In vivo characterization of *Linc-p21* reveals functional *cis*-regulatory DNA elements

Supplemental Information

Contents

This file contains the following:

- Figures S1-S4
- Extended Experimental Procedures
- Supplemental References
- Supplemental Files 1-4

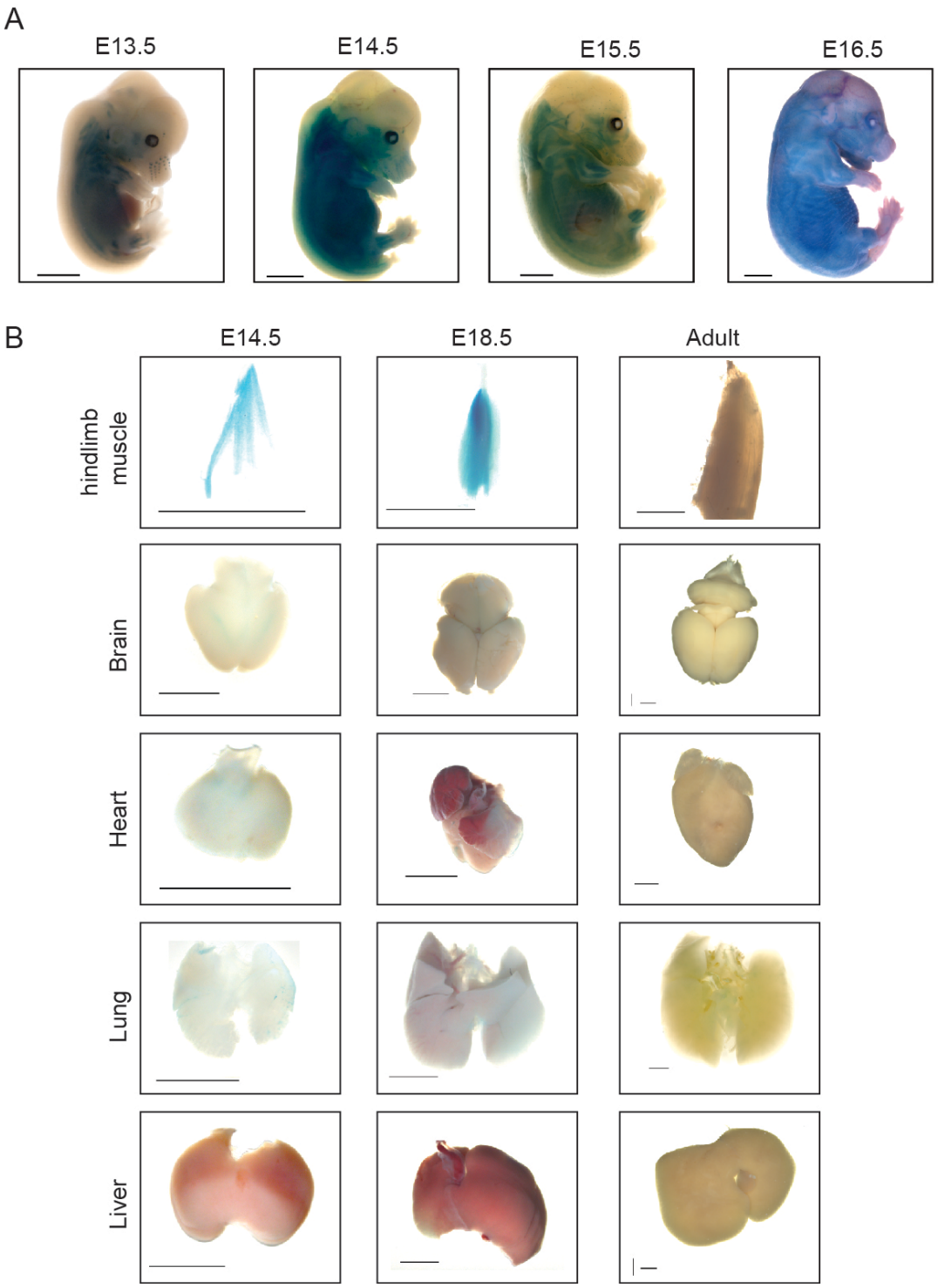


Figure S1: Spatiotemporal analysis of *Linc-p21* expression *in vivo,* related to Figure 1
(A) Representative *Linc-p21* +/- E13.4 (n=10), E14.5 (n=9), E15.5 (n=2), and E16.5 (n=2) mouse embryos. Blue stain is ß-galactosidase activity depicting regions of *Linc-p21* locus activity (scale bar= 2mm). (B) Organs of *Linc-p21* +/- mice at various developmental stages showing ß-galactosidase (*lacZ*) activity (scale bar = 2mm). (B.1-3) Hindlimb tissue for E14.5 (n=2), E18.5 (n=2), and adult (n=2),. B.4-6 Brain: E14.5 (n=2), E18.5 (n=2), and adult (n=2); B.7-9 Heart: E14.5 (n=2), E18.5 (n=2) and adult (n=2); B.10-12 Lung: E14.5 (n=2), E18.5 (n=2), and adult (n=2); and B.13-15 Liver: E14.5 (n=2), E18.5 (n=2) and adult (n=2).

