Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

As Published
http://dx.doi.org/10.1073/pnas.1411263112

Publisher
National Academy of Sciences (U.S.)

Version
Final published version

Accessed
Mon Apr 01 01:19:12 EDT 2019

Citable Link
http://hdl.handle.net/1721.1/100788

Terms of Use
Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Detailed Terms

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain

Loyal A. Goff\textsuperscript{a,b,c,1,2,3}, Abigail F. Groff\textsuperscript{a,c,d,1}, Martin Sauvageau\textsuperscript{a,c,e,1}, Zachary Trayes-Gibson\textsuperscript{3}, Diana B. Sanchez-Gomez\textsuperscript{2}, Michael Morse\textsuperscript{a}, Ryan D. Martin\textsuperscript{a}, Lara E. Elcavage\textsuperscript{a}, Stephen C. Liapis\textsuperscript{a,2}, Meryem Gonzalez-Celeiro\textsuperscript{a,c,g}, Olivia Plana\textsuperscript{a}, Eric Li\textsuperscript{a}, Chiara Gerhardinger\textsuperscript{a,c}, Giulio S. Tomassy\textsuperscript{a,3}, Paola Arlotta\textsuperscript{a,c,3}, and John L. Rinn\textsuperscript{a,c,e,f,3}

Departments of \textsuperscript{a}Stem Cell and Regenerative Biology and \textsuperscript{b}Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138; \textsuperscript{c}Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; \textsuperscript{d}The Broad Institute of MIT and Harvard, Cambridge, MA 02142; \textsuperscript{e}Department of Systems Biology, Harvard Medical School, Boston, MA 02115; \textsuperscript{f}Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA 02215; and \textsuperscript{g}Institute of Molecular Health Sciences, Swiss Federal Institute of Technology in Zurich, 8093 Zurich, Switzerland

Edited by Donald W. Pfaff, The Rockefeller University, New York, NY, and approved April 16, 2015 (received for review July 31, 2014)

Long noncoding RNAs (lncRNAs) have been implicated in numerous cellular processes including brain development. However, the in vivo expression dynamics and molecular pathways regulated by these loci are not well understood. Here, we leveraged a cohort of 13 lncRNA-null mouse models to investigate the spatiotemporal expression of lncRNAs in the developing and adult brain and the transcriptome alterations resulting from the loss of these lncRNA loci. We show that several lncRNAs are differentially expressed both in time and space, with some presenting highly restricted expression in only selected brain regions. We further demonstrate altered regulation of genes for a large variety of cellular pathways and processes upon deletion of the lncRNA loci. Finally, we found that 4 of the 13 lncRNAs significantly affect the expression of several neighboring protein-coding genes in a cis-like manner. By providing insight into the endogenous expression patterns and the transcriptional perturbations caused by deletion of the lncRNA locus in the developing and postnatal mammalian brain, these data provide a resource to facilitate future examination of the specific functional relevance of these genes in neural development, brain function, and disease.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Epigenetic Changes in the Developing Brain: Effects on Behavior,” held March 28–29, 2014, at the National Academy of Sciences in Washington, DC. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/Epigenetic_changes.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence reported in this paper has been deposited in the Gene Expression Omnibus database (accession no. GSE67116).

1L.A.G., A.F.G., and M.S. contributed equally to this work.

*Present address: Department of Neuroscience and Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21205.
1To whom correspondence may be addressed. Email: loyalagoff@jhmi.edu, giulio_tomassy@harvard.edu, Paola_Arlotta@harvard.edu, or john_rinn@harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1411263112/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1411263112

PNAS | June 2, 2015 | vol. 112 | no. 22 | 6855–6862
stage when key brain developmental events are taking place; neurogenesis already has reached its peak in both the nascent diencephalon and the telencephalon (2, 18, 19). Conversely, by 7–8 weeks of age, all major postnatal developmental milestones have been achieved, including myelination, and the mouse brain can be considered mature (20, 21).

The incorporation of a lacZ reporter in the targeting constructs of each lncRNA locus KO mouse strain allowed us to determine the precise location of expression of the lncRNA genes using heterozygote mice. We collected whole brains at E14.5 and adult time points for each of the 13 lncRNA mutant strains (Dataset S1). LncRNA expression patterns were assessed by selecting coronal sections collected every 80 μm (E14.5) or 240 μm (adult) and by staining to detect β-galactosidase (β-gal) activity. A series of rostro-caudal images spanning the entire length of the brains from these strains were also collected and made available as Dataset S2.

