time quantitative PCR (RT-qPCR) was performed with Power SYBR Green PCR Master Mix (Applied Biosystems) and 2–4 µL cDNA per reaction, with 250 nM each primer on a Bio-Rad CFX384 instrument. Three or four technical replicates of each reaction were performed, and all independent replicate targeting and corresponding negative control samples were queried on the same plate.

For CRISPRi in **Figure 3**, a total of six unique reference gene candidates (18S rRNA, GAPDH, PGK1, PPIB, TBP, SDHA were evaluated for their relative stability via calculation of the geNorm M value[62] parameter (qbase+ software, Biogazelle). Although these genes exhibit differing absolute stability ranks between the aforementioned groups of samples (i.e. BONIFACIO-targeting samples versus ILYICH-targeting samples), the scale of differences between these candidates across different samples was generally small, and the relative relationships of target genes of interest between samples were generally robust to the reference gene choice. *PPIB* was ultimately chosen for use across all **Figure 3** sample groups on the basis of consistency of its relative expression level ($2^{ΔCt}$) across negative control experiments. The amplification efficiency and factor of all primers corresponding to **Figure 3** were measured using a 5- to 10-point two-fold

dilution series of select cDNA samples, where three independent dilution series replicates were performed for *PPIB* and the amplification factor was calculated as the average across replicates. The amplification factor of each primer set was used as the base for exponentiation of the respective amplicon's mean C_t value when calculating "$2^{ΔCt}$". 18S rRNA was used as the reference gene for all other RT-qPCR displayed.

Fold differences ($2^{ΔΔCt}$) of targeting samples were calculated relative to the respective negative control samples, with scaling of the mean negative control expression level to 1. Throughout figures displaying relative fold changes, the s.e.m. for targeting samples includes the propagated uncertainty of the mean negative control expression level used for calculating fold differences.

Statistics. For data presented in **Figure 1** P values were calculated via the Mann-Whitney-Wilcoxon test in R. For data presented in **Figures 3**, **4c**, and **5c,d** and **Supplementary Figure 6c,d**, P values were calculated via two-tailed Welch's t test in R. Except for **Figure 5c,d**, distributions of raw RT-qPCR data ($2^{ΔCt}$, in all instances averaged from three or four technical qPCR replicates per plate) from all measurements of all independent experiments, before conversion to fold-change values, were compared for significance testing. For **Figure 5c,d**, the compared distributions consisted of data following conversion to relative fold-change values. Data shown in **Figures 3d,f** are calculated from averaging across all independent experiments and two (**Fig. 3d,f**, sgRNAs 1–4 and '–') or four (**Fig. 3f**, sgRNAs 5–6 and '–') independent qPCR plates assaying these experiments. Otherwise, data from multiple independent experiments are calculated from averaging of single qPCR plate measurements across the replicate experiments alone (i.e. **Figs. 3e** and **4c**), or across qPCR technical replicates if only single independent experiments were performed (i.e. **Fig. 5b,e**). **Figure 3d,e** correspond to the n = 3 and n = 6 targeting and negative control experiment counts, respectively, in the stated **Figure 3n** range. **Figure 3f** corresponds to the stated n = 4 and n = 7, or n = 1 and n= 2, experiment counts. For **Figure 4c** and **Supplementary Figure 6c–d**, t = 0, 2, and 4 h data points correspond to four independent experiments, while t = 8, 12, and 24 h data points correspond to three independent experiments.

Characterization of *HIDALGO* transcripts. The initial evidence for several *HIDALGO* transcripts from our CPE sequencing and splice sites detected therein was further supported by 5′ and 3′ RACE using gene-specific primers, CAGE-seq peaks[60], and RT-qPCR. 5′ RACE was performed using SMARTER 5′/3′ RACE kit (Clontech) following manufacturers protocols. In brief, reverse transcription was performed with random-hexamer primers, and then PCR was performed with Clontech adapters and imaged on a 1% agarose gel stained with ethidum bromide. RACE was performed on either total RNA or chromatin pellet, which yielded similar results. Relative *HIDALGO* transcript amounts were assessed by RT-qPCR with several primer sets (**Supplementary Fig. 4c,e,f**), some of which are isoform specific in that they span spatially disparate exon-exon junctions, and some of which should detect all isoforms. Measurement of relative isoform abundance requires synthesis of direct and indirect evidence: primer sets that detect both isoforms #1 and 2, isoform #3 alone, and the composite of #1–4, respectively. As there are no unique splice sites within *HIDALGO* isoform #4 that enable selective detection, its levels are inferred by comparison of primer set #1–4 to those that detect #1–2 and #3.

