Title
The role of long noncoding RNAs in neurogenesis

Author
Ramos, Alexander Daniel

Publication Date
2014

Peer reviewed|Thesis/dissertation

The Role of Long Noncoding RNAs in Neurogenesis

by

Alexander Ramos

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

[redacted]

For my parents

Acknowledgements

Shortly after I first joined the lab four years ago, I told my mentor, Daniel Lim, that I wanted to do RNA-seq on the SVZ and characterize lncRNAs. It was a crazy proposal; no one in the lab had ever done RNA-seq, and the technology was still in the early stages with little analysis tools available. The fact that Dan supported this, enthusiastically gave me a tremendous amount of confidence, and is typical of his mentorship. He never stood in the way of a project or an idea, and left me free to explore while gently nudging in the right direction. He is also a great storyteller, and I was able to learn so much from writing manuscripts and grants with him. He will continue to be a role model for me long after this.

I was also fortunate to have scientific mentorship from my thesis committee: Robert Blelloch, Jun Song, and John Rubenstein. Their guidance helped shape this thesis and will continue to shape my thinking moving forward. When I first came to UCSF, I was constantly told about the collaborative environment, and the experiences I've had here far exceeded my expectations. Members of the Kriegstein, Oldham, Rowitch, and Alvarez-Buylla labs were a constant source of advice, borrowed reagents, and support.

I could not have done this without my friends. I would not have been here without the continued support of my family.

Chapter 3 of this thesis appears in published form as:

Ramos, A.D., Diaz, A., Nellore, A., Delgado, R.N., Park, K.-Y., Gonzales-Roybal, G., Oldham, M.C., Song, J.S., and Lim, D.A. (2013). Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. Cell Stem Cell *12*, 616–628.

Aaron Diaz performed analysis of Capture-Seq, and Abhi Nellore built the reference website. Ryan Delgado assisted with cell culture and quantification, KiYoub Park performed SVZ ChIP-seq, Gabriel Roybal performed ChIP-qPCR experiments, Michael Oldham performed

gene coexpression analysis, and Jun Song supervised bioinformatics analysis. Daniel Lim supervised all research and helped write the manuscript.

Chapter 4 is a manuscript prepared for submission with the following authors: Alexander D. Ramos, Rebecca E. Andersen, Siyuan John Liu, Tomasz Jan Nowakowski, Caitlyn Gertz, Ryan D. Salinas, Arnold R. Kriegstein, Daniel A. Lim.

Rebecca Andersen is listed as a co-first author, and played an equal role in experimental design and data analysis. John Liu performed RNA-seq analysis. Tom Nowakowski performed in utero electroportations and human tissue in situ hybridization, Caitlyn Gertz assisted with time-lapse imaging, Ryan Salinas helped to develop RIP assays, Arnold Kriegstein supervised research, and Daniel Lim supervised research and helped write the manuscript.

The Role of Long Noncoding RNAs in Neurogenesis

Alexander Ramos

Long noncoding RNAs (lncRNAs) have been described in cell lines and various whole tissues, but lncRNA analysis of development *in vivo* is limited. Here, we comprehensively analyze lncRNA expression of the adult mouse subventricular zone neural stem cell lineage. We utilize complementary genome-wide techniques including RNA-seq, RNA CaptureSeq, and ChIP-seq to associate specific lncRNAs with neural cell types, developmental processes, and human disease states. By integrating data from chromatin state maps, custom microarrays, and FACS purification of the subventricular zone lineage, we stringently identify lncRNAs with potential roles in adult neurogenesis. shRNA-mediated knockdown of two such lncRNAs, *Six3os* and *Dlx1as*, indicate roles for lncRNAs in the glial-neuronal lineage specification of multipotent adult stem cells. Using our lncRNA pipeline, we identify *Pinky* (*Pnky*) as a long non-coding RNA (lncRNA) that regulates the transition between NSCs and neurogenic progenitors. *Pnky* is expressed in NSC populations and becomes downregulated during neurogenesis. shRNA-mediated knockdown of *Pnky* expands the pool of neurogenic progenitors, increasing neurogenesis. *Pinky* is conserved and expressed in NSCs of the developing mouse and human cortex. In embryonic mouse brain, *Pnky* knockdown accelerates neurogenesis and depletes this embryonic NSC population. *Pnky* physically interacts with PTBP1, a regulator of mRNA splicing, and *Pnky* depletion leads to the differential splicing of PTBP1 targets. These data thus identify *Pinky* as a conserved lncRNA that regulates a critical stage of neurogenesis from NSCs in the adult and embryonic brain. Taken together, our data forms the foundation of a resource for the study of lncRNAs in development and disease. In depth study of the novel lncRNA *Pnky* highlights the utility of this resource for the identification of lncRNAs that play key roles in neural lineage progression.

Table of Contents

Table of Figures

Chapter 1: Introduction

Historical Perspective: RNA-based gene regulation and association with chromatin

Since the initial characterization of chromatin more than 50 years ago, investigators have appreciated the existence of various species of intimately associated RNA. Early work in the biochemical purification of histones led to the discovery of histone-bound chromosomal RNA or 'cRNA' (Holmes et al., 1972; Huang and Bonner, 1965). Further studies of cRNA revealed its ability to hybridize with DNA (Bekhor et al., 1969), demonstrated that *in vitro* reconstitution of chromatin requires the presence of RNA (Bekhor et al., 1969; Huang and Huang, 1969), and found distinct populations of cRNA in different tissues (Mayfield and Bonner, 1971). In the era before the discovery of eukaryotic transcription factors, these data led some to suggest that cRNA molecules were responsible for tissue-specific gene activation (Mayfield and Bonner, 1971).

Later studies of a distinct RNA population, chromatin-associated RNA (caRNA), or heterogeneous nuclear RNA (hnRNA) demonstrated that this RNA species was enriched in heterochromatin, and suggested it was a structural component (Paul and Duerksen, 1975). Chemical treatment to deplete ribonucleotide particles in the nucleus and subsequent electron microscopy revealed a disorganized and misshapen nuclear matrix (Fey et al., 1986), further demonstrating that RNA could play a key structural role in the nucleus. Subsequent studies using RNAse treatment on isolated nuclei or treatment of cells with inhibitors of RNA synthesis caused disorganization of chromatin and a collapse of the nuclear matrix (Nickerson et al., 1989).

These early studies of nuclear RNA pointed to a structural and gene-regulatory role for RNA in the nucleus. In an early genome-wide approach using purified nuclear poly-A RNA and

hybridization to cDNA libraries, Salditt-Georgieff and Darnell could show that a majority of nuclear RNA does not contribute the cytoplasmic RNA, leading them to conclude: “It appears that many long capped heterogeneous nuclear RNA molecules are of a different sequence category from those molecules that are successfully processed into mRNA.” (Salditt-Georgieff and Darnell, 1982) What were these RNAs? Were they responsible for gene expression regulation and/or chromatin structure? How?

The earliest specific examples of the intersection of noncoding RNAs and chromatin were those RNAs that contribute to dosage compensation, or the mechanism that keeps gene expression from sex chromosomes balanced between males and females (Bernstein and Allis, 2005). The first such characterized RNA is *Xist*, is a ~17 kb spliced and poly-adenylated transcript with six short conserved repeat elements, but otherwise low sequence conservation. In female cells, the X chromosome destined for inactivation/silencing begins to transcribe high levels of *Xist*, which eventually spreads *in cis* and coats the entire inactive chromosome (Plath et al., 2002).

Investigators next sought to characterize the state of chromatin at the inactive X chromosome. Covalent histone modifications are major determinants of the transcriptional state of a genomic locus. Methylation of histone 3 at lysine 4 (H3K4me3) is typically associated with actively transcribed genes and euchromatin; the major ‘writers’ of the H3K4me3 mark are the Trithorax group proteins (trx-G). In contrast, methylation of histone 3 at lysine 9 (H3K9me3) and lysine 27 (H3K7me3) are typically associated with transcriptionally silent genes and heterochromatin; the major ‘writers’ of these marks are the Polycomb group proteins (Pc-G) (Jenuwein and Allis, 2001). The inactive X chromosome is heterochromatic and marked with H3K9me3 and H3K27me3, and these marks appear subsequent to *Xist* spread. Study of *Xist* action revealed that a 1.3 kb conserved repeat domain (RepA) could directly bind and recruit EZH2, the catalytic domain of the Polycomb repressive complex 2 (PRC2), and writer of the

H3K27me3 mark (Zhao et al., 2008). This study suggested that a long noncoding RNA (lncRNA) could recruit a histone-modifying enzyme for proper silencing and dosage compensation of the chromosome. Importantly, the role of long RNA in dosage compensation is also seen at imprinted gene loci; imprinted genes are an example of dosage compensation in which one allele of a gene is silenced depending on the parent-of-origin. The lncRNA *Air* is paternally imprinted; this RNA can bind and recruit the H3K9 methyltransferase G9a to silence neighboring genes *in cis* (Nagano et al., 2008). Similarly the lncRNA *Kcnq1ot1* can recruit PRC2 and G9a for silencing of the imprinted *Kcnq1* locus (Pandey et al., 2008).

Early studies into the expression and function of lncRNA trasncripts clearly implicated these genes in the creation and maintenance of heterochromatin. The first identified and studied lncRNAs tended to be very long (>17 kb), and very highly expressed. They were usually studied in the context of dosage compensation, either from the sex chromosome or imprinted gene loci. It wasn't until the advent of high-throughput sequencing technologies almost 20 years after the initial discovery of *Xist* that the full range of structure and function of lncRNAs could be appreciated.