Of the 13 lncRNA strains analyzed by these methods, 10 showed clear β-gal signal in the adult brain. Consistent with our previous RNA-seq analysis of WT brains and ES cell-derived neuronal stem cells (12, 17), little or no β-gal expression was observed for Mannr, Halr1, and Ttip33cor1 in the adult brain. Among those lncRNA loci with detectable β-gal signal in the adult brain, β-gal activity was also detected at E14.5 for the Lincenc1, Eldr, Pantr1, Pantr2 (12), and Peril loci. Several of these lncRNA genes demonstrated embryonic expression in regions known to give rise to the corresponding β-gal+ cell populations seen in the adult. At E14.5, Lincenc1 is expressed in the ventricular zone of the ventral telencephalon (VTel), but not of the dorsal telencephalon (DTel) (Fig. 1A). In the DTel, Eldr is expressed in the intermediate zone (IZ) and lateral cortical plate (CP) (Fig. 1A, Inset). At this developmental stage, the lateral CP contains postmitotic neurons destined for the upper layers of the neocortex (22).

The expression pattern of Lincenc1 at E14.5 is developmentally consistent with its expression in specific structures in the adult brain. Specifically, in the adult, Lincenc1 is expressed at high levels in the upper cortical layers II/III and IV (Fig. 1B), with highest number of labeled cells distributed in the primary somatosensory (S1) cortex. In this region, expression respects specific laminar and areal boundaries (Fig. 1A and B). For example, Lincenc1 is restricted to layers II/III in the S1 region (representing sensory input from the trunk; S1Tr, Fig. 1B), whereas its distribution is restricted to layer IV within the S1 barrel field (S1BF, Fig. 1B). Outside this region, Lincenc1 is expressed in the CA1, CA2, and CA3 fields of the hippocampus and in the dentate gyrus, both in the granular and molecular cell layers and the hilus (Fig. 1B). Together these findings describe a lncRNA locus with an intricate expression pattern that is established during development and is maintained in corresponding structures of the adult brain.

Eldr is another lncRNA demonstrating correlated expression at E14.5 and in the adult brain. In the E14.5 embryonic brain, Eldr is expressed in the germinial and mantle zones of the VTel (Fig. 1C). In the adult brain, Eldr+ cells are found scattered throughout the neocortex in a pattern resembling that of cortical interneurons, which are developmentally generated in the VTel (Fig. 1D) (23, 24). In addition, Eldr is expressed in a specific subgroup of cortical layer VI cells lining the subcortical white matter (arrowheads in Fig. 1D, Inset). Lineage-fate mapping and combinatorial marker analysis will be required in the future to support a direct lineage relationship between Eldr+ cells in the embryonic and the adult brain.

Like Eldr, Tug1 expression was observed in scattered cells within the neocortex of the adult brain (Fig. 1F). However, unlike Eldr, in the embryonic E14.5 brain, Tug1 is expressed, albeit at low levels, only in the developing choroid plexus with no expression detected in the VTel.

The lncRNA Peril shows distinct expression in both the developing and adult brain (Fig. 1G and H). At E14.5, Peril is expressed only in the VTel (Fig. 1G). In the adult brain, Peril is expressed in various structures of the telencephalon.
expressed primarily in the ependymal layer of the lateral ventricle, a region associated with adult neurogenesis (Fig. 1H) (25, 26). Expression of Peril was also detected in the olfactory bulb and dentate gyrus, albeit at lower intensity (Dataset S2). Interestingly, Peril is located ~110 kb downstream of the protein-coding gene Sox2, a known neuronal stem cell regulator which is also expressed in similar regions of the brain (27), suggesting the possible involvement of Peril in the regulation of Sox2 transcription. These results highlight a few specific examples of IncRNAs whose expression maps to specific brain regions through development and into adulthood and that may be poised to exert functional roles for the establishment and function of these structures.

**Regional Specificity of IncRNA Expression in the Adult Brain.** The majority of the IncRNAs in this study demonstrate some level of expression in the adult brain. Several of the IncRNAs analyzed showed restricted patterns of expression within multiple brain structures in the forebrain, midbrain, and hindbrain. Specific examples include the IncRNAs Pgs2os2, Crme, Kantr, and Celr (Fig. 2).

In the adult brain, Pgs2os2, an immune-response-activated IncRNA located proximal to the Prostaglandin-endoperoxide synthase 2 gene (Pgs2) (17, 28), is expressed in the neocortex, hippocampus, and thalamus (Fig. 2A and Dataset S2). In the cortex, its expression profile resembles the distribution of excitatory pyramidal neurons in layers II/III and IV (Fig. 2A, Inset) (2). In the hippocampus, Pgs2os2 is expressed in the CA1, CA2, and CA3 stratum pyramidale (Fig. 2A).