The consensus TATA box is "TATAWAWR" (where W = A/T, R = A/G[63]), and there is support for binding of this motif by TFIIB of the PIC 10–30 bp upstream of the exact site of initiation with 0 or 1 mismatches[64]. The cryptic TATA box for *HIDALGO* TSS #4 is "TATAAGTA" which has one purine→purine mismatch relative to consensus, and both 5′-RACE evidence (**Supplementary Fig. 4e**) and CAGE-seq (**Fig. 4a**) suggest that this element is 137 bp upstream of the +1 base. Moreover, the hallmarks of transcriptional initiation[21,63,65] are present at this site in ENCODE data sets (H3K4me3, a Pol II peak, DNAse I hypersensitivity, histone acetylation, TF binding sites, **Fig. 4a**).

cheRNA knockdowns. CRISPRi was performed in K562 or HEK293 cells with dCas9-KRAB integrated in the genome[38,66]. K562 CRISPRi cells were generously provided by L. Gilbert and J. Weissman (UCSF), and HEK293 CRISPRi cells were generated by Lipofectamine 2000 (Invitrogen) transfection of a modified dCas9-KRAB vector flanked by an FRT site and containing a hygromycin resistance gene, into HEK293 Flp-In (Invitrogen) cells, followed by greater than two weeks of continual hygromycin resistance (100 µg/ml). sgRNAs were designed

by eCRISP(http://www.e-crisp.org/), cloned into a modified px330 (refs. 38,67) vector containing eGFP and a modified stem loop designed to increase binding to dCas9 (ref. 68), and verified by Sanger sequencing. Twenty-four hours before transfection, cells were seeded to 6-well plates with $5.0–8.0 \times 10^5$ cells/well (K562 CRISPRi) or 1.2×10^6 cells/well (HEK293 CRISPRi) in either RPMI 1640 (Mediatech Inc./Corning Cellgro), 2mM Glutamine, 10% FB Essence (Seradigm), 1% penicillin/streptomycin for K562 CRISPRi cells, or DMEM (Gibco) 10% FB Essence (Seradigm), 1% penicillin/streptomycin for HEK293 CRISPRi. For each transfection, 10 µl of Lipofectamine 2000 reagent was diluted in 250 µl Opti-MEM Reduced Serum Media (Gibco), and 4 µg of plasmid DNA was diluted in 250 µl Opti-MEM. Diluted Lipofectamine and DNA were combined, mixed by pipetting, and incubated for 10 min at room temperature, then added drop-wise to cells without removing media. After two days cells were re-plated on 10 cm plates.

Four to six days post-transfection, cells were removed from plates (we found this an optimal time-span for knock-down, before which there might not have been enough time, and after which there were too few remaining GFP+ cells for subsequent experiments (i.e. RT-qPCR, ChIP)), centrifuged 5 min 500 × g, 4 °C, and re-suspended in 1-2 ml fresh media, and then GFP positive cells were isolated by FACS (Aria II/III, BD). Transfection conditions, outgrowth and sorting for a given experiment with all relevant controls were performed identically side-by-side. The majority of K562 CRISPRi experiments were GFP+ sorted and harvested five days post-transfection.

Sorted cells were pelleted (5 min, 500 × g, 4 °C, and re-suspended in 500 µl Trizol (Life Technologies). The aqueous layer from Trizol extraction was applied to RNA Clean & Concentrator-25 columns (Zymo Research), and then converted into cDNA as described above. Polyclonal K562 cell-lines with sgRNAs incorporated into the genome were transfected as above with Lipofectamine 2000 with the same vector used for transient transfections except eGFP was replaced with a puromycin resistance gene. Selection was performed with puromycin (6.7 µg/ml) two days after transfection and continued for two weeks. All comparisons (**Figs. 3, 5 and 6** and **Supplementary Figs. 3a–c** and **5c,d**) were made relative to an off-target 21 nt negative control sgRNA referred to as "negative control 1" in Gilbert *et al.* 2013 (ref. 66), and referred to in figures in this manuscript as "–", except for the experiments shown in **Figure 5e**, where hemin-induction of stable cell lines were made relative to "normal" K562 cells. The closest match to the negative control sgRNA is 15 nt of hybridization to a protocadherin (*PCDH17*), that is missing a PAM sequence and is not expressed in K562 cells.