Modern Rediscovery of lncRNAs

With the advent of microarray technology and next generation sequencing (NGS), several groups were able to capture noncoding transcripts that were previously unknown. One of the earliest examples of a lncRNA found in the modern '-omics' era was uncovered using high-density tiling arrays of the *Hox* gene loci. The *Hox* genes are highly conserved loci consisting of 39 protein-coding genes that play key roles in body patterning and cellular positional identity. Adult fibroblasts isolated from different parts of the body maintain expression of *Hox* genes in culture corresponding to their position *in vivo*. Proper combinatorial expression of *Hox* genes is known to be dependent on H3K27me3-mediated repression and H3K4me3-mediated activation in the appropriate cell type. The use of high-density tiling arrays allowed for

the discovery of actively transcribed DNA between protein-coding *Hox* genes, including one named *H*ox *A*ntisense *I*ntergenic *R*NA (HOTAIR) (Rinn et al., 2007). HOTAIR is transcribed from a HoxC locus, binds PRC2, and directs the repression of the opposing *HoxD* locus. For the first time, a lncRNA was identified that could work *in trans*; that is, target loci on a distant chromosome. Later studies found HOTAIR to be overexpressed in various types of cancer, where it could target PRC2 to loci outside of *HoxD* and execute a metastatic transcriptional program (Gupta et al., 2010).

Although high-density tiling arrays allowed for the discovery of novel intergenic transcripts at the *Hox* loci, the technology was still dependent on having a particular genomic target region. ChIP-seq and RNA-seq allowed for the discovery and annotation of thousands of lncRNAs in mouse, human, and zebrafish. In one of the first genome-wide screens for intergenic noncoding RNAs, Guttman and colleagues used ChIP-seq maps of several diverse cell types to identify intergenic regions with the combination of H3K4me3 and H3K36me3 marks, a combination known to signify actively transcribed regions (Guttman et al., 2009). In combination with RNA-tiling arrays, they were able to identify 1600 lncRNAs that are transcribed in intergenic regions. Coexpression analysis with protein-coding genes placed these lincRNAs in a myriad of essential processes, including immune response, cell cycle, and alternative splicing. Subsequent RNA-seq studies, combined with ab initio transcript reconstruction, allowed for the identification of several thousand additional long noncoding RNAs across 24 tissues and cell types (Cabili et al., 2011; Guttman et al., 2010). A combination of chromatin maps and RNA-seq were used to annotate lncRNAs in zebrafish (Ulitsky et al., 2011), providing valuable information about the evolutionary conservation and putative conserved functions of lncRNAs.

Extensive sequencing and annotation efforts allowed for the elucidation of general principles of long noncoding RNAs (Cabili et al., 2011; Guttman et al., 2009; Ulitsky et al., 2011): 1) lncRNAs are arbitrarily defined as being > 200 bp long 2) They are often spliced, poly-adenylated, and capped 3) They show poor evolutionary conservation at the primary sequence

level, although many have short regions of high conservation. 4) They tend to have lower expression levels than mRNAs 5) They are extremely tissue- and cell type-specific, more than mRNAs.

LncRNAs and repressive chromatin

Initial studies of lncRNA action focused on their interaction with PRC2. Several genome-wide immunoprecipitation studies confirmed that hundreds of lncRNAs could interact with PRC2 components EZH2 and SUZ12 (Guttman et al., 2011; Khalil et al., 2009; Zhao et al., 2010). Careful analysis of HOTAIR binding partners revealed that the transcript bound EZH2 at its 5' end and histone demethylase LSD1 at its 3' end (Tsai et al., 2010). LSD1 removes the activating H3K4me3 mark, while EZH2 places the repressive H3K27me3 mark; the tethering together of these two factors therefore ensures the proper formation of repressed chromatin. Genome-wide studies demonstrated that there was significant overlap of gene promoters occupied by SUZ12 and LSD1, and localization of both is lost upon HOTAIR knockdown, suggesting the tethering activity of HOTAIR is required for proper localization of both of these modifiers at a subset of genes.

HOTAIR or the lncRNA Meg3 can also tether EZH2 to JARID2, a core PRC2 component. (Kaneko et al., 2014). The RNA-binding domain of JARID2 is required for the proper localization of EZH2 to chromatin, providing further evidence that scaffolding of proteins through lncRNAs is required for targeting of histone modifications. High-throughput RNA immunoprecipitation (RIP) assays of 12 chromatin-modifying enzymes revealed that lncRNAs can bind histone mark 'readers,' 'writers,' and 'erasers,' suggesting lncRNAs serve to scaffold several different histone-modifying enzymes with complimentary functions (Guttman et al., 2011).

The development of Chromatin isolation by RNA purification (ChIRP) and other similar methods allowed investigators to begin to answer uncover how a lncRNA can target histone-

modifying enzymes to chromatin (Chu et al., 2011; Simon et al., 2011). ChIRP-seq of HOTAIR revealed it binding to over 800 regions throughout the genome. HOTAIR binding sites are flanked by widespread EZH2 and H3K27me3 signal, suggesting HOTAIR can recruit EZH2, which subsequently spreads. Intriguingly, ChIRP-seq of *Ezh2* -/- cells demonstrated that HOTAIR remains localized to its genomic targets independently of EZH2, suggesting HOTAIR binding is the initial event required for EZH2 recruitment to certain loci. Importantly, while HOTAIR's binding patterns and binding to specific loci is reminiscent of a transcription factor, analysis of HOTAIR binding sites does not produce a sequence motif, suggesting that binding could be dependent on RNA secondary structure or a set of unidentified proteins.

Careful analysis of the *Xist* localization and silencing of the X chromosome using a technique similar to ChIRP provided further insight into the localization of lncRNAs (Engreitz et al., 2013). *Xist* localization was not sequence-specific; rather, *Xist* spreads to loci that are nearby in three-dimensional space. Polycomb recruitment and subsequent repression of these loci changes the three-dimensional conformation of the X chromosome, allowing *Xist* to continue its spread. Importantly, the continued spreading of *Xist* is dependent on the EZH2-binding RepA domain, demonstrating active repression is required for its spread. This suggests a model in which a lncRNA can establish a sub-nuclear domain of repressed chromatin. Consistent with this notion, the PRC2-associated lncRNA *Tug1* is responsible for establishing a repressive sub-nuclear domain in response to mitogen signaling (Yang et al., 2011). LncRNAs can therefore serve as a scaffold on which chromatin-modifying enzymes can assemble to established repressed chromatin domains in three-dimensional space.

LncRNAs and active chromatin

The first examples of lncRNAs a role in the establishment and maintenance of active chromatin also came from studies of homeotic gene loci. In *Drosophila,* both polycomb and

trithorax group members are recruited to trithorax response elements (TREs) or polycomb response elements (PREs). TREs upstream of the *ultrabithorax* (*Ubx)* homeotic gene encode lncRNAs in a cell-type specific manner (Sanchez-Elsner, 2006). These lncRNAs recruit the H3K4 histone methyltransferase Ash1 and to allow for the proper activation of *Ubx*. Importantly, this recruitment is abrogated by RNAse H treatment, suggesting the requirement of a DNA-RNA hybrid. Furthermore, ectopic expression of TRE RNA *in trans* is sufficient to recruit *Ubx*, suggesting the RNA species, and not the act of transcription, is required for recruitment.

Mammals do not have known TREs, however the experiments in *Drosophila* suggested to the intriguing possibility that lncRNAs could target activating histone-modification complexes to chromatin. *Mistral* is a lncRNA expressed from an intergenic gene region between homeotic genes *HoxA6* and *HoxA7* (Bertani et al., 2011). Like TRE RNAs, *Mistral* recruits a trithorax group member (MLL1) for the activation of nearby *HoxA* genes. This interaction is also RNAseH sensitive, and MLL1 was found to specifically bind a *Mistral* RNA/DNA complex. These data suggest an evolutionarily conserved mechanism by which lncRNAs can recruit trithorax group members to form active chromatin domains. Similar to *Mistral*, the lncRNA HOTTIP can bind and recruit trithorax group member WDR5 to the distal *HoxA* genes to allow for their activation (Wang et al., 2011). HOTTIP is expressed at very low levels, perhaps as little as one copy per cell. Unlike in the case of TRE-associated lncRNAs, overexpression of HOTTIP does not cause Trx recruitment and target activation, perhaps because ectopic HOTTIP cannot ‘find’ its proper genomic target.

Immunoprecipitation of WDR5 followed by RNA-sequencing has revealed that over 1000 mRNAs and lncRNAs can associate with this core trithorax group member (Yang et al., 2014). Overexpression of a mutant WDR5 lacking the RNA binding domain revealed that this domain is necessary for H3K4me3, target gene activation, and WDR5 protein stability. This suggests a model whereby lncRNAs are required for the recruitment and stability of WDR5 on chromatin,

which then facilitates the subsequent binding of RbBP5 and the rest of the Trithorax complex at gene targets.

In addition to recruiting trithorax group members, lncRNAs can also mediate nearby gene expression through the recruitment of mediator proteins and subsequent chromosomal looping (Lai et al., 2013). First described in human keratinocyte cell lines, enhancer-like lncRNAs cause the down regulation of specific genes within a 300 kb window (Ørom et al., 2010). Interestingly, these enhancer-like lncRNAs can potentiate the activation of a reporter when placed in an artificial construct. Overexpression of this class of lncRNA does not have an effect on gene expression, suggesting they function entirely in *cis*. Later studies demonstrated that activating lncRNAs bind mediator, and this interaction is required for chromosomal looping and target gene activation. Thus, a class of activating lncRNAs can affect three dimensional chromatin structure and intrachromosomal looping.

LncRNAs can also form and maintain localized sub-nuclear domains associated with active gene transcription (Batista and Chang, 2013). The lncRNA *Firre* coordinates the activation and spatial localization of five unique loci that reside on 5 different chromosomes (Hacisuleyman et al., 2014). The association of these loci is dependent on both Firre and HnRNP U, an RNA-binding protein also associated with the nuclear matrix. Similarly, the lncRNA MALAT1 associates with MLL1 and other chromatin-remodeling enzymes to establish a sub-nuclear domain of activated genes in response to mitogen signaling (Yang et al., 2011).

From the early characterization of XIST to the present, lncRNAs are perhaps best studied for their ability to associate with chromatin, regulate histone modifications, and affect the three-dimensional organization of the genome. The rest of this section will focus on other functions of lncRNAs, both in the nucleus and cytoplasm.