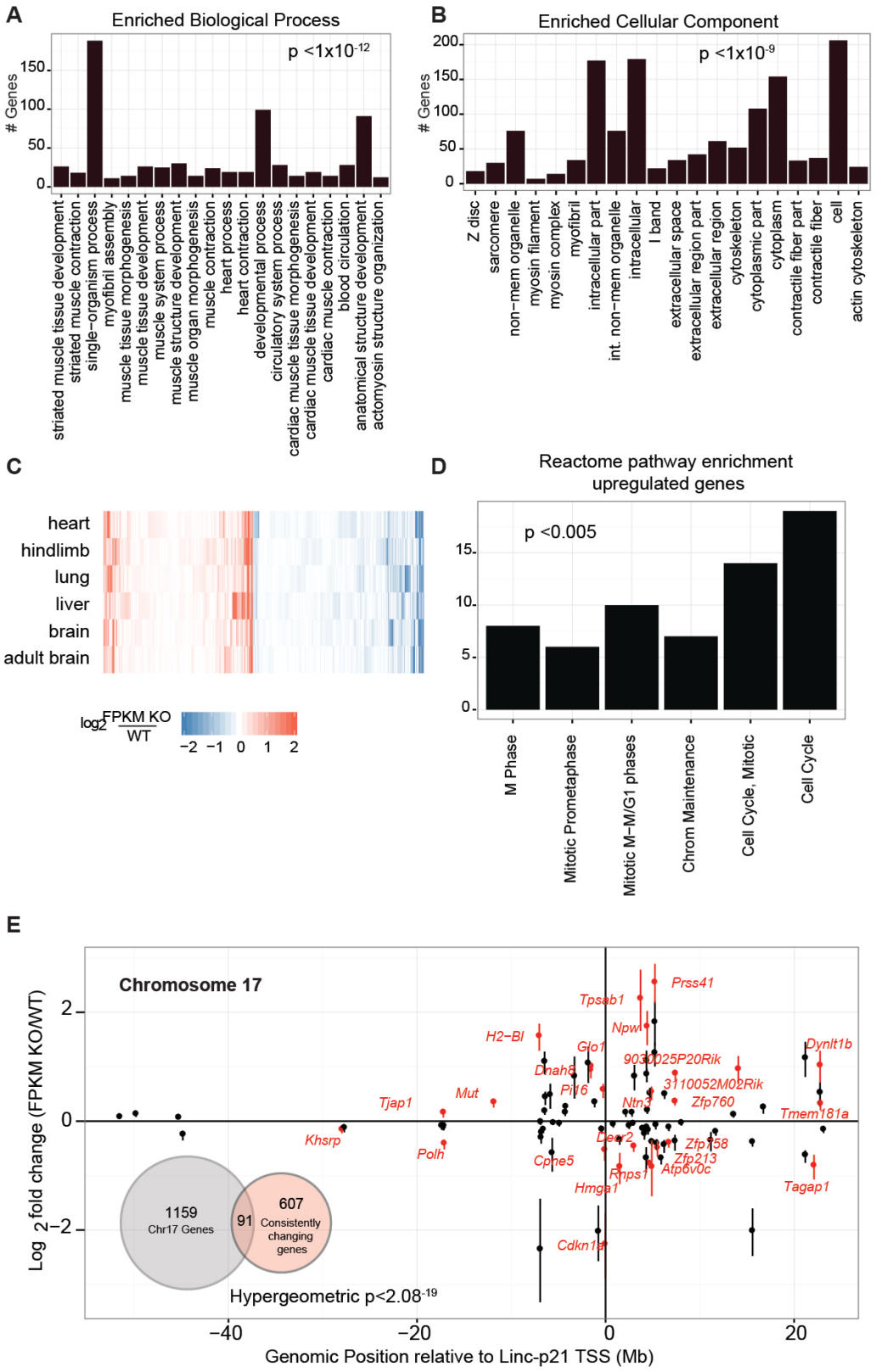


Figure S2: Global and local effects of *Linc-p21* deletion, related to Figures 1-2

(**A**) Summary of enriched biological process GO terms for the 311 genes called significant and aggregated across all tissue comparisons. (**B**) Summary of enriched cellular component GO terms for the 311 aggregated significant genes across tissues. (**C**) An overview of all genes, not just those called significant, that change in the same direction in each tissue comparison. Blue represents a decrease and red represents an increase in *Linc-p21* knockout. Color/intensity represents the $\log_2$ fold change of KO/WT FPKM averaged across replicates for each tissue. (**D**) Reactome pathway enrichment of genes consistently upregulated across all tissues in the knockout mice. (**E**) Effect of *Linc-p21* deletion upon genes on chromosome 17. Plot depicts the genomic position of genes (x axis) that are significantly differentially expressed (in at least one tissue) following *Linc-p21* deletion and is centered at the *Linc-p21* locus. The y-axis represents the average $\log_2$ fold change across all tissues and error bars represent the standard error. Gene names in red are significantly differentially expressed in at least one tissue. Inset venn diagram depicts the overlap between genes that are consistently dysregulated by *Linc-p21* deletion and those on chromosome 17. Statistical significance assessed by hypergeometric test.