ASO knock-downs were performed with 50 nM (final concentration) LNA-FAM gapmers (Exiqon) transfected with Lipofectamine 2000. Oligos were diluted in 250 µl Opti-MEM, and then incubated 10 min room temperature with 10 µl Lipofectamine 2000 reagent diluted in 250 µl Opti-MEM (500 µl final) before being added to 3 ml of cells in RPMI 1640 media in 6-well culture plates. Similar to CRISPRi transfections, cells were passaged to 10 cm plates in 10 ml fresh media after two days, and then grown for 3 more days (5 total post-transfection). GFP positive cells were FACS-sorted on the FAM channel (Aria II/III, BD), and subjected to RNA isolation, ChIP, or 3C. Subtle effects on *HBG2* were also observed for ASO-1, potentially attributable to substantial sequence similarity between the two mRNAs, a similar transcribed region of homology (15/16 bp match) downstream of *HBG2* (**Fig. 4a**), or more widespread effects of *HIDALGO*. All ASO comparisons are made relative to a non-targeting LNA-FAM (Exiqon) that was transfected in parallel, referred to as "ASO –" in figures and **Supplementary Table 1**.

Chromosome conformation capture (3C). 3C was performed on LNA ASO HIDALGO KD samples (**Fig. 6a**) mostly as described[69] with the following modifications. K562 cell samples were crosslinked with 2% (w/v) formaldehyde (made fresh from paraformaldehyde, Sigma-Aldrich P6148) in 1X PBS supplemented with 10% (v/v) FB Essence (Seradigm) at room temperature for 10 min. After quenching with 125 mM glycine, cells were pelleted (225 rcf, 8 min, 4 °C) and lysed via resuspension and incubation in ice-cold lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 5 mM MgCl$_2$, 0.1 mM EDTA, 0.1% Triton X-100, 1X Roche Protease Inhibitor Cocktail) for 10 min on ice. Nuclei were pelleted (400 rcf, 5 min, 4 °C), frozen in liquid nitrogen, and stored at −80 °C. Thawed nuclei pellets were resuspended in 0.5 mL Buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 4 mM MgCl$_2$, 340 mM sucrose, 10% (v/v) glycerol, 1 mM DTT, 1X Roche Protease Inhibitor Cocktail) + 0.12% (w/v) SDS and incubated at 37 °C for 1 h while shaking at 900 rpm. Triton X-100 was added to 1% (v/v) final and samples

were incubated at 37 °C (1 h, 900 rpm in Eppendorf Thermomixer). Homogeneity of the nuclei suspension was maintained via pipetting every 20 min during both detergent incubations. Nuclei were pelleted (500 rcf, 10 min, room temperature), and 0.4 mL of the supernatant was removed. The pellet and remaining volume of supernatant were resuspended in 0.25 mL final of 0.5X Buffer A + 1X CutSmart Buffer (NEB) with the concentration of protease inhibitor adjusted to 1X final. DNA digestion was performed at 37 °C (12 h, 900 rpm) with initially 100 units of NspI (NEB), plus 20 more units after 1.5 h.

Following digestion, SDS was added to 1.6% (w/v) final and samples were incubated at 65 °C (25 min, 900 rpm). Each sample volume was then mixed with 3.063 mL of 1.15X T4 DNA Ligase Reaction Buffer (NEB, 1X = 50 mM Tris-HCl pH 7.5, 10 mM MgCl$_2$, 1 mM ATP, 10 mM DTT) + 1% (v/v) final Triton X-100 and incubated at 37 °C for 1 h with rotation. Ligation was performed at 16 °C for 4 h followed by a 30 min incubation at room temperature using, initially, 100 units of T4 DNA Ligase (Promega HC), plus 30 units more after 2 h. After ligation, the volume of each sample was increased to 7.0 mL with 50 mM Tris-HCl pH 8.0, with addition of EDTA pH 8.0 to a final concentration of 5 mM and adjustment of SDS concentration to 0.5% (w/v). De-crosslinking was performed with 300 µg Proteinase K (Invitrogen) and incubation at 60 °C for 14 h, plus another 300 µg Proteinase K after the first 3 h (600 µg total). No RNase step was performed.

Afterwards, samples were frozen in liquid nitrogen and lyophilized. Lyophilized samples were resuspended in 1 mL 10 mM Tris-HCl pH 8.0 and heated briefly at 65 °C until fully solubilized. DNA was purified via phenol-chloroform extraction, with three chloroform back-extractions of the aqueous phase, and subsequent ethanol precipitation with glycogen. A second ethanol precipitation of the resuspended pellet was required to eliminate residual contaminating species with absorbance near 230 nm.