Nuclear lncRNAs

In addition to binding chromatin modifiers, there is increasing evidence that lncRNAs can bind and positively regulate transcription by both transcription factors and hormone receptors: Neural lineage specific lncRNAs RMST (Ng et al., 2013; 2012), Paupar (Vance et al., 2014), and Evf-2 (Feng et al., 2006) can bind and positively regulate the function of transcription factors Sox2, Pax6, and Dlx2, respectively, and the lncRNAs PRNCR1 and PCGEM1 (Yang et al., 2013) can cooperatively bind the androgen receptor and positively regulate gene activation in prostate cancer cells. In contrast there is also evidence that lncRNAs can serve to bind and antagonize transcriptional regulators and chromatin-modifiers: lncRNA PANDA binds transcription factor NF-YA and causes a failure to activate transcriptional targets (Hung et al., 2011); SchLAP1 binds chromatin-remodeling enzyme SNF5 and antagonizes its binding to targets in prostate cancer cell lines (Prensner et al., 2013); finally, lncRNA Gas5 serves contains a 'decoy' binding site for the glucocorticoid receptor and inhibits its binding to downstream targets (Kino et al., 2010).

Activation of a lncRNA can actively remove transcriptional repressors from DNA. *Jpx* is a lncRNA that is transcribed divergently ~10 kb upstream of the *Xist* locus. Before the onset of X-inactivation, the *XIST* promoter is bound by CTCF. *Jpx* is upregulated upon X-inactivation and binds CTCF. This binding allows Jpx to 'evict' CTCF from the *Xist* promoter, and X-inactivation proceeds (Sun et al., 2013b). Therefore, in addition to allowing for the active regulation of chromatin state, lncRNAs can also cooperate or antagonize a myriad of transcriptional regulators in the nucleus to control gene expression programs.

Alternative splicing represents a mechanism for generating extraordinary diversity from a limited set of gene loci. Differential use of splice isoforms can, for example, change the DNA-binding affinity of a transcription factor to allow for the maintenance of pluripotency in ESCs (Gabut et al., 2011). The lncRNA *Gomafu* has a conserved UACUAAC tandem repeat motif, which mimics the intron branch point sequence (Tsuiji et al., 2011). *Gomafu* can therefore bind

and sequester splicing factors away from pre-mRNA. *Gomafu* can bind the splicing factors SRSF1 and QKI, causing alternative splicing of at least two mRNA targets (Barry et al., 2013). Knockdown or overexpression of *Gomafu* causes opposite changes in exon selection of these targets, further suggesting that *Gomafu* expression levels alone can directly influence alternative splicing of specific transcripts.

The lncRNA MALAT1 localizes to nuclear paraspeckles, a subnuclear domain that are thought to be centers for mRNA splicing (Lamond and Spector, 2003). Indeed, MALAT1 binds SRSF1, an RNA splicing factor (Tripathi et al., 2010). Antisense oligo-mediated depletion of MALAT1 causes widespread alternative splicing of hundreds of exons. It can achieve this in part through the regulation of splicing factor recruitment to nuclear paraspeckles and pre-mRNA targets. Interestingly, three independently generated transgenic mice that lack MALAT1 do not have any changes in alternative splicing or SR protein localization (Eißmann et al., 2012; Nakagawa et al., 2012; Zhang et al., 2012a). These mouse studies suggest that lncRNA phenotypes may be highly dependent on cell cultures conditions and cellular context, and highlight the need to carefully design knockout strategies for lncRNAs. It is clear that a straight genetic knockout may not phenocopy the results of an acute depletion in culture using siRNA or similar methods.

Cytoplasmic lncRNAs

While most early studies of lncRNAs focused on their functions in the chromatin regulation in the nucleus, an expanding body of work implicates roles for lncRNAs in the cytoplasm. Many cytoplasmic lncRNAs can affect cellular function and gene expression by utilizing an array of different mechanisms to regulate mRNA stability, mRNA translation, and protein-protein interactions (Fatica and Bozzoni, 2014). Interestingly, several lncRNAs with nuclear function also function with separate mechanisms in the cytoplasm.

LincRNA-p21 was initially described as being a downstream repressor of gene targets in the p53 response (Huarte et al., 2010). The nuclear fraction of this transcript can bind and recruit transcriptional repressor and RNA-binding protein HnRNP K to hundreds of loci *in trans*. In the cytoplasm, LincRNA-p21 can associate with the polysome fraction, directly bind to at least two unique target mRNAs, and decrease their translation (Yoon et al., 2012). LncRNAs can also increase target translation; a lncRNA transcribed antisense to *Uchl1* can form a sense-antisense pair, and can enhance UCHL1 translation (Carrieri et al., 2012). Surprisingly, this ability is dependent on only an overlapping 5' sequence and an inverted SINEB2 element, and artificial transcripts can be engineered with this feature to enhance the translation of GFP.

Embedded retrotransposon repeats may play key roles in other mRNA-lncRNA pairs. ½-sbsRNAs are cytoplasmic lncRNAs that form imperfect duplexes with Alu elements embedded in a subset of target mRNAs (Gong and Maquat, 2011). This duplex is then bound by staufen1 and targeting for destruction via the nonsense-mediated-decay pathway. In this way, one lncRNA can regulate a number of target mRNAs.

The lncRNA TINCR binds and destabilizes target mRNA transcripts through a non-*Alu* binding motif called the 'TINCR-box.' (Kretz et al., 2013) Intriguingly, while TINCR and its targets bind staufen-1, targets are not destroyed via the nonsense-mediated decay (NMD) pathway, and instead TINCR-bound targets are stabilized by the interaction. These data suggest lncRNAs may confer a previously unappreciated function to cytoplasmic RNA-binding protein staufen1.

The STAT3 transcription factor is sequestered in the nucleus in a dephosphorylated state, and translocates to the nucleus upon phosphorylation. In dendritic cells, the lnc-DC1 noncoding transcript binds STAT3 and effectively prevents its dephosphorylation by phosphatase SHP1 (Wang et al., 2014). Therefore, in addition to posttranscriptional regulation of mRNAs, lncRNAs can block protein-protein interaction in the cytoplasm.

In the nucleus, lncRNAs serve as flexible scaffolds for the assembly of chromatin-modifying enzymes. There are now an increasing number of examples of lncRNAs serving as scaffolds in the cytoplasm for a variety of enzymatic activities. A high-throughput screening assay identified the lncRNA *NRON* as a cytoplasmic lncRNA that regulates the activation of NFAT (Willingham et al., 2005). Phosphorylated NFAT is cytoplasmic, and upon desphosphorylation it translocates into the nucleus to act as a transcription factor. Studies demonstrated that NRON exists in a cytoplasmic protein complex with NFAT, three of its kinases, and the scaffold protein IQGAP1 (Sharma et al., 2011). *NRON* knockdown causes enhanced NFAT dephosphoylation and downstream target activation. *HOTAIR* can also serve as a molecular scaffold in the cytoplasm; cytoplasmic *HOTAIR* interacts with E3 ubiquitin ligases and enhances the ubiquitination and subsequent degradation of target protein (Yoon et al., 2013). Interestingly, HOTAIR levels and target ubiquitination were increased dramatically in senescent cells, hinting that lncRNA subcellular localization and mechanism of action can be dynamically regulated by external cues.

A final emerging role for lncRNAs in the cytoplasm is serving as 'sponges' for microRNAs. MicroRNAs are short 21-25 bp small RNAs that bind target mRNAs in a sequence-specific manner and negatively regulate their expression (He and Hannon, 2004). Linc-RoR was originally discovered as a noncoding RNA that enhances reprograming of fibroblasts into induced pluripotent stem cells (iPSCs) (Loewer et al., 2010). Linc-RoR shares microRNA binding sites with core pluripotency transcription factors including Nanog, Sox2, and Oct4 (Wang et al., 2013b). High expression of this 'sponge' lncRNA allows for elevated expression of these factors and maintenance of the pluripotent state.

H19 is a cytoplasmic lncRNA expressed from an imprinted locus with key roles in growth and embryonic development (Gabory et al., 2010). A recent study demonstrated that *H19* can be processed and serve as a reservoir for miR-675, an anti-growth microRNA (Keniry et al., 2012). *H19* additionally harbors binding sites for for the let-7 family of microRNAs and can

function as a sponge, allowing for the upregulation of let-7 targets (Kallen et al., 2013). *Lnc-MD1*, discussed below in the section 'Tissue-and-lineage-specific lncRNAs', can also serve as both a reservoir and sponge for microRNAs, and regulated muscle cell differentiation (Cesana et al., 2011; Legnini et al., 2014). Therefore, lncRNAs can integrate into complex circuits with microRNAs to regulate gene expression and ultimately cellular fate.

Tissue-and lineage-specific lncRNAs

Most of the lncRNA molecular mechanisms discussed were studied in cultured cell lines. This section will focus on lncRNAs expressed in particular lineages, often with correlates *in vivo.* Genome-wide sequencing studies across 24 human tissues and cell types have revealed that lncRNAs are exquisitely tissue-specific, even more so than protein-coding genes (Cabili et al., 2011). High throughput sequencing of specific tissues and cellular lineages have been performed in blood (Alvarez-Dominguez et al., 2014; Hu et al., 2013), pancreas (Morán et al., 2012), adipocytes (Sun et al., 2013a), and adult and embryonic cerebral cortex (Aprea et al., 2013; Belgard et al., 2011). A high-throughput siRNA study of lncRNAs in ESCs revealed particular transcripts can repress the differentiation into each of the germ layers (Guttman et al., 2011). In addition, lncRNAs can be bound by both pluripotency and lineage-specific transcription factors (Ng et al., 2012). This suggests a model whereby lncRNAs can play key roles in both the fate specification and differentiation of cell types throughout the body.