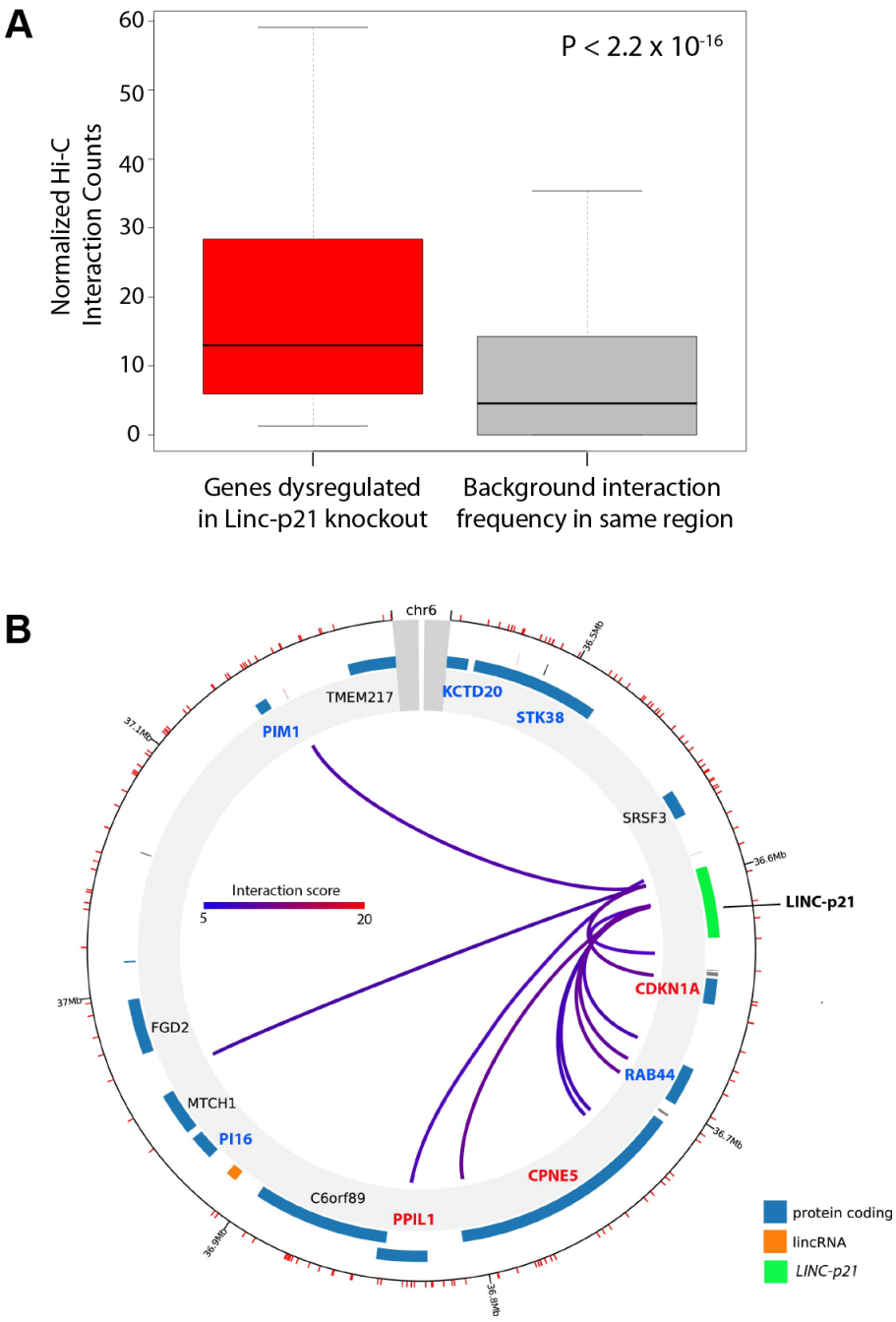
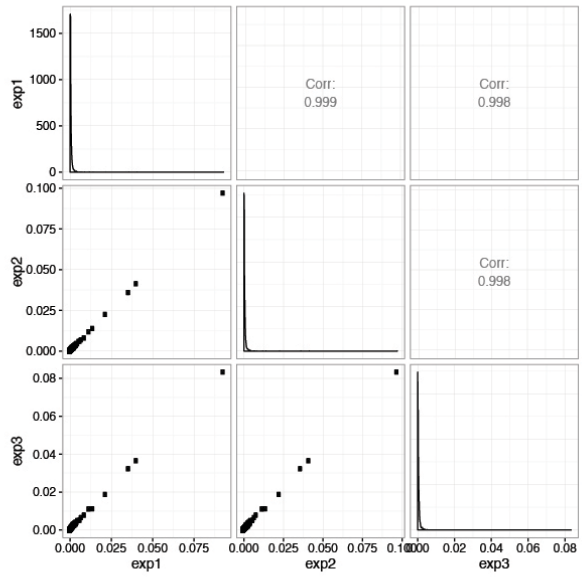