Similarly, multiple regions of the adult brain express Crme. This IncRNA, which shares a bidirectional promoter with the Iroquois homeobox gene lnx3, is expressed in distinct nuclei of the hypothalamus (Fig. 2B, Inset). It is also weakly expressed in the adult ependymal layer of the lateral and third ventricles and in the habenula (Dataset S2).

We observed a distinct expression pattern in the adult brain for Kantr, an IncRNA located ~20 kb downstream of the histone demethylase gene Kdm5c on the X chromosome. We find that Kantr is expressed in the adult neocortex, hippocampus, thalamus, and hypothalamus (Fig. 2C and Dataset S2). Within these regions, Kantr expression is spatially restricted to defined subregions. For example, in the hippocampus it is expressed largely in the CA1 stratum pyramidale (Fig. 2C, Inset), a region containing CA1 pyramidal neurons of different physiological and molecular identities (29–31).

Finally, we found that the IncRNA Celr, located proximal to the Insulin-induced gene 2 (Insig2), is specifically expressed in the adult substantia nigra, where it colocalizes with tyrosine hydroxylase-positive (TH+) dopaminergic neurons (Fig. 2D).

Collectively, these findings indicate that IncRNAs can have very precise and restricted spatial expression in key structures of the adult brain. This broad catalog of brain images of IncRNA gene expression provides a unique resource to facilitate future investigation.

**Gene-Expression Perturbations in Mutant Brains.** We next asked whether the deletion of these specific IncRNA genomic loci, while maintaining the act of transcription, contributed significantly to changes in gene expression. To address this question, whole brains from a minimum of three homozygous KO mice per strain were collected at both E14.5 and adult stages (age range 7.9–13.6 wk) (Dataset S3). Whole brains from WT E14.5 littermates and adults (age range 7.9–14 ± 2.0 wk) were also collected as control. To avoid potential sources of bias in sample collection and preparation, KO and WT brains were selected for each strain to maintain a balanced experimental design with respect to sex, litter (E14.5 only), and genetic background (Dataset S3). RNA was isolated from each sample as detailed in Materials and Methods and was used as input to generate Illumina TruSeq libraries.

Libraries were indexed and sequenced on the Illumina HiSeq 2500 platform across multiple flow cells, generating a total of 1.385 × 10^10 50-bp paired-end reads at a median depth of 1.18 × 10^6 aligned reads per sample (Dataset S4). Reads were mapped and quantified as described in Materials and Methods and SI Materials and Methods. For quantification, a modified version of the GENCODE (vM2) transcriptome (Dataset S5) was used as a reference. For quality control, a preliminary expression analysis was conducted on all samples using Cuffnorm.

Through a combination of expression analysis and visual inspection of the RNA-seq read alignments to the presumptive deleted regions, we confirmed the genotype and sex of each IncRNA KO and WT sample (Dataset S6). Following this “digital genotyping” via lacZ expression, the 109 remaining samples passing our quality control were subjected to differential analysis using Cuffdiff2 (v2.2.1).

We first conducted a global analysis across all strains. Differential transcriptional regulatory events were aggregated from all pairwise IncRNA KO versus WT comparisons across all strains at E14.5 or adult time points. For both time points, we observed exceptionally low variability across all conditions with a moderate increase in cross-replicate variability at E14.5 (Fig. S1). To identify IncRNA KO strains with similar differential expression profiles, hierarchical clustering of either E14.5 or adult samples was performed using the Jensen–Shannon distance across the normalized expression profiles for the universe of differentially expressed genes in all pairwise comparisons. We observed a greater total number of differentially expressed genes at E14.5 (n = 2,389) than in the adult comparisons (n = 669) (Fig. S2A). Similar results were observed for differential isoforms and protein-coding sequences (CDS) (Fig. S2A). In adults, most IncRNA mutant strains demonstrated some significant transcriptional regulatory events, but the total number of differentially expressed genes, isoforms, and CDS was lower than observed at E14.5 for individual strains (Fig. S2A). Approximately 10% of the significant genes (n = 214), isoforms (n = 124), and CDS (n = 192) identified at E14.5 were also differentially expressed in the adult comparisons (Fig. S2B).

Next we globally assessed the gene pathways perturbed upon deletion of IncRNA loci in vivo. Gene ontology (GO) enrichment analysis was performed for the superset of differentially expressed genes at both E14.5 and adult stages. We observed a significant enrichment of gene sets associated with neuronal differentiation and cell fate commitment at E14.5 alongside gene sets associated with cell-cycle regulation (Fig. S2C). This finding is congruous with the observed expression of several IncRNAs in various germinal zones of the developing brain (Fig. 1A, C, and G). In adult mice, the top significant GO gene sets included positive and negative regulators of cell differentiation and gene sets associated with PDGF and MAPK signaling pathways (Fig. S2D).