Lnc-MD1 is a muscle-specific transcript that is induced upon myoblast differentiation into myocytes (Cesana et al., 2011). *Lnc-MD1* acts as a microRNA 'sponge' for miR-133 and miR-135, which allows for the upregulation of muscle-specific transcription factors. Interestingly, lnc-MD1 also serves as the host transcript for miR-133, and participates in a feed-forward regulatory loop with RNA-binding protein-HuR (Legnini et al., 2014). HuR is negatively regulated by miR-133, but binds and stabilizes lnc-MD1, thereby causing additional 'sponging,' lower miR-

133 biosynthesis, and an increase in HuR levels. Through miRNA sponging and regulation of HuR, a single lncRNA transcript can control the timing of differentiation.

Braveheart (Bvht) is a long noncoding transcript specifically expressed in embryonic stem cells and is required for differentiation into cardiac cells (Klattenhoff et al., 2013). *Bvht* regulates a core cardiac transcriptional network at least in part through its targeting of PRC2 to cardiac lineage-specific gene promoters. Additionally, *Bvht* is required for the maintenance of cardiac fate in primary cell cultures of cardiac fibroblasts.

Fendrr is a lncRNA expressed in in the lateral plate mesoderm during embryonic development, and in high levels in the mesoderm derivatives of the adult lung (Grote et al., 2013; Sauvageau et al., 2013). Inactivation of *Fendrr* through insertion of a premature Poly-A signal leads to embryonic lethality at E13.75 due to heart and body wall defects (Grote et al., 2013). Interestingly, a second knockout mouse was generated that entirely deletes the *Fendrr* gene; this mouse exhibited perinatal lethality likely due to a vascularization defect in the lungs (Sauvageau et al., 2013). *Fendrr* has the ability to bind both PRC2 and TrxG protein complexes and affect chromatin modifications at lineage-specific loci.

In the skin, two lncRNAs with opposing roles in progenitor differentiation have been identified. ANCR expression is strongly enriched in skin progenitor populations and becomes downregulated during differentiation (Kretz et al., 2012). Depletion of this lncRNA causes rapid gene expression changes consistent with differentiation as well as ectopic differentiation of progenitors in an organotypic cell culture system. In contrast, the lncRNA TINCR (also discussed above) becomes induced upon progenitor differentiation, and its depletion causes a loss of differentiated cell types (Kretz et al., 2013). TINCR carries out this function in part by binding staufen-1 and a several pro-differentiation mRNAs that are subsequently stabilized.

In the blood lineage, two lncRNAs have been described that are essential for proper differentiation and functioning of dendritic cells. LincRNA-Cox2 is induced by immunogenic stimuli and regulates the expression of key cytokines (Carpenter et al., 2013). This action

achieved at least in part through its interaction with HnRNP A/B and HnRNP A2/B1, RNA-binding proteins with known roles in transcriptional activation and splicing. Lnc-DC1 is specifically induced during monocyte differentiation into dendritic cells, and depletion of this transcript causes both a loss of proper differentiation and an inability to respond to immunogenic stimuli (Wang et al., 2014). Lnc-DC1 is a predominantly cytoplasmic lncRNA that functions through modulating the phosphorylation state and therefore nuclear translocation of STAT3. lncRNAs therefore play key roles in both the differentiation and function of a key immune cell type.

While there is ample evidence for lncRNA regulation of diverse lineages, I sought to further investigate the role of lncRNAs in neurogenesis. What follows is a brief introduction to the ventricular-subventricular zone (V-SVZ) adult neural stem cell system and lncRNAs that have been characterized in neurodevelopmental model systems *in vitro* and *in vivo*.

The V-SVZ as a model of neural development

The ventricular-subventricular zone (V-SVZ) of the adult mouse brain represents an ideal system for the study of lncRNAs *in vivo*. The neural stem cells (NSCs) of the V-SVZ are adult derivatives of radial glia, the stem cells of the embryonic brain responsible for cortical neuro- and glio- genesis. The V-SVZ represents a unique niche with several soluble and matrix-associated factors that allow for continued self-renewal and neurogenesis (Alvarez-Buylla and Lim, 2004). Stem cells of the V-SVZ express the glial markers GFAP and GLAST and the neural stem cell marker nestin. Activated stem cells express the epithelial growth factor receptor (EGFR), and retain GFAP expression. After activation, stem cells generate DLX2+, MASH1+, EGFR+ transit-amplifying cells. These undergo several symmetric divisions and progress through the lineage to migratory neuroblasts, which maintain Dlx2 expression and also stain for surface markers CD24, PSA-NCAM, and the cytoskeleton component Tuj1 (also called βIII

tubulin). These neuroblasts travel through the rostral migratory stream (RMS), and eventually terminally differentiate into interneurons of the olfactory bulb (OB). Stem cells acutely isolated from the V-SVZ can be grown in culture for several passages in the presence of EGF, FGF, and serum. Upon growth factor withdrawal, these cultures robustly produce Tuj1+ ,PSA-NCAM+ neuroblasts, and smaller numbers of GFAP+ astrocytes and Olig2+,O4+ oligodendrocyte precursors (OPCs) (Scheffler et al., 2005).

In contrast to the embryonic brain wherein multipotent precursor cells are inherently transient, continually changing their developmental potential and location over time and with organ morphogenesis, the adult V-SVZ retains its NSC population in a stable, spatially restricted niche, producing neurons and glia for the life of the animal (Kriegstein and Alvarez-Buylla, 2009). This enduring population of multipotent stem cells and its well-characterized daughter cell lineages make the V-SVZ a particularly tractable *in vivo* model for molecular-genetic studies of development. The V-SVZ has been used to elucidate key principles of neural development including the role of signaling molecules, transcription factors, microRNAs, and chromatin modifiers (Ihrie and Alvarez-Buylla, 2011).

LncRNAs and neurogenesis

Several genome-wide expression studies and siRNA screens in embryonic stem cells identified lncRNAs required for the maintenance of pluripotency and repression of the neural lineage. Genome-wide analysis of lncRNA expression during neuronal-glial fate specification revealed unique sets of lncRNAs that are differentially regulated during neuronal vs. oligodendroglial differentiation *in vitro*. Ng, et al. used custom lncRNA microarrays to identify 35 lncRNAs that were upregulated during neuronal differentiation from embryonic stem cells (ESCs) (Ng et al., 2012). Four of these candidates were chosen for shRNA-mediated knockdown, and all depletion of any of the four lead to a decrease in neurogenesis. Two of

these lncRNAs (lnc_N1 and lnc_N3) were largely nuclear, and bound key chromatin modifiers SUZ12 and REST. One of the identified lncRNAs, RMST, was found to localized to chromatin at the promoters of key neurogenic regulators, including *Sp8* and *Dlx2* (Ng et al., 2013). Intriguingly, RMST binds directly to SOX2 as well as RNA-binding protein HnRNPA2/B1. RMST co-occupies loci bound by SOX2, and SOX2 occupancy is lost upon RMST knockdown, suggesting that a long-noncoding RNA, perhaps through an hnRNP adapter, can modulate the recruitment of transcription factors to downstream targets.

Paupar is a lncRNA transcribed upstream from *Pax6*, a key homeobox transcription factor expressed in neural progenitor populations (Vance et al., 2014). *Paupar* associates with chromatin, and depletion of *Paupar* causes cell-cycle arrest and drives the neuronal differentiation of the N2A neuroblastoma cell line. Genome-wide studies indicate that *Paupar* and *Pax6* regulate a common set of transcriptional targets, and CHART-seq for *Paupar* demonstrates that it binds hundreds of gene promoters, including key neural differentiation genes. Interestingly, *Paupar* peaks are enriched for the DNA-binding motif of *Pax6*, and further studies demonstrate that *Paupar* directly binds *Pax6*, and the lncRNA and transcription factor co-occupy several gene promoters. Unlike the role of HOTAIR in targeting PRC2, knockdown of *Paupar* does not affect *Pax6* localization; rather, *Paupar* can enhance transcriptional activation or repression by *Pax6*. These data suggest that *Paupar* is recruited to loci genome-wide by *Pax6*, where it then acts to modify *Pax6* function through as of yet undefined mechanisms.

The developing retina is one of the first systems in which lncRNA function was studied *in vivo*. *Tug1* is expressed throughout the central nervous system and in retinal progenitors. siRNA-mediated depletion *in vivo* causes improper development of photoreceptor cells and cell death throughout the retina (Young et al., 2005). *Gomafu (also called RNCR2)* is expressed in retinal progenitor cells (Rapicavoli et al., 2010). Knockdown of *Gomafu* causes an increase in the production of amacrine cells and Muller glia from retinal progenitor cells, suggesting *Gomafu* can repress specific lineages from a multipotent progenitor *in vivo*.

High-throughput sequencing efforts and *in situ* hybridization analysis have identified a series of homeodomain associated opposite strand transcripts (HOSTs) (Rapicavoli and Blackshaw, 2009). These are putative lncRNAs that are transcribed divergently from known homeobox transcription factors, including those essential for neural development such as *Pax6*. These transcripts can show either reciprocal or coordinated expression with their protein-coding neighbor *in vivo*, suggesting different modes of regulation or molecular mechanism. One such lncRNA, *Six3os*, can modulate *Six3* function in the developing retina (Rapicavoli et al., 2011), reminiscent of the ability of *Paupar* to regulate *Pax6* action. Interestingly, *Paupar* is also transcribed divergently from its homeobox gene neighbor and interacting partner, further hinting that HOST transcripts may play a conserved role unique to these essential transcription factors.

DLX2 is a homeobox transcription factor that is required for interneuron development in both the OB (Long et al., 2007) and forebrain (Anderson et al., 1997), and its proper expression is critical for postnatal neurogenesis in the V-SVZ (Lim et al., 2009). *Dlx1* and *Dlx2* are oriented in an inverted configuration separated by a 8.3 kb intergenic region with several ultraconserved elements that are known binding sites for pro-neurogenic transcription factors (Poitras et al., 2007). The lncRNA *Dlx1as* is also transcribed from this intergenic region. The transcriptional start site overlaps an ultraconserved enhancer region, while its 3' end partially overlaps *Dlx1* on the opposite strand. This transcript has been described in embryonic development (Liu et al., 1997), where it demonstrates increased expression in more mature cell types vs. *Dlx1/2*. In adult neurogenesis, *in situ* hybridization data demonstrates that *Dlx1as* is preferentially expressed in the migratory neuroblasts of the RMS and throughout the OB of adult mice, while *Dlx1* and *Dlx2* are strongly expressed throughout the V-SVZ, RMS, and OB (Dinger et al., 2008).