RT-qPCR was performed using 200 ng per reaction of this material in triplicate, with normalization to GAPDH amplicon. GAPDH reactions were performed using Power SYBR Green PCR Master Mix (Applied Biosystems) and final primer concentrations of 250 nM (each). Ligation products were detected using a TaqMan probe that anneals to a common region and differing primer pairs corresponding to the different contact loci (see **Supplementary Table 1** for all sequences). TaqMan reactions were performed using TaqMan Gene Expression Master Mix (Applied Biosystems) with final primer and probe concentrations of 500 nM and 250 nM, respectively. 3C was performed using four independent experiments of each the HIDALGO-targeting ASO and non-targeting negative control ASO ("ASO 1" and "ASO –", respectively, see **Supplementary Table 1**), and a single "ASO 1" outlier was detected and excluded at 95% confidence using Dixon's Q test. Shaded regions of each curve in **Figure 6a** correspond to s.e.m. from independent experiments (solid line is arithmetic mean), where for the "ASO 1" curves this s.e.m. includes the propagated uncertainty of the "ASO –" mean value used for conversion of the contact frequencies into relative values (with scaling of the maximum "ASO –" mean value to 1).

3C was performed on CRISPRi HIDALGO KD samples (**Supplementary Fig. 5c**) as described above with the following differences. Following the first 0.12% (w/v) SDS and 1% (v/v) Triton X-100 incubations, nuclei were spun through a 5 mL sucrose cushion (10 mM HEPESKOH, pH 7.9, 30% (w/v) sucrose, 4 mM MgCl$_2$), then resuspended in 0.5X Buffer A + 1X CutSmart Buffer (NEB). De-crosslinking was performed with addition of 200 µg Proteinase K (Invitrogen) and incubation at 65 °C for 5 h. Samples were then treated with 150 µg RNase A for 45 min at 37 °C. Ethanol precipitation was performed as described[69].

Single molecule FISH. HEK293 cells were grown on acid-washed coverslips in 6 well plates. Stellaris FISH Probes (Biosearch Technologies, Inc., Petaluma, CA) against PVT1 exons and the two MYC introns, labeled with Cy5 and Cy3, respectively, were designed with the Stellaris FISH Probe Designer (http:// www.biosearchtech.com/stellarisdesigner). HEK293 cells were hybridized with the PVT1 and MYC intron smFISH Probe sets following the manufacturer's instructions available online at http:// www.biosearchtech.com/stellarisprotocols, and imaged on a Zeiss Axiovert 200M inverted wide-field microscope in the UChicago Microscopy Core Facility. Three-dimensional *z*-stacks of images were flattened and background-subtracted using ImageJ software. Foci were identified as local maxima, then regions of interest were drawn around cells containing MYC intron or PVT1 foci. Cells that contained an apparent MYC intron were recorded and compared to cells containing PVT1 foci. In cells bearing foci for each of the RNA, distances between centers of mass of each MYC intron 2 focus and the nearest

PVT1 foci were measured using Object-based methods in the JACoP plugin[70], and a histogram of nearest-distances was plotted in R.

Luciferase assays. Luciferase response assays were performed as previously described[71]. Candidate elements were amplified from HEK293T genomic DNA or synthesized via Gibson Assembly. Sequence was verified and then cloned into the pGL4.23 enhancer luciferase response vector with minimal promoter. K562 immortalized cells were co-transfected with luciferase response vector and a pRL renilla luciferase control using Lipofectamine 3000, cultured for 48 h after transfection, then lysed and assayed using the Dual-Luciferase Reporter Assay system (Promega).

Internal standard calibrated chromatin immunoprecipitation (ICeChIP). ICeChIP was conducted as previously described[45], with some modifications. Sorted cell pellets (see cheRNA knockdowns), once harvested, were washed with 1 mL ice-cold PBS, then with 1 ml ice-cold Buffer N (15 mM Tris-HCl pH 7.5, 15 mM NaCl, 60 mM KCl, 8.5% w/v sucrose, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM DTT, 200 uM PMSF, 50 µg/mL BSA, 1× Roche Protease Inhibitor Cocktail) and pelleted at 500 g for 5 min at 4 °C. The cell pellets were then resuspended in 2 volumes of Buffer N and lysed by adding 1 volume (i.e. 3 PCV) of 2× Lysis Buffer (Buffer N supplemented with 0.6% v/v NP-40 Substitute, Sigma) and incubating for 10 min at 4 °C. Following lysis, nuclei were pelleted at 500 g for 5 min at 4 °C, and the nuclei were resuspended in 6 volumes of Buffer N. To quantify nuclei, 2 µl of the nuclei suspension was added to 98 µl of 2 M NaCl in triplicate and vortexed vigorously. Total nucleic acid of the nuclei-salt mixtures was then determined spectroscopically.