We further investigated the individual IncRNA mutant strains to determine which specific gene-expression pathways and programs are perturbed between WT and KO mice. We conducted differential expression analysis on individual KO versus WT comparisons for each strain at both E14.5 and adult time points. In these analyses, pathway and gene set enrichment studies were conducted using publicly available tools and annotated gene sets (detailed in SI Materials and Methods and Dataset S7).

One of the most abundant IncRNAs within this cohort is Kantr. Deletion of the Kantr locus results in a significant increase in gene sets involved in extracellular matrix organization, Ncam1 and L1cam interactions, NOTCH signaling, NGF and PDGF signaling, axon guidance, and neuronal development in the E14.5 brain (Fig. S3A and B). We also observed significant enrichment for gene sets associated with neurodegeneration (oxidative phosphorylation, Alzheimer’s, Huntington’s, and Parkinson’s disease) (Fig. S3B). In the adult mouse brain, deletion of the Kantr locus resulted in a significant increase in gene sets associated with axon guidance and PDGF signaling.
Peril is expressed in regions associated with neural stem and progenitor cells in both E14.5 and adult mouse brains (Fig. 1 G and H), deletion of the Peril locus (Fig. S2D) revealed specific effects on cell-cycle gene sets and key pathways involved in neural stem cell maintenance and differentiation. At E14.5, Peril−/− brains were enriched for genes associated with SHH and TGFβ/WNT signaling pathways and demonstrated a significant increase in the neural stem cell marker genes Notch1-3, Pax3, and Nestin (Fig. S2 E and F). Additionally, we detected a significant down-regulation of the VTel-restricted progenitor marker genes Dlx1, Dlx2, and Dlx5 (33, 34) relative to WT brains (Fig. S2E). Interestingly, and unlike E14.5 brains, adult brains of Peril−/− mice have a significant and specific reduction in the expression of the canonical neural progenitor marker Nestin and the cyclin-dependent kinase inhibitor Cdkn1a (p21) (Fig. S2G).

These select examples highlight the breadth of information available in this RNA-seq dataset. Detailed reports for each KO versus WT brain comparison at E14.5 and adult time points are provided for each individual IncRNA locus in Dataset S6.

**Cis Transcriptional Regulatory Dynamics of Specific lncRNA Loci.** Because many lncRNA genes have been implicated in cis-regulatory functions (35–38), we asked whether we could observe any significant changes in expression from the local genomic region neighboring the selected lncRNA loci. We first asked whether we could observe a bias in the transcriptional output of these loci as a result of replacing the endogenous lncRNA with our lacZ reporter. To determine whether the expression level of the knocked-in lacZ reporter was similar to that of the endogenous lncRNA, we compared expression levels for lncRNAs across all WT samples with the lacZ reporter gene from each KO strain. We found that lacZ mRNA expression is positively but not strongly correlated with endogenous lncRNA expression (r² = 0.4555; E14.5 = 0.47; adult = 0.29) (Fig. 3A). Although lacZ mRNA in several of the mutant strains such as Crnde−/−, Celrr−/−, and Pantr1−/− demonstrated expression levels consistent with that observed for the endogenous lncRNA (Fig. 3A), other strains exhibited more disparate expression levels. For example, lacZ was detected at higher levels than endogenous Lncenc1 and Peril when comparing KO with WT.

Next, we assessed the effect of each lncRNA locus on its local genomic environment by examining whether ablation of the lncRNA gene affected the expression of the closest protein-coding gene. For each strain and time point, we analyzed the KO vs. WT differential expression of nearest-neighbor protein-coding genes (Fig. 3B and Fig. S4). In 5 of the 13 mutant strains, deletion of the lncRNA genomic locus significantly affected the expression of the adjacent protein-coding gene at one or both time points (Fig. S5).
Pantr2 mice confirm deletion of each respective LncRNAs at the genomic locus are specifically and dynamically regulated in cis. We next asked whether this finding could represent a significant regional cis-regulatory effect as a result of the loss of the lncRNA locus. We detected five instances corresponding to four lncRNAs that demonstrated a significant (P < 0.05; bootstrap test) regional effect (Fig. 3D and Fig. S4). Interestingly, the majority of conditions, including those with a significant impact on the entire transcriptome, demonstrated no significant cis-regulatory effect. Collectively, these data point to variability among different lncRNAs regarding transcriptional regulation of neighboring loci. Further mechanistic work will be required to define cis regulation by each of the lncRNAs described here.