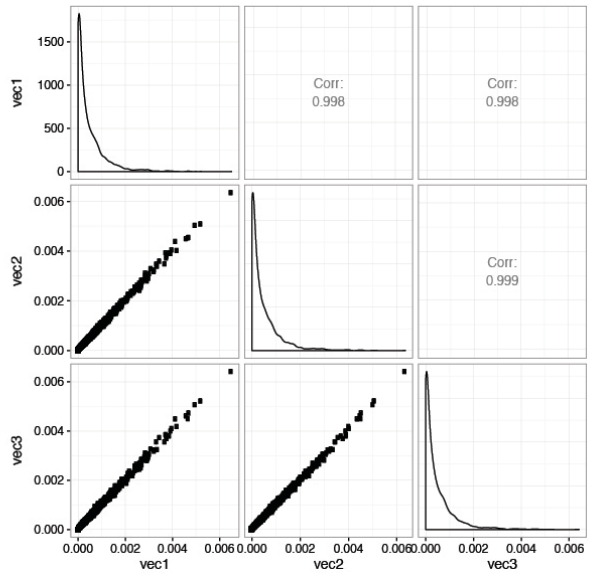
Figure S3: Physical interactions of *Linc-p21* in human and mouse, related to Figure 3

(**A**) Frequency of physical interactions between murine *Linc-p21* and local genes (+/- 2Mb of *Linc-p21*) that were dysregulated across multiple tissues following *Linc-p21* deletion, compared to the background interaction frequency in that region. Mouse embryonic stem cell Hi-C data was binned at 10kb resolution (Dixon et al. 2012). Statistical significance was assessed by Wilcoxon rank sum test. (**B**) Circle plot depicting local interactions of human *LINC-p21* in CD34+ hematopoietic stem cells and GM12878 cells - as identified by capture Hi-C (Mifsud et al. 2015). Gene names in red represent genes that were significantly dysregulated in one or more murine tissue following *Linc-p21*deletion. Gene names colored in blue represent genes that demonstrated a non-significant but similar direction of effect across tissues. Interactions are displayed as colored lines between restriction fragments and are colored according to the study interaction score. Plot adapted from plot generated by CHiCP (Schofield et al. 2016).

A



B



C

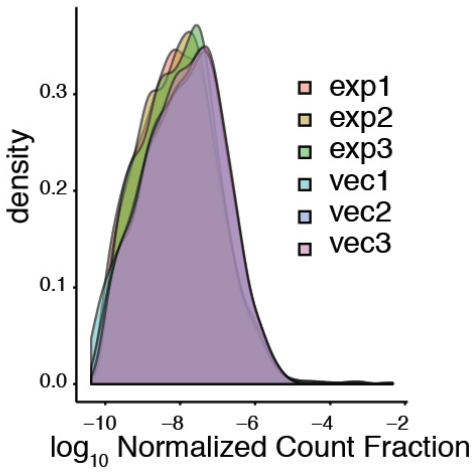


Figure S4: MPRA Quality Control, related to Figure 4

(**A**) Biological replicate correlations of depth-normalized experimental MPRA libraries (from GFP RNA).
(**B**) Replicate correlations of depth-normalized vector control MPRA libraries. (**C**) Density plots of tag distributions for each replicate.

Extended Experimental Procedures

Mice

Linc-p21 (*Trp53cor1*) knockout mice were generated in collaboration with Regeneron Pharmaceuticals, crossed with B6.C-Tg(CMV-cre)1Cgn/J mice to remove the neomycin resistance cassette and further backcrossed into C57BL/6J, as previously described (Sauvageau et al. 2013; Goff et al. 2015). The resulting Trp53cor1$^{tm1.1Vlcg}$ strain was maintained by heterozygous breeding. Mutant mice were identified by genotyping for the loss of the *Linc-p21* allele and gain of the *lacZ* cassette (Transnetyx, Cordova, TN).

Heterozygous E13.5 through E18.5 embryos and adult mice (2-3 months old) were used to determine the tissue expression pattern of the *Linc-p21* locus via knocked-in *lacZ* reporter gene. E14.5 knockout and wildtype embryos were used to assess the effects of *Linc-p21* ablation on global gene expression (RNA-seq) in multiple tissues.