A *Dlx1as*-null mouse was generated by inserting a poly-A cassette in the first intron of *Dlx1as* (Kraus et al., 2013). Importantly, this mouse does not produce full-length *Dlx1as*, however transcription can still initiate at the TSS. *Dlx1as* mutant mice are born at expected

ratios, survive to adulthood, and are grossly normal. *Dlx1as* null brains have a slight increase in *Dlx1* and *Mash1* expression as measured by ISH, however there is no obvious increase in interneuron generation. The mild phenotype found in the *Dlx1as* could be caused by several factors: 1) The premature-polyA strategy did not produce a strong enough depletion 2) The act of transcription at the *Dlx1as* promoter is sufficient for its partial function 3) The straight knockout could be compensated by redundant mechanisms, and finally, 4) The *Dlx1as* lncRNA is not required for proper neurogenesis or development. For future studies of lncRNAs, especially in complicated developmental systems, it may be required to use more sophisticated transgenic tools, such as inducible or cell-type specific deletions. In particular, with homeobox-adjacent lncRNAs, it may be necessary to study the lncRNA knockouts on a sensitized background, for example crossing the *Dlx1as* transgenic to a *Dlx1/2* heterozygous mutant.

Interestingly, the *Dlx* family of transcription factors lie on the same chromosomes as *Hox* gene family members, and it has been suggested that the *Dlx* genes and *Hox* genes may employ similar mechanisms for regulation (Ghanem et al., 2003). Similar to the *Dlx1/2* locus, the *Dlx5/6* locus encodes a lncRNA, *Evf2*, from an ultraconserved intergenic enhancer region. Initial studies in cell lines suggested that *Evf2* forms a complex with *Dlx2* and enhances activation of the *Dlx5/6* intergenic enhancer (Feng et al., 2006), consist with models proposed for *Paupar* and *Six3os*. An *Evf-2* knockout mouse was generated using a premature poly-A signal strategy (Bond et al., 2009). *Evf-2*-null mice had reduced numbers of interneurons in the embryonic hippocampus, however numbers returned to normal in the adult. Nevertheless, electrophysiological recordings of the adult hippocampus revealed loss of synaptic inhibition, demonstrating that an early lack of interneurons led to aberrant circuitry in the adult brain.

In contrast to what would be expected based on *in vitro* results, *in vivo* deletion of *Evf-2* led to an increase of *Dlx5* and *Dlx6* levels, similar to what was seen at the *Dlx1* locus with *Dlx1as* knockout. Consistent with *in vitro* findings, *Evf-2* null mice had a loss of DLX2 recruitment to the intergenic enhancer, however they also lacked the recruitment of MECP2,

methyl-CpG-binding protein and transcriptional repressor of the *Dlx5/6* locus. These data demonstrate that the lncRNA *Evf-2* can recruit both activators and repressors to the *Dlx5/6* locus, and further highlights the need for genetic knockout strategies to complement *in vitro* knockdown experiments in cell culture.

To further characterize the interaction between MECP2, DLX1/2, and *Evf-2*, Berghoff and colleagues conducted genetic epistasis experiments by crossing the *Evf-2*-null mouse with *Mecp2* -/- and *Dlx1/2*+/- mutant strains (Berghoff et al., 2013). MECP2 mice display a 2-fold increase in Dlx5 transcript levels. This increase in abrogated on a *Dlx1/2* +/- background, suggesting that DLX1/2 and MECP2 are antagonistic at the *Dlx5/6* locus. Analysis of methylation of the DLX5/6 enhancer revealed that *Evf-2* could inhibit the methylation of the DLX5/6 enhancer *in trans.* Taken together, these data suggested a model in which *Evf-2* can regulate the methylation state of the Dlx5/6 enhancer, which modulates the recruitment of the methyl-CpG repressor MECP2 and DLX1/2 transcriptional activators. Interestingly, a recent study identified hundreds of transcripts that can bind DNA methyltransferase DNMT1 and negatively regulate DNA methylation (Di Ruscio et al., 2013), suggesting a mechanism through which *Evf-2* may function at the *Dlx5/6* enhancer.

Sequencing and annotation of lncRNAs in the zebrafish genome lead to the identification of two brain-enriched lncRNAs, *Megamind* and *Cyrano* (Ulitsky et al., 2011). Depletion of *Cyrano* during development *in vivo* results in morphological defects in the head and eyes, and neural tube defects. Embryonic depletion of *Megamind* lead to small heads with hydrocephalus and loss of neuronal populations throughout the brain. Treatment with lower doses allowed survival to adulthood, however these fish exhibited locomotor defects (Lin et al., 2014). Remarkably, the developmental phenotypes caused by depletion of *megamind* or *Cyrano* could be rescued with the mouse or human homologues of these transcripts, suggesting lncRNAs play a conserved role in development of the nervous system despite their relative lack of primary sequence conservation.

Further study of the mammalian homogue of *Megamind,* called TUNA, revealed that this lncRNA is required for neuronal differentiation from ESCs (Lin et al., 2014). TUNA was found to bind RNA-binding proteins PTBP1, hnRNP K, as well as NCL. These three proteins exist in a complex that requires RNA to form, suggesting TUNA can serve as a scaffold for complex assembly. TUNA localizes to chromatin at the promoters of several key neurogenic genes, including SOX2. TUNA depletion causes a failure to recruit HNRNP K to the SOX2 promoter, and subsequent down-regulation of SOX2 and a loss of neurogenesis. TUNA therefore represents a second example of neurogenesis being controlled through the cooperation of a lncRNA and hnRNP proteins.

The developing cortex has also provided invaluable insight into the roles of lncRNAs and their molecular mechanisms *in vivo. Neurog1* is a neurogenic transcription factor expressed in the developing cortex. Like *Dlx1as* and *Evf-2*, *utNgn1* is a long noncoding RNA transcribed from a conserved enhancer element near *Neurog1* (Onoguchi et al., 2012). This transcript's expression is tightly correlated with *Ngn1* expression *in vivo*, and depletion of this lncRNA in acutely dissected neural progenitor cultures results in a failure to upregulate *Ngn1* upon differentiation.

Neurogenesis in the cortex follows a stereotyped series of transitions and cell divisions, similar to the V-SVZ lineage. Radial glial cells are ventricle-contacting stem cells that reside in the ventricular zone (VZ) that can either self renew or divide to give rise to TBR2+ transit-amplifying progenitors which differentiate and migrate through the SVZ and intermediate zone (SVZ/IZ) (Lui et al., 2011). These cells divide once or more to give rise to neurons that will then populate the six cortical layers. The layers are born in an inside out fashion, that is, the inner most layers are born at the onset of neurogenesis at E11.5, with the outer layers being born at the end of the neurogenic period 6 days later at E17.5 (Molyneaux et al., 2007). Aprea and colleagues used a transgenic fluorescent reporter system to separate and purify proliferating progenitors (PPs), differentiating progenitors (DPs), and newborn neurons (N) from the

embryonic cortex (Aprea et al., 2013). These cells were subject to whole-transcriptome sequencing, and lncRNAs from each of these cell types were identified, including 9 'switch genes,' defined as genes either enriched (on-switch) or depleted (off-switch) in DPs compared to other cell types. Among these 9 switch genes were *RMST* and *Gomafu* (*Miat*), which had previously been shown to play key roles in neural development in other experimental systems.

Gomafu overexpression or knockdown via *in utero* electroporation of developing cortex lead to an expansion of TBR2+ progenitors in the VZ, a decrease in the number of differentiating progenitors, and a decrease of neurons reaching the cortical plate (Aprea et al., 2013). This loss of neurons was chiefly caused by an increase in cell death of neurons in the IZ, suggesting *Gomafu* can regulate progenitor production and neuronal survival *in vivo*. Mechanistically, *Gomafu* can affect the splicing of *Wnt7b*, a known regulator of progenitor proliferation and differentiation in the embryonic cortex. Further studies are required to fully characterize the differential splicing induced by *Gomafu* manipulation, but differential splicing of *Wnt* raises the intriguing possibility that a lncRNA can control cell fate decisions through the splicing of the ligands and possibly receptors of key signaling pathways.

A recent study reported the generation and partial characterization of 18 lncRNA knockout mice (Sauvageau et al., 2013). In contrast to the *Fendrr* and *Evf-2* mice, these knockouts were generated by complete or near-complete deletion of the targeted lncRNA, and knock-in of a Lac-Z reporter. Of the 18, three lncRNA knockout mice, *Fendrr -/-, Mdgt -/-, and Peril -/-* demonstrated neonatal or perinatal lethality with variable penetrance. *Peril* is located 110 kb downstream of key neural transcription factor *Sox2* and is expressed in germinal zones of E14.5 and E18.5 brain and spinal cord. RNA-seq of *Peril* -/- brains vs. controls revealed a downregulation of pathways involved in cell cycle, energy processing, and protein translation processes, suggesting that *Peril* may be involved in the regulation of the cell cycle or metabolism of neurogenic progenitors.

Brn1 (*Pou3f3)* and *Brn2 (Pou3f2)* are neurogenic transcription factors critical for the proper differentiation and migration of cortical neurons during brain development (Sugitani et al., 2002). Two lncRNAs are transcribed near the *Brn1* locus: *linc-Brn1a* is bi-directionally transcribed from the *Brn1* promoter, while *linc-Brn1b* is ~10 kb downstream of *Brn1*. *Linc-Brn1b* is expressed in germinal zones of the developing brain starting at E13.5 (Sauvageau et al., 2013). By E18.5 expression is absent in the VZ and SVZ and is restricted to neurons of the cortical plate, and in adult mice expression is maintained in upper cortical layers of the somatosensory and visual cortex.