**Differential Regulation and Spatial Distribution of lncRNAs Derived from the Pou3f3 Genomic Locus.** To gain specific insights into the distinct transcriptional programs that establish spatiotemporal segregation of lncRNAs within the brain, we focused on two lncRNAs, Pantr1 and Pantr2, that flank both sides of the POU-domain gene family member Pou3f3 (Bml1), a key transcription factor involved in cortical development (39, 40). Pantr1 shares a bidirectional promoter with Pou3f3, whereas Pantr2 is located ∼13.2 kb downstream of Pou3f3 in the opposite orientation (Fig. 4.1).

We first compared the gene expression perturbations in Pantr1 and Pantr2 KO mice relative to WT mice. RNA-seq quantification indicated no significant effect on the expression level of Pou3f3 in either Pantr1−/− or Pantr2−/− E14.5 or adult mouse brains (Fig. 4B). However, we observed that loss of the Pantr1 genomic locus resulted in a significant, albeit modest, increase in the expression of Pou3f3 in the E14.5 brain (Fig. S6).

Conversely, deletion of the Pantr2 locus did not result in significant changes in expression of these paralogous regions. Generally, we observed that, relative to all other lncRNA mutant mouse strains, the loss of the Pantr1 locus resulted in the most dramatic effects on the transcriptome of the whole brain at E14.5 (Fig. S2A). Notably, ablation of the Pantr1 locus resulted in significant up-regulation of the neuronal progenitor markers.

![Figure 3. Cis effects of lncRNA deletion in the developing and adult brain. (A) Scatter plots of average lacZ expression (fragments per kilobase of exon per million fragments mapped, FPKM) in IncRNA KO versus endogenous lncRNA expression averaged across WT samples for all embryonic and adult samples. The line y = x is a reference for perfect correlation. (B) Summary of lncRNAs that significantly regulate their closest protein-coding neighbor. LncRNA mutant strains are shown in parenthesis. The x axis shows gene start distance (in kilobases) from the IncRNA transcriptional start site. The y axis shows log₂-fold change of expression levels between WT and IncRNA KO. Genes in quadrants III and IV are down-regulated in the KO. The lncRNA is shown in blue, and the closest protein-coding neighbor is shown in red. The orientation of each gene is indicated with arrows. (C) Example of a cis-region plot. The x axis shows gene start distance (in megabases) from the IncRNA transcriptional start site. Here we show a truncated region spanning only a 2-Mb window rather than 4 Mb around the IncRNA of interest, Tug1. The y axis shows the test statistic (Cuffdiff2). Red color indicates significant differential expression between WT and KO. (D) Table of significance: P values for every cis-region plot. The five highlighted conditions have regional effects with significantly differentially regulated neighboring genes.

![Figure 4. LncRNAs at the Pou3f3 genomic locus are specifically and dynamically regulated in the developing and adult brain. (A) Pou3f3 genomic locus and targeting strategy of the adjacent IncRNAs Pantr1 and Pantr2. The RNA-seq representative read density profiles for E14.5 and adult brains collected from Pantr1−/− and Pantr2−/− mice confirm deletion of each respective lncRNA compared with WT samples. (B) RNA-seq expression estimates (average from triplicates) in E14.5 and adult brains collected from WT, Pantr1−/−, and Pantr2−/− mutant mice for Pantr1, Pou3f3, and Pantr2. (C) β-Gal staining of IncRNA Pantr1−/− (Left) and Pantr2−/− (Right) adult brain coronal sections shows strong expression of Pantr1 in the CA1 region of the hippocampus (Pantr1, Lower Left), in both upper (III, IV) and deep (V, VI) layers of the neocortex, and in the caudate putamen, whereas Pantr2 is expressed most strongly in the medial habenula and at lower levels in the lateral habenula, layer V of the neocortex, and the CA1 of the hippocampus (Pantr2, Lower Right). Images are representative of staining performed on two animals. CA, cornu ammonis; CPU, caudate putamen; Hip, hippocampus; Hyp, hypothalamus; LHB, lateral habenular nucleus; MHB, medial habenular nucleus; NCx, neocortex; O, stratum oriens of the cornu ammonis; P, stratum pyramidale of the cornu ammonis; R, stratum radiatum of the cornu ammonis; Th, thalamus. [Scale bars: 500 μm (lower magnification whole brain panels), 200 μm (higher magnification middle and bottom panels).]
Ascl1, Nestin, Otx2, Notch1, Notch3, Ms1, Pax3, and Pax6, and significant down-regulation of mature neuronal cell-type markers including many canonical astrocyte markers such as Gfap, Sl100b, Aldh111, Fabp7, and Gap43 (Dataset S6). The observation that Pantr3 and neighboring gene loci are not significantly affected in adulthood suggests that this effect may not be cis-mediated. Interestingly, despite their genomic proximity to Pantr3, we observed strikingly different expression patterns for these two lncRNAs in the adult brain. Although Pantr1 is strongly expressed in layers II/III−IV and at lower layers in layers V−VI (Fig. 4C, Upper), Pantr2 cells are present only in low numbers in layer V (Fig. 4C, Upper). Similar differences are found in the hippocampus, where Pantr1 is strongly expressed in the pyramidal layer of CA1 and CA2, whereas Pantr2 is faintly present in a few scattered cells (Fig. 4C, Lower). Conversely, Pantr2 is very strongly expressed in the medial habenular nucleus, where Pantr1 is almost completely absent (Fig. 4C, Lower). These results, in conjunction with the previously demonstrated role for Pantr2 in development and organization of the neocortex (12), suggest that loss of the Pantr1 and Pantr2 loci may impact neuronal development programs via distinct cellular and context-specific mechanisms. Collectively, these examples demonstrate the utility of tracking lncRNAs during development and monitoring perturbations in gene expression upon their deletion to facilitate the selection and development of additional experiments to disentangle the molecular mechanisms of these loci.