Beta-galactosidase Staining and Imaging

Expression of the *lacZ* reporter gene was assessed by histochemical detection of ß-galactosidase (X-gal staining) as previously described (Sauvageau et al. 2013; Goff et al. 2015). Embryos were harvested and immediately fixed in 4% paraformaldehyde (PFA) at 4^{o}C overnight. Tissue samples were collected prior to fixation for genotyping. Adult heterozygous mice were transcardially perfused with 4% PFA prior to tissue collection. Dissected organs and tissues were then fixed in 4% PFA at 4^{o}C overnight. After fixation, the embryos and dissected adult tissues were washed three times with wash buffer (2mM MgCl2, 0.01% Deoxycholic acid, 0.02% NP-40 dissolved in 1X PBS), incubated for 24 hours in X-gal staining solution (5mM potassium ferricyanide, 5mM potassium ferrocyanide, in wash buffer), post-fixed in 4% PFA for 2 hours, and then stored at 4^{o}C in 70% ethanol (Sauvageau et al. 2013).

In the case of the E18.5 embryos, the skin was removed prior to X-gal staining to allow for penetration of the staining solution. Organs and tissues dissected from these embryos were stained for 12 hours to ensure optimal staining. Whole embryos were imaged at 3.5x magnification and E14.5 dissected organs at 70x for muscle, 15x for brain, 20x for heart, 20x for lung, 15x for liver and for E18.5 at 20x for muscle, 9.0x for brain 10x for heart, 10x for lung and 9.0x for liver and 3.0X for adult organs (scale bar=2mm) using an AxioZoom.V16 microscope equipped with an AxioCam HRm camera. Images were processed with the Zen Pro (Zeiss) Imaging Software.

RNA-Seq analysis

Reads from fastq samples were aligned to the mouse genome (mm10) using Tophat2 with non-standard options "--no-coverage-search --max-multihits 10 -p 8" (Kim et al. 2013). Each sample was quantified using Cuffquant with nonstandard options "-p 8 --no-update-check", and differential analysis was performed for each wildtype-vs-knockout tissue comparison using Cuffdiff2 with nonstandard option "-p 8". We also performed a Cuffdiff2 analysis in which all wildtype-vs-knockout samples were assessed together using nonstandard option "-p 8" (Trapnell et al. 2012). All analysis scripts are available as Files S2 and S4, all code is available on Github, and we frequently used Cummerbund for analysis and to generate figures (Goff et al 2013, https://github.com/rinnlab/lincp21).

Local region analysis

Genes within a 4 megabase window surrounding the *Linc-p21* locus were assessed for significant differential expression in each tissue. Significant differential expression for the entire region was assessed by comparing the number of significantly differentially expressed genes in this region to the number of differentially expressed genes in 10,000 size-matched randomly selected regions and assigning the corresponding bootstrapped p-value. A p-value < 0.05 was used as the significance threshold. The reproducibility of the differential expression of particular genes was assessed by

generating a summary image across all tissues. Each point is the average fold-change of a gene within the 4Mb region surrounding *Linc-p21*, and the error bars represent the standard error of the fold-change for this gene across all tissues.

Expression and Fold-Change Correlation analysis

We assessed the correlation of gene expression values of *Cdkn1a* to *Linc-p21* as well as the expression response (1- KO/WT) of *Cdkn1a* to *Linc-p21* transcript level. Each correlation was assessed by a linear regression (in R: lm(geneResponse ~ *Linc-p21*_wildtype_expression)).

Hi-C analysis

Mouse embryonic stem cell Hi-C data (Dixon et al., 2012) was downloaded from GEO (GSE35156), analyzed as previously described (Barutcu et al., 2015), and binned at 10kb resolution. The biological replicates were pooled for downstream analysis. Interaction frequencies were compared between the 10kb bins containing the *Linc-p21* gene locus and (1) the bins containing genes similarly affected by *Linc-p21* deletion across multiple tissues or significantly differentially expressed in at least one tissue and (2) all the 10kb bins within the same region (+/- 2Mb of the *Linc*-p21 gene locus).