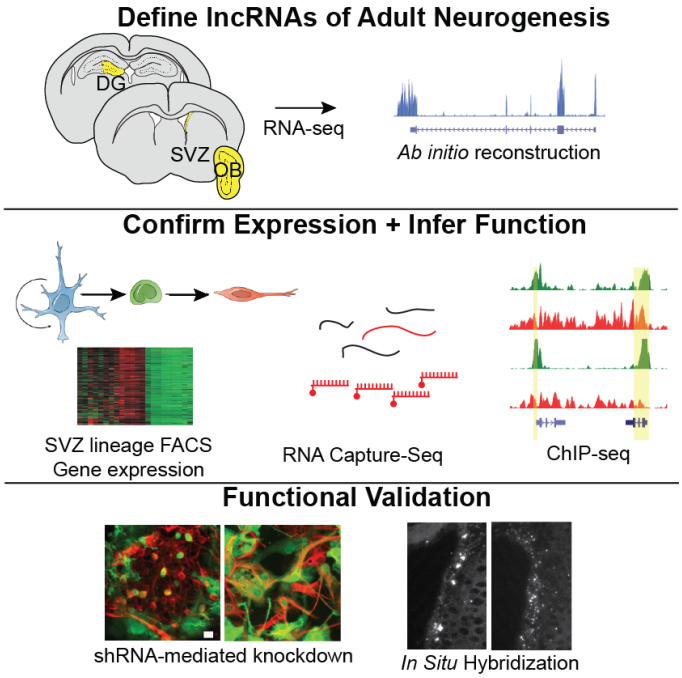
Linc-Brn1b-knockout mice exhibit a significant decrease in *Brn1* mRNA and protein expression. RNA-seq of null-cortices revealed downregulated genes were enriched in gene ontology terms associated with cellular proliferation and uprgulated genes were enriched for terms related to neuronal differentiation. Consistent with these results, there was a decrease in proliferating intermediate progenitors in the embryonic brain and a selective reduction of upper layer neurons. Interestingly, this was accompanied by an expansion of deep layer neurons, suggesting a mis-specification of cortical progenitor fate.

Introduction to current work

Most studies of lncRNAs in neurogenesis have focused on ESC-derived NSCs. While these cells are a good model system, results obtained from these cell culture studies are not always representative of neurogenesis and NSC biology *in vivo*. I have turned to the V-SVZ NSC system for its durable stem cell population, well-defined lineage, and the ability to derive primary stem cell cultures that faithfully recapitulate neurogenesis. In CHAPTER 3, I describe the use of genome-wide techniques to establish a lncRNA catalogue. Because lncRNAs are lowly expressed and cell-type specific, it was very likely that existing catalogues were incomplete. In order to study the function of lncRNAs, we first used a series of genome-wide methods to identify putative neurogenesis lncRNAS. These studies were done with both

cultured stem cell and stem cells isolated *ex vivo*, ensuring that the studied transcripts were likely to be have *in vivo* relevance. By adding additional filters derived from analysis of ChIP-seq peaks, we were able to narrow a list of thoudands of lncRNAs to a list of just 100 that were likely to play an important role in V-SVZ neurogenesis. We further functionally validated two of these candidates with shRNA experiments and *in situ* hybridization and demonstrated they could have differential effects on V-SVZ NSC differentiation. A summary of the approach for CHAPTER 3 is presented in **Figure 1.1**.

Figure 1.1



In CHAPTER 4, we chose one candidate lncRNA that we named *Pnky* for further study. Unlike most neural lncRNAs to date, knockdown of *Pnky* enhances neurogenesis from V-SVZ NSCs. We are able to show that this enhanced neurogenesis is a result of enhanced generation of transit-amplifying progenitors. Time-lapse imaging demonstrates these progenitors undergo and increased number of divisions and therefore generate more neurons. *Pnky* is also evolutionarily conserved and expressed in NSCs in the human and mouse cortex. Remarkably, *Pnky* knockdown in the embyronic cortex depletes the NSC pool and accelerates neurogenesis.

Finally, we show that *Pnky* interacts with the known splicing regulator PTBP1, and *Pnky* can influence alternative splicing and isoform selection. In sum, the work presented here begins by building a catalogue of novel lncRNAs that may play a role in neurogenesis; narrows this list down to high-confidence targets, and finally fully characterizes the role of one such lncRNA in both embryonic and adult neurogenesis.

Chapter 2: Materials and Methods

Brain Dissection for RNA extraction

The brain from adult (>P60) male C57/B6 mice was removed from the skull and placed in ice-cold L15 media and a 0.5 mm thick coronal slab was obtained. The lateral V-SVZ and striatum were then microdissected, avoiding contamination from the corpus callosum. FACS of V-SVZ cells was performed as described (Pastrana et al., 2009). DG was microdissected in ice-cold L15 media from 300 μm thick coronal sections obtained with a Vibratome. All Experiments were performed in accordance with protocols approved by Institutional Animal Care and Use Committee at UCSF.

High-Throughput sequencing

Illumina libraries were prepared using standard protocols. Sequencing was performed with a Genome Analyzer IIx or Hi-Seq instrument (Illumina). CaptureSeq followed by 454 sequencing (GS-FLX Titanium, Roche) was performed essentially as described (Mercer et al., 2012).

Sequencing Data Processing

Sequencing analysis was performed on the Galaxy platform (Goecks et al., 2010), using TopHat (Trapnell et al., 2009) and Bowtie (Langmead et al., 2009) for read alignment. **Supplementary file 2** describes all novel and previously published datasets used in this work. Cufflinks and Cuffdiff (Trapnell et al., 2010) were used for transcript reconstruction, quantification, and differential expression analysis. MACS was used to call Chip-Seq peaks (Zhang et al., 2008). For longer 454 reads, BLAT (Kent, 2002) was used to map the 454 reads to the transcriptome, and Newbler (Roche) was used to construct contigs.

lncRNA Catalog Assembly

.gtf files from Cufflinks were filtered using the UCSC genome browser to remove all single-exonic transcripts as well as any transcripts overlapping RefSeq genes, known protein-coding genes from other species using the 'Other Ref seq' track and the 'XenoRefGene' table. These filtered .gtf files were used as input for CuffCompare, and the complete Ensembl gene annotation .gtf file (Downloaded from: www.ensembl.org/info/data/ftp on 5/23/2011) was used as reference. We kept any transcript with the CuffCompare designation "I," which marks all novel intergenic sequences. To these sequences, we added annotated lincRNAs and processed_transcripts from Ensembl. Finally, we excluded all genes known to RefSeq, giving the 8992 transcripts in **Supplementary File 1**, a .bed file containing our putative lncRNAs. This file was merged with the Illumina iGenomes UCSC mm9 gtf annotation of all RefSeq genes (http://cufflinks.cbcb.umd.edu/igenomes.html on 1/5/2012), generating a final comprehensive annotation of non-coding and coding transcripts, which was then used for all subsequent analysis of RNA-seq data.

CaptureSeq library and Expression Arrays

Capture-Seq and Expression Array designs are available upon request. For the Capture-Seq probe set, probes were designed to tile across ~100 MB of putative noncoding loci and ~30 MB protein-coding control loci, for a total of 6287 genomic targets. For expression array probe selection, a FASTA file was generated representing all putative lncRNAs. A description of probe selection strategies for both sequence-capture and expression arrays can be found in *Technical Note: Roche Nimblegen Probe Design Fundamentals.*

In situ hybridization

In situs for Chapter 3 were performed using the QuantiGene ViewRNA ISH tissue Assay (Affymetrix). Probes were designed based on lncRNA sequences determined by Cufflinks. Images were taken on a DMI4000B fluorescent microscope (Leica).

For Chapter 4, Adult V-SVZ and cell culture staining, branched DNA in situ was performed on adult tissue according to manufacturers instructions using the RNAScope 2.0 high definition BROWN kit (ACD). In situ was performed on adult tissue according to manufacturers instructions using the RNAScope 2.0 high definition BROWN kit. For V-SVZ cultures, cells were grown on laminin and poly-d-lysine coated Labtek 8-well chamber slides and fixed with 4% PFA, and the RNAScope 2.0 high definition BROWN kit was used with the following modification: Pretreat solution 2 was added to cells and they were steam-treated for 10 mins. No protease was used.

For embryonic brain staining in Chapter 4, *in situs* were performed as described (Wallace and Raff, 1999). Briefly, DIG-labeled RNA probes were synthesized using the T7 high yield RNA synthesis kit (NEB) and a Digoxigenin RNA labeling mix (Roche), probes were hybridized overnight at 65 degrees. After washing, anti-digoxigenin Fab fragment (Roche) was added and incubated overnight at 4 degrees. Slides were washed and color reaction was carried out in 10% polyvinyl alcohol with NBT/BCIP (Roche) for 24-48 hours.