Discussion

We have explored the spatiotemporal expression patterns of 13 lncRNA genes and analyzed gene-expression changes resulting from the deletion of each locus at the whole-brain level. The publicly available resource of spatial lncRNA expression patterns and differential RNA-sequencing data generated in this study provides an initial map to facilitate the understanding of lncRNA loci in the development and function of the mammalian brain.

Although many strategies for genetic loss-of-function studies exist, the whole-gene ablation method used here is often a first approach to determine the functionality of a locus irrespective of its molecular nature. Although each of these loci contains a lncRNA, it is important to consider that any observation resulting from this strategy could reflect the loss of any regulatory element in the deleted region. Thus, we do not ascribe the expression perturbations identified in this work to a functional lncRNA molecule but rather to the genetic loci containing the targeted lncRNA. Despite this limitation, one advantage of this approach is the incorporation of the lacZ reporter, which can be used to determine the precise expression patterns of the lncRNA within the tissue. Moreover, because the reporter maintains transcriptional activity at the locus, it also serves to control, on some level, for phenotypes resulting from the loss of transcription.

Many questions remain as to the global mechanistic roles for lncRNA transcripts in regulating gene activity. The rate of lncRNA gene discovery has significantly outpaced our ability to evaluate both the physiological significance and function of these genes. It is difficult to predict whether the loss of any particular lncRNA locus will present a phenotype, but crucial information on the spatiotemporal dynamics of expression from each locus can provide significant direction and focus to downstream mechanistic studies by highlighting those loci most likely to have a physiological impact. Genetic approaches such as the gene-ablation method used in this study, and more broadly in large efforts such as the Knockout Mouse Project (KOMP) (41) and the International Knockout Mouse Consortium (IKMC) (42), have proven utility in determining if a genomic locus has any biological role even if broadly defined as the sum of multiple regulatory elements.

More studies now have used whole-gene ablation to determine whether a lncRNA locus contributes to a particular phenotype or, as in the case of Visc−2, whether a lncRNA whose expression pattern in the developing brain is highly conserved across mammals results in no overt phenotype (43). The results from these whole-gene ablation studies inform the selection and focus of future studies to dissect a given locus further into its functional elements. It is important to note, however, that this approach cannot ascribe any observed biological activity to a functional lncRNA transcript, particularly in the absence of rescue experiments. Other in vivo loss-of-function approaches have also been applied to address the question of lncRNA functional roles (11, 44–48). However, it is important to stress that no single method exists that can account for all possible mechanisms of action of a noncoding loci (49). Within these limits, the phenotypes observed after ablation of specific lncRNA loci confirm that expression of this class of non-coding RNAs can serve as a proxy signal to identify functional genomic loci with physiological relevance to disease and development (12), independent of whether this activity is directly ascribed to a functional lncRNA molecule. For example, the perturbed pathways most commonly affected by knocking out the lncRNA loci analyzed here suggest that some of our mutant strains may have defects in neuronal differentiation during development.