The nuclei were then spiked with 2 µL of approximately 5 nM nucleosome standards. The spiked nuclei were pre-warmed at 37 °C while shaking at 900 r.p.m. for 2 min, and then 1 Worthington unit of micrococcal nuclease (MNase) was added for every 1 µg of chromatin in the nuclei suspension, and the suspension was incubated at 37 °C while shaking at 900 r.p.m. for 12 min. After digestion, 1/9 volume of 10× MNase Stop Buffer (10 mM EDTA, 10 mM EGTA) was added while vortexing. The nuclei were then lysed by adding 5 M NaCl to a final concentration of 600 mM NaCl while vortexing. The insoluble debris was pelleted at 18,000 g for 1 min at 4 °C. The soluble chromatin was diluted with 1 volume of ChIP Buffer 1 (25 mM Tris pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 10% v/v glycerol, 0.1% v/v NP-40 Substitute).

For the H3K27me3 ICeChIP, 10 µl of Protein G Dynabeads (Invitrogen) were washed twice by resuspension into 200 µl of ChIP Buffer 1, collecting on a magnetic rack. CST9733 antibody (0.6 µg, Cell Signaling) was diluted to 100 µl in ChIP Buffer 1 and added to Dynabeads before incubating on a rotator for at least 1 h at 4 °C. After conjugation, the beads were washed twice with ChIP Buffer 1 and then resuspended in 50 µl of ChIP Buffer 1. For H3K4me3 and H3K9me3 ICeChIP experiments, the biotinylated recombinant Fab was conjugated to M-280 Streptavidin Beads (10 µl, Invitrogen) as previously described[72]. Briefly, 0.6 µg of each recombinant Fab (either 304M3B for H3K4me3, or 309M3B for H3K9me3) were conjugated to pre-washed M-280 resin by incubation in 100 µl of ChIP Buffer 1 with 50 µg/µl BSA for one hour, followed by two washes of the conjugated beads with ChIP Buffer 1 with 50 µg/µL BSA and 5 µM biotin for 15 min each.

Each bead suspension was added to 800 ng of chromatin and incubated on a rotator for 15 min at 4 °C. The beads were then washed twice with 200 µl ChIP Buffer 2 (25 mM Tris pH 7.5, 5 mM $MgCl_2$, 300 mM KCl, 10% v/v glycerol, 0.1% v/v NP-40 Substitute), transferred to a new tube, and incubated on a rotator for 10 min at 4 °C. Washing was repeated two more times with 200 µl ChIP Buffer 3 (10 mM Tris pH 7.5, 250 mM LiCl, 1 mM EDTA, 0.5% sodium deoxycholate, 0.5% v/v NP-40 substitute). The beads were then rinsed with 200 µl ChIP Buffer 1, then 200 µl TE buffer, and resuspended in 50 µl ChIP Elution Buffer (50 mM Tris pH 7.5, 1 mM EDTA, 1% w/v SDS), and incubated at 55 °C for 5 min. The supernatant was collected. ChIP Elution Buffer was also added to the inputs and processed as other samples in all downstream steps.

Samples were adjusted to a final concentration of 200 mM NaCl and 10 mM EDTA. Proteinase K (10 µg) was then added to each elution, and incubated at 55 °C for 2 h. The DNA was then recovered by adding 1.5 volumes of Serapure HD[73] and incubating at room temperature for 15 min, then collecting on a magnetic rack, washing twice on the magnetic rack with 70% ethanol, and eluting into 50 µL of 0.1× TE.

Data availability. All RNA-seq data sets have been deposited in Gene Expression Omnibus (GEO), Series GSE83531. Source data for **Figure 1a** are available in the online version of the paper. Other data are available upon reasonable request.

61. Ichetovkin, I.E., Abramochkin, G. & Shrader, T.E. Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme family and localization to the trypsin-resistant domain. *J. Biol. Chem.* **272**, 33009–33014 (1997).
62. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3** research0034.1–research0034.11 (2002).
63. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**, 233–245 (2012).
64. Siebert, M. & Söding, J. Universality of core promoter elements? *Nature* **511**, E11–E12 (2014).
65. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
66. Gilbert, L.A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
67. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
68. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
69. Hagège, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2**, 1722–1733 (2007).
70. Cordelieres, F.P. & Bolte, S. JACoP v2.0: improving the user experience with co-localization studies. in *Proceedings of the 2nd ImageJ User and Developer Conference* 174–181 (2008).
71. Nadadur, R.D. *et al.* Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. *Science Translational Medicine* **8**, 354ra115 (2016).
72. Hattori, T. *et al.* Antigen clasping by two antigen-binding sites of an exceptionally specific antibody for histone methylation. *Proc. Natl. Acad. Sci. USA* **113**, 2092–2097 (2016).
73. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).