In situ probe sequences

For mouse *Pnky,* the full-length transcript was cloned using SMARTer RACE cDNA amplification kit with RNA isolated from V-SVZ NSC cultures and cloned into the PGEM-Teasy vector (Promega). For human *Pnky*, the following 3' sequence was cloned and used as an *in situ* probe:

TACGTTGACGTGGAGAGGATTTCAAACAACGCTAAAATGCTTTGAACTGACAAGGTGTCTT
GATATCTCCCTCACTCCATCCAGCACAGCTCCTCGAGATCACTCGCTAGGACAATGGCTGA
GCAGGCGATTCGTGCGGGCCTCGCCACCTCGGGGCGCGGACTGCGGGGTGTCCTAAGC
CCCTTCCGCAAGGACAGGATGGAGGCACCTGTAAGGAGATGCTGGCGCCACCCCAGCTT

CTCCCAGGTCCGGAGGAACCTCTACTCAGTCAATACCCTGAGCTGGACTTGTCTGAAGAAA
CGGAGCCGACTCCCTCTTGCCGGGGTGCGCTGAGTGGAGGGGAAACATCCTCGAATAAC
AGAACTACACCAAAAAGACACCCATGTTATCTCTCACACTTTCACACTCCTCGAGATAGTGA
GCCGGACCTGGGTCTTAGTAGCACCCAGTACCTTGACACAAACCTCCCAAATTTCCACCTG
AGTAACAGTTATGGGGTCAGTCCATGCACTGTAACTTGAACTCTAATTTATTAACTATTTCAT
CTAGTAAACACACTCACACCATATATAAAATAGCATTTATTTATTTCTATATACCAGGAGTTG
GCAGAAAACCCACCGTGACCACTCCCATACATTGAGCTGGAGGCACACAATTACTAAAACA
GAGGTGAGATGGTATTCATTTGATCTTAATTTTTTCTTATTTATGTAGTCCCAGGATAATAGA
AATCAGGAAACAAAAGAAAACAAAGAATTTTCTGAGGAGATGGCCATTGGGGGAGTGGAG
GTAGCAGCTGGTTTAAACCTAAGTAAAACTAGAAAAAGAAACTGCTGTTTCCTTTTTCTTATA
TCCACCTTAGAGGATCATGTTTGAACGTCCCTACTCCTCCTCCTCTTTTTAAAAAGCCTTGT
CTCAGTCATTCATTCCTGTGCTTCCTGCTCTTCTGCTAGACCCCAGCAGCTGTTTGATTTGG
TGAGGCCCCCCTCCAACCTCTGAGTGGAACTTCTTTTCTAAGGGCCTGCAGAATGTCAAAA
CTGAGGCTCTGGCTTCGGAGCTAGAGCTTTGAACAGCCAATCCACACAAAAAGGCAGCTG
GCTGCTTTAATGAAAACTGCTATAAAGCTTCAAGAACTTTAGCCTTGGGGGATGCATTTATA
AGGAACATGGAAAATGCATTTCCAAGTTGCTGGTTCTTGGGAGAGGCATAATAAACATTTA
CC

RNA extraction

All tissue samples were snap-frozen in liquid nitrogen at the time of dissection and stored at -80° C until RNA extraction. At the time of RNA extraction, frozen sections were resuspended in the Trizol reagent, and passed through a 21 G needle 10 times and a 28 G needle 3 times to ensure tissue homogenization. After ethanol precipitation, RNA was loaded on RNAeasy columns and subjected to on-column DNAse digestion followed by elution. RNA concentration and quality were assessed via Nanodrop spectrophotometer and an Agilent Bioanalyzer using an RNA nano chip. All RNA used for library preparation had an RIN > 8.8.

Cell Culture

For V-SVZ-NSC cultures, V-SVZ dissection was performed as described above. After dissection, V-SVZ tissue was rinsed with fresh PBS plus antibiotics and incubated in 0.25% trypsin (diluted in PBS at 1:5 to 1:3) at 37°C with occasional agitation for 15 min. Cells were plated in neurosphere medium (Neurobasal supplemented with N2, B27, glutamine, 20ng/ml EGF and 20ng/ml FGF) on a low-attachment plate and incubated in Every 3-4 days, fresh EGF

and FGF were added to the culture. V-SVZ-NSCs were dissociated with Accutase for passaging.

For monolayer culture and differentiation assay, V-SVZ was dissected from P5-P7 C57/B6 mice and dissociated as above, and resuspended in N5 media (DMEM/F12 with Glutamax, 5% Fetal Bovine Serum, N2 supplement, 35 ug/mL bovine pituitary extract, 20 ng/ml EGF, 20 ng/ml FGF, antibiotic/antimycotic). Cells were split 1:2 to passage 5 or 6 before switch to differentiation media (DMEM/F12 with Glutamax, 2% FBS, N2 supplement, 35 ug/mL bovine pituitary extract, antibiotic/antimycotic).

cDNA library preparation and Illumina Sequencing

cDNA libraries were constructed using Illumina's RNA-seq library kit per manufacturer's instructions. Briefly, RNA went through two rounds of poly-A selection using Oligo-dT beads. RNA was fragmented, and then purified by ethanol precipitation. First strand cDNA was synthesized using random primers. Second strand synthesis was performed, and the resulting double stranded cDNA was end-repaired with T4 DNA polymerase, Klenow polymerase, and T4 PNK. Klenow exo was used to adenylate the 3' ends, and Illumina PE adapters were ligated to the fragments. Libraries were then separated on a 2% agarose gel. For the single end runs, fragments of 200 bp +/- 25 bp were excised from the gel and purified. For the paired end runs, fragments of 300 bp +/- 25 bp were isolated. The purified cDNA was amplified by 15-18 cycles of PCR, and the resulting library was analyzed on an Agilent Bioanalyzer high sensitivity DNA chip before being sent for quantification and sequencing at the University of California Davis Genomics core on either a Genome Analyzer IIx or Hi-Seq instrument.

cDNA library preparation, sequence capture, and 454 sequencing

cDNA libraries were constructed using Roche's cDNA Rapid Library Preparation protocol for the GS FLX Titanium platform. Briefly, RNA was subjected to two rounds of poly-A selection using Oligo-dT beads (Thermo). RNA was fragmented at 70°C for 30 seconds in 100 mM Zinc

chloride, 100 mM Tris-HCL, pH 7.5. Fragmented RNA was purified by ethanol precipitation before random hexamer primed first- and second-strand cDNA synthesis. Fragments were end-repaired and 454 adapters ligated. The library was then purified using AMPure beads and resuspended in nuclease-free water.

Sequence capture was performed according to the Nimblgen SeqCap EZ Library User's guide, with the NimbleGen SeqCap EZ Hybridization and Wash Kit. Briefly, the cDNA library was amplified with 12 cycles of LM-PCR. The amplified library was purified on a QIAquick PCR purification column and eluted in nuclease-free water. Mouse COT DNA (Roche) (5 ug), 454 Rapid-HE1 (1 nmol) and 454 Rapid-HE2 (1nmol) oligos (Ordered from IDT) were added to the library, and the mixture was dried using a vacuum concentrator on high heat. The pellet was resuspended in 2X hybridization buffer and hybridization component A, and heated at 95°C for 10 minutes to denature the DNA. This mixture was added to an aliquot of our custom-designed SeqCap EZ library, and the hybridization reaction was carried out in a thermocycler set at 47°C (heated lid set at 57°C) for 72 hours. After hybridization, the library was thoroughly washed and purified using Strepadvidin Dynabeads (Invitrogen). The captured library was amplified with 15 cycles of LM-PCR, and the final library was purified on a QIAquick column. The final library was run on a 2.5% agarose gel to check size distribution and a Nanodrop spectrophotometer to determine concentration. The final library migrated at 500-700 bp. Sequencing was performed at the 454 Sequencing Center in Branford, CT on a GS-FLX Titanium instrument.

Chromatin Immunoprecipitation

Cultured V-SVZ-NSCs were dissociated and fixed in 1% paraformaldehyde for 10 min. Paraformaldehyde was quenched with 125 mM glycine for 5 min. After rinsing cells with PBS three times, the cell pellet was resuspended in swelling buffer (0.1 M Tris, pH7.6, 10 mM KOAc, 15 mM MgOAc) and incubated on ice for 20 min. Cell resuspension was dounced on ice 20 times and the nuclei pelleted via centrifugation at 2500xg at 4°C for 5min. The nuclear pellet

was resuspended in lysis buffer (50mM Tris, pH8, 10 mM EDTA, pH8, 1% SDS, 1X proteinase inhibitor cocktail (Roche Applied Science)) and incubated on ice for 10 min. The nuclear resuspension was sonicated with 25 cycles of: 30 sec on/30 sec off using the high setting of a Bioruptor (Diagenode Inc.). The sonicated chromatin was diluted at 1:10 in dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris pH8, 167 mM NaCl, 1X proteinase inhibitor cocktail) before immunoprecipitation. For input, sonicated chromatin equivalent of approximately 100,000 cells was spared. For IP with H3K27me3 antibody (Millipore, 07-449) and H3K4me3 antibody (Millipore, 17-614), the sonicated chromatin equivalent of 500,000 cells and 2,000,000 cells were used, respectively. After overnight incubation for IP, chromatin-protein-antibody complex was pulled down with Dynabeads protein G (Invitrogen) for 2 h at 4°C, followed by 5 washes with wash buffer (0.1 M Tris, pH9, 0.5 M LiCl, 1% Igepal, 1% deoxycholic acid, 1X proteainse inhibitor cocktail) and one wash with TE buffer (10 mM Tris, pH8, 1 mM EDTA). Chromatin-protein complex was eluted in elution buffer (1% SDS, 50 mM $NaHCO_3$) by vortexing for 15 min and then reverse-crosslinked in 0.2 M NaCl at 65°C overnight. RNA and protein was removed by treating with RNase (Roche Applied Science) at 37°C for 30 min and proteinase K (Roche Applied Science) at 55°C for 30 min. DNA was purified with phenol:chloroform:isoamyl alcohol (Invitrogen) in Phase Lock Gel Light tube (5PRIME), followed by a standard ethanol precipitation procedure with glycogen (Roche Applied Science).

Chip-seq Library Generation

Library was generated using reagents from NEBNext DNA Sample Prep Master Mix Set 1 (New England BioLabs) and following the manufacturer's instructions. Briefly, ChIPed DNA was end-repaired, the ends were adenylated, and then DNA was ligated to adaptor oligonucleotides. DNA was purified after each reaction using AMPureXP beads (Beckman Coulter Inc.). DNA was separated on a 2% agarose gel and DNA 250 to 550 bp was excised.

DNA was purified from the gel and PCR-amplified for 15 cycles. Absence of adapter contamination and the size range of the DNA sample was confirmed with an Agilent Bioanalyzer DNA chip. Libraries were run on an Illumina Genome Analyzer IIx at the Genomics Core at the University of California, Davis.

Oligonucleotides for Chip-Seq library generation

The sequences of adapter oligonucleotides are 5'-phosphorylation-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3' and 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATC*T-3' (*phosphorothioate bond). Before adding adapters to the library generation reaction, adapters were annealed at a concentration of 10 µm each by heating at 95°C for 5 min and cooling down slowly to ambient temperature (Quail et al., 2008). The sequences of PCR primers are 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' and 5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3'. All oligonucleotides were HPLC or PAGE purified (Integrated DNA Technologies).