The cellular heterogeneity of the mammalian brain can confound interpretation of RNA-seq studies of the whole brain for lncRNAs with discrete spatial expression. Examination of the whole brain allows the detection of very strong effects that result from the deletion of a genomic locus, including indirect effects in cell types that do not express the lncRNA in question. However, ablation of the lncRNA locus may have a direct impact on a relatively small population of cells, resulting in signal dilution over the entire brain. This limitation can be addressed only by transcriptional profiling of purified populations of cells. Despite the known cellular heterogeneity of the brain samples that we profiled, we observe large transcriptional impacts in several mutant strains (e.g., Kantr, Pantr1, and Peril), suggesting either a broad effect across multiple cell types or strong differential expression within a cell population. Although this lack of resolution can make direct interpretation of gene-expression perturbations difficult, it can be useful for identifying broad trends in differential transcriptional regulation and can begin to provide a picture of the impact of specific lncRNA loci with the potential to serve as a proxy signal for the function of the lncRNA within the brain. In the future, sequencing of specific subpopulations of cells from these mutant strains should identify additional affected pathways. The availability of expression profiles for lncRNAs, as provided here, will aid in the selection of specific populations of cells for purification.

Our differential RNA-seq analysis provides biological contexts and interpretations that are, in most cases, borne out by the selective and specific positive β-gal stains. Thus, the combination of spatiotemporal in vivo expression and differential analysis can provide additive contextual clues and permit future hypothesis-driven analyses that can be used to determine whether a functional role exists in neural development and disorders for these lncRNAs.

The power of this combinatorial approach is apparent in several lncRNA mutant strains. For example, Peril RNA-seq and β-gal staining both suggest that this lncRNA is expressed in neural stem cell populations and may have a significant impact on their biology. Embryonically, Peril is expressed in regions containing neural progenitor cells. This distribution is maintained in Peril+/− adult mice, with β-gal+ staining observed in the ependymal lining of the ventricles (Fig. 2) and in the dentate gyrus of the hippocampus (Dataset S2), both regions that are associated with adult neurogenesis (50–52). Consistent with these expression patterns, RNA-seq analysis of the Peril+/− embryonic brain showed misregulation of cell-cycle genes critically important for proper maintenance and differentiation of neural progenitors (53). In adult whole brains, Peril−/− mice maintain a
significant misregulation of Nestin (Fig. S2G). Interestingly, Peril/T−/− whole brains demonstrate a significant reduction in Cdkn1a (Fig. S2G) relative to WT, and the specific loss of this protein has been shown to induce premature terminal differentiation and quiescence in adult neural stem cells (54).

The Pou3f3 locus lncRNAs Pantr1 and Pantr2 provide other notable examples in which the in vivo expression pattern combined with RNA-seq analysis provides intriguing insights. We previously have demonstrated a functional role for the Pantr1 lncRNA locus in contributing to development of the neocortex in mice (12). However, the mechanism by which this locus mediates this effect remains unclear. Modulation of Pantr1 has been shown to affect Pou3f3 expression levels in cancer cells (55). However, in the context of whole-brain expression analysis in E14.5 and adult mutant mice, we observe no significant effect on the expression of the neighboring Pou3f3 gene for either Pantr1+/− or Pantr2+/− (Fig. 4B). Surprisingly, loss of the Pantr1 locus resulted in a significant, albeit modest, increase in the expression of the other POU3-family paralogs located on different chromosomes. These genes may act redundantly to coordinate neurogenesis cellular identity and migration in the developing cortex (39). Thus, although speculative at this stage, the Pantr1 locus might act to regulate coordinately the cellular processes governed by the POU3-family genes during brain development.

Our lncRNA KO strains are comprised of deletions of varying distance from the target promoter. However, we found no obvious trend related to gene-targeting design that could justify the discrepancy in some strains between lacZ expression levels in the brain and WT lncRNA levels. It is also important to note that the knocked-in lacZ reporter is driven off the endogenous lncRNA promoter and thus, in theory, experiences the same transcriptional regulatory dynamics as the lncRNA.

Here we have identified several lncRNA loci whose ablation in vivo results in significant differential expression of local (± 2 Mb) or neighboring (immediately adjacent) protein-coding genes. Different mechanisms could be proposed for such effects. A first possibility is that these specific lncRNAs may act in cis. For example, Trp53cor1 is expressed in the brain, and Trp53cor1−/− whole brains do show significant up-regulation of several proximal genes (Inpp5s, Selrn, Rnf185, 8430429K09Rik), suggesting that this locus could exhibit cis-like regulatory effects.

A second possibility is that DNA regulatory elements are contained within lncRNA loci and that these elements are responsible for local gene regulation. A potential example for this modality is Tps3cor1 (linc-p21). We observed numerous differentially regulated protein-coding genes surrounding this locus in Tps3cor1−/− adult brains (Fig. S4). However, although this finding is consistent with a cis-regulatory effect for this locus (56), we did not detect expression of Tps3cor1 locus activity by β-gal staining or RNA-seq (Fig. 3A, Dataset S2, and Dataset S6). The absence of lncRNA expression combined with differential expression of surrounding genes points to a possible functional role for DNA elements within the locus. Of note, our KO strategy deletes a p53-binding motif a few base pairs upstream of the transcription start site of Tps3cor1. This observation suggests that this and potentially other DNA elements that reside within the Tps3cor1 locus may contribute to the local changes in expression levels.