Cloning and mutagenesis

The *Linc-p21* locus was cloned using a BAC plasmid (RP24-248L4) obtained from Children's Hospital Oakland Research Institute. Acc651 and Xho1 restriction sites were added to amplification primers to enable ligation into the multiple cloning site of the pGL4.23 vector (Promega) containing the luciferase gene. The p53 binding site was perturbed using inverse PCR and 5' phosphorylated primers containing mutations amplified off of the Exon 1 clone in pGL4.23 similar to the Quick Change protocol (Agilent Technologies). The pGL4.73 vector was used in co-transfection as a transformation control expressing Renilla luciferase.

Transfection

For reporter experiments with luciferase, 100ng of pGL4.23 and pGL4.73 were transfected in triplicate using 0.2ul of TransfeX reagent (ATCC ACS-4005) in OptiMEM (Life Technologies 31985-062) with no antibiotics. 100µl of cells were plated into black 96 well plates with optically clear bottoms at a concentration of $1x10^4$ cells/ml. After 24 hours, transfection media was removed and full growth media containing antibiotics was added with 10% FBS. Cells were allowed to grow for an additional 60 hours with additions of fresh media at 24 and 48 hours.

Luciferase Luminescence

At 60 hours, the Dual Glow Luciferase Assay System (Promega E2920) was used to measure luciferase and renilla luminescence on a 96-well luminometer. Growth media was removed and 20µl of Luciferase Reagent was added. After incubating for 10 minutes at room temperature, plates were read using pre-defined settings to quantify firefly luminescence. 20µl of Stop and Glow reagent was added and incubated for 10 minutes. Renilla luminescence was read similar to above. Analysis was performed by calculating the ratio of firefly to Renilla luminescence. All sections were normalized to ratio of Exon 1 luminescence, which was set to 100%.

MPRA Analysis

We counted tags originating from reads containing GFP sequence and with a perfect match to the barcodes we designed, and then normalized each sample to the total number of counts from that sample. We calculated the median ratio of RNA library signal to vector library signal for each base-pair and calculated rolling signal means across the locus with different windows and slides (File S5). For all analysis reported in this paper, we used a window of 500bp and a slide of 50bp. To generate p-values for significance of any given region, we permuted the signal ratio values across the entire locus 1000 times and repeated the sliding window analysis for each permutation, generating a randomized permutation p-value for each 500bp window.

Accession numbers

All newly sequenced data has been deposited in GEO with accession number GSE73472. Data on embryonic and adult brains was processed from publically available data set GSE61716. Hi-C data was processed from data set GSE35156.

Supplemental References

Barutcu, A.R., Lajoie, B.R., McCord, R.P., Tye, C.E., Hong, D., Messier, T.L., Browne, G., van Wijnen, A.J., Lian, J.B., Stein, J.L., et al. (2015). Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. Genome Biol. *16*, 214.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Goff L, Trapnell C and Kelley D (2013). *cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data.*. R package version 2.14.0.

Goff, L.A. et al., 2015. Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences*, 112(22), pp.6855–6862.

Kim, D. et al., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4), p.R36.

Sauvageau, M. et al., 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. D. Reinberg, ed. *eLife*, 2, p.e01749.

Trapnell, C. et al., 2012. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, 31(1), pp.46–53.

Supplemental Files

- **Supplemental File 1. RNA-Seq master sheet --** description of each sequencing library and mouse used in this paper.
- **Supplemental File 2. Report Rmd --** R markdown code used to generate the individual tissue reports (available on github).
- **Supplemental File 3. Tissue sequencing reports --** summary analysis files of each wildtype-vs-knockout tissue comparison.
- **Supplemental File 4. Summary and MPRA analysis Rmd --** R code used to generate any of the "summary analysis" or MPRA figures (available on Github).