Quantitative ChIP (qChIP)

Eluted DNA samples were added to 2X SYBR Green I Master Mix (Roche, Cat. No. 04887352001) and primer mix (0.5 µM final concentration). Primers used: Dlx1asTSS_F: GAGATGCAAAAAGCCTGCGG; Dlx1asTSS_R: CCAGCCCCATCTTCCTATGC. Dlx2 upstream: F- TTGGCTAAGGAAGGCCTAGA; R- CACCAGGGAGCGTTTCTAAT. The plate containing samples was run on a Roche LightCycler 480.

Antibodies used for qChIP

Antibody	Company	Catalog #	[µg/ml]
H3K4me3	Active Motif	39159	1µg/ml
H3K27me3	Active Motif	39155	1µg/ml
JMJD3	ABGENT	AP1022a	1µg/ml

Gal4 (control Ab)	Santa Cruz Biotech	sc-577	1µg/ml

RT-qPCR

RNA was reverse transcribed using the Transcriptor First Strand cDNA synthesis kit (Roche), and qPCR performed using SYBR Green I master mix (Roche) on a Light Cycler 480. *Dlx1as* expression was determined with delta delta Ct method, normalizing to housekeeping gene *Rplp0* and expression levels in proliferating cultures.

FACS purification of V-SVZ lineage

Prospective purification of the V-SVZ lineage was performed as described (Pastrana et al., 2009). Briefly, V-SVZ was dissected from 20-25 adult hGFAP-GFP mice (Jackson Laboratory) and tissue was digested with papain for 10 minutes at 37°C with agitation. The reaction was quenched with DMEM+10% serum, and digested tissue was spun down and cells resuspended in DMEM/F-12. Cells were passed through a 22% Percoll gradient, and resulting pellet was resuspended in labeling buffer (1% BSA, 0.1% Glucose in HBSS). EGF-A647 (Invitrogen) 1:300 and CD24-PE (BD) 1:100 were added and the labeling reaction was incubated for 15 minutes on ice. Labeled cells were washed 3x in labeling buffer and resuspended in DMEM/F-12 for FACS. FACS was performed on a FACS ARIA II YG at the Broad Center Cell Analysis and Sorting Facility. Samples were sorted directly into Trizol reagent, and RNA was extracted, amplified, labeled, and hybridized to arrays.

CaptureSeq and Expression Array Design

Up to 8 60mer unique probes were selected for 9462 of our lncRNA transcripts. 53 transcripts had fewer than 8 probes, 657 had duplicate probes, and no probes could be selected for 77 transcripts. 70,206 probes total were chosen and printed in duplicate for a total of 140,412 experimental probes per array. The array also included 159 protein-coding control probe sets (3 probes per gene). A description of probe selection strategies for both sequence-

capture and expression arrays can be found in *Technical Note: Roche Nimblegen Probe Design Fundamentals,* available at: http://www.nimblegen.com/products/lit/probe_design_2008_06_04.pdf. For protein-coding arrays, we used the Roche Nimblegen Mouse Gene Expression 12x135K Array.

Microarray Hybridization and Analysis

All RNA samples were amplified with a Whole Transcriptome Amplification Kit (Thermo), labeled, and hybridized according to manufacturers' protocol as published in the Nimblegen Arrays User's Guide, v. 6.0. Arrays were hybridized on a Maui Hybridization System and scanned on a GenePix 4000B scanner. Images were analyzed with DEVA v. 1.2.1. All array data was normalized using the Robust Multiarray Analysis (RMA) method (Irizarry et al., 2003), and differential expression was assessed using ArrayStar software v. 5.0.0. Samples were clustered and visualized with GeneE software v. 2.0.53. For statistical significance of differential expression of lncRNAs, student's t-test was used with a Benjamini-Hochberg corrected p value (See **Supplementary File 4**).

Alignment of Illumina RNA-seq Data

Alignments of RNA-seq data were performed using TopHat v. 1.2.0 (Trapnell et al., 2009) on the Galaxy platform (Goecks et al., 2010). For previously published data, FASTQ files were downloaded from GEO. For the paired-end reads from the V-SVZ, OB, and DG, reads were aligned by the Gladstone Institute Bioinformatics core using Tophat v. 1.2.0 and default parameters. **Supplementary File 2** summarizes all sequencing data generated for this paper. V-SVZ1, OB, and DG technical replicate lanes were run once on a Genome Analyzer II with 85 bp reads, and once on a Hi-Seq with 100 bp reads.

Transcript assembly, quantification, and filtering

Transcripts were assembled using Cufflinks v. 1.0.0 on the Galaxy platform with default parameters using paired-end reads generated from SVZ, OB, DG, NPC, and ESCs. The resulting .gtf files were filtered using the UCSC genome browser to remove all single-exonic transcripts as well as any transcripts overlapping RefSeq genes, and known protein-coding genes from other species using the 'Other Ref seq' track and the 'XenoRefGene' table. These filtered .gtf files were used as input for CuffCompare, and the complete Ensembl gene annotation .gtf file (Downloaded from: www.ensembl.org/info/data/ftp on 5/23/2011) was used as reference. We kept any transcript with the CuffCompare designation "I," which marks all novel intergenic sequences. To these sequences, we added annotated lincRNAs and processed_transcripts from Ensembl. The final result was 9554 transcripts. We performed a final filter to exclude all genes known to RefSeq, giving the 8992 transcripts reported in **Supplementary File 1**, a .bed file containing all of our putative lncRNAs, viewable on the UCSC genome browser. This file was merged with the Illumina iGenomes UCSC mm9 gtf annotation of all RefSeq genes (downloaded from http://cufflinks.cbcb.umd.edu/igenomes.html on 1/5/2012). This generated a final comprehensive annotation of non-coding and coding transcripts, which was then used for all subsequent analysis of RNA-seq data. For transcript quantification and differential expression, we used outputs from Cuffdiff v 1.2.1, with the following parameters: cuffdiff -N -b. This enables the bias correction algorithm, and normalizes FPKM values by the upper quartile of mapped fragments.

To generate heatmaps and clustering results in **3.2**, Cufflinks' isoform FPKM tracking output was filtered in the following way: only isoforms with an FPKM > 0, confidence_lo > 0, and status = OK across all samples were retained. This kept 9220 transcripts. FPKM values were transformed by log2(1+FPKM). Rows and columns were clustered with GeneE software, using a Pearson correlation metric and complete linkage.

Region Specificity Score Calculation

FPKM values for genes in SVZ, OB, DB, STR, PFC, POA, CTX, CP, SVZIZ, VZ, ES, and NPC were obtained by averaging replicate data. For each gene, tissue specificity score was then computed based on the Jensen-Shannon divergence between the relative abundance in these tissue types and the extreme distribution of being expressed in only one tissue type where the gene has the greatest expression value (Cabili et al., 2011).

Gene Ontology Analysis and Analysis of Protein-Coding Neighbors

Gene ontology analysis was performed using DAVID (Huang et al., 2009). To determine protein coding neighbors, we used GREAT (McLean et al., 2010) with our lncRNA's putative TSS as input.

Module Analysis

Gene co-expression modules were identified using a four-step approach. First, pairwise Pearson correlation coefficients (cor) were calculated for 29,334 transcripts across all samples. Second, transcripts were clustered using the *flashClust* (Langfelder and Horvath, 2008) implementation of a hierarchical clustering procedure with complete linkage and 1 – cor as a distance measure. The resulting dendrogram was cut at a static height of ~0.112, corresponding to the top 1% of pairwise correlations for the entire dataset. Third, all clusters consisting of at least 8 members were identified and summarized by their module eigengene (i.e. the first principal component obtained via singular value decomposition) (Horvath and Dong, 2008; Oldham et al., 2006). Fourth, highly similar modules were merged if their Pearson correlation coefficients exceeded an arbitrary threshold (0.85). This procedure was performed iteratively such that the pair of modules with the highest correlation > 0.85 was merged, followed by recalculation of all module eigengenes, followed by recalculation of all correlations, until no pairs of modules exceeded the threshold. Following these steps, 56 co-expression modules were identified. The strength of module membership (k_{ME}) for each transcript was calculated by

correlating its expression pattern across all samples with each module eigengene (Horvath and Dong, 2008; Oldham et al., 2008). These values are reported in **Supplementary File 3**. Module enrichment analysis with curated gene sets was performed using Fisher's exact test with gene symbols as a common identifier. Modules were defined as consisting of all transcripts that were positively correlated with the module eigengene at a significance threshold of P < 2.61e–08. This threshold corresponds to a Bonferroni-corrected P-value of .05 / (the number of transcripts X the number of modules). Enrichment analysis results for all gene sets and all modules are reported in **Supplementary File 3**.

Chip-seq alignment and peak calling

All Chip-seq reads were aligned using Bowtie v. 1.1.2 on the Galaxy platform with the default parameters. **Supplementary File 2** summarizes SVZ-NSC data and downloaded data used for Chip-seq analysis.

Chip-seq peaks were called using MACS v. 1.0.1 (Zhang et al., 2008) on the Galaxy platform, using chromatin input as the Chip-seq control file. Default parameters were used for all samples with the exception of NPC_K27 and MEF_K27. For these, the 'Build Shifting Model' option was turned off. Resultant output BED files of peaks were used to call chromatin modifications at gene promoters. Monovalent modifications were determined by finding any peaks that were within a 1 kb window of genes' transcriptional start sites, while excluding overlap with the opposing mark. Coding genes were defined as any gene having an accession number= NM_*, and lncRNAs were defined as described above.

454 Alignment and Comparison to Illumina

Roche's Newbler version 2.6 was used to construct a *de novo* assembly of the transcriptome from the 454 RNA-seq reads for the subventricular zone (SVZ). All command-line parameters were set to their default values. Bowtie version 0.12.8 was then used to align these reads to the transcriptome assembled by Newbler. All alignments were kept (using the -a

option). Because the current stable version of Bowtie cannot handle long 454 reads, BLAT (Kent, 2002) was used to map the 454 reads to the transcriptome