Any investigation into the activity and impact of a lncRNA locus requires, as a foundational step, an examination of the in vivo expression patterns and resulting perturbations on gene expression in a loss-of-function context. In this study, we report the in vivo expression profiles for 13 lncRNAs during mammalian brain development and the consequent changes in gene-expression patterns after genomic ablation. However, the stable and reproducible production of an RNA transcript from a particular locus cannot be confounded with the assumption that this transcription results in a functional RNA molecule but rather serves as evidence of biological activity within the lncRNA locus. The precise molecular mechanisms then must be validated on a per-locus basis and require incorporating multiple genetic loss- and gain-of-function strategies. Here we provide a resource to facilitate further inquiry.

Materials and Methods

Mice were housed under controlled pathogen-free conditions in Harvard University’s Biological Research Infrastructure. All procedures were carried out in accordance with the guidelines of the National Institutes of Health’s Guide for the Care and Use of Laboratory Animals (57) and were approved by the Harvard University Committee on the Use of Animals in Research and Teaching.

Mouse Tissue Collection and Processing. LncRNA KO mice were generated in collaboration with Regeneron Pharmaceuticals by replacing the selected lncRNA gene with a lacZ reporter cassette as previously described (12). For β-gal expression, E14.5 whole brains were harvested from embryos fixed by immersion in 4% (vol/vol) paraformaldehyde (PFA) at 4 °C overnight. Adult brains were perfused with 4% (vol/vol) PFA and postfixed in 4% (vol/vol) PFA at 4 °C for 12 h. For RNA isolation, E14.5 whole brains were homogenized in Trizol (Life Technologies) (1 ml per brain). Adult brains were snap frozen in liquid nitrogen, and stored at −80 °C. Frozen brains were pulverized and homogenized in TRIzol (5 ml per brain).

β-Gal Staining and Immunostaining. Brain-wide expression of the lacZ reporter gene was assessed in all mutant strains by histochemical detection of β-gal (X-gal staining) and in biological replicates. Sequential sections, obtained at every 80 μm for E14.5 brains and at every 240 μm for adult brains, were imaged at 5× and 10× magnification using a Zeiss Axioscan.Z1, a Nikon 90i microscope. Immunohistochemistry for β-gal and immunostaining for the interneuron marker TH was performed using standard methods. Primary antibodies and dilutions were chicken anti-α-jigal (1:500; G4L-45; ICL) and rabbit anti-TH (1:5,000; A0152; Chemicon). Appropriate secondary antibodies were from the Molecular Probes Alexa series. Images were acquired and processed with the Zen (Zeiss) or Volocity analysis software v4.0.1 (Improvis).

mRNA-Seq Library Preparation and Sequencing. For all samples, RNA-seq was performed as previously described (TruSeq RNA Sample Preparation Kit v2; Illumina) (12). Libraries were sequenced using a paired-end, 50-bp read-length sequencing protocol (NWl Bauer Core, Harvard University FAS Center for System Biology). Sequencing reads were aligned to mm10 using TopHat2 (39) with the following alignment parameters: sensitivity = medium, multimatches 10–8. To standardize the analysis of each differential comparison, we created a report template that integrated components from several R/Bioconductor (59) packages as detailed in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank Venus Lai, David Frendewey, and David Valenzuela from Regeneron Pharmaceuticals for generating the lncRNA KO mice; Manolis Kellis (Computer Science and Artificial Intelligence Laboratory, MIT) for mentorship and support of L.A.G.; all members of the J.L.R. and P.A. laboratories for advice and input on the manuscript, and David Kelley for helpful comments; the Harvard Center for Biological Imaging for microscopy support; and the Bauer Sequencing Core within Harvard University’s Faculty of Arts and Sciences Division of Science for sequencing support. This study was principally supported by National Institutes of Health (NIIH)/National Institute of Mental Health Grant R01MH102416-0 (to J.L.R.) and in part by NINH/National Institute of Neurological Disorders and Stroke Grants R01NS062849 and R01MH101268 (to P.A.). L.A.G. is supported in part by National Science Foundation/Postdoctoral Research Fellowship in Biology DBI-0905973. A.F.G. is supported by National Science Foundation Graduate Research Fellowship DGE1146152. P.A. is a New York Stem Cell Foundation Robertson Investigator.


57. Committee on Care and Use of Laboratory Animals (1996) Guide for the Care and Use of Laboratory Animals (Natl Inst Health, Bethesda), DHHS Publ No (NIH) 85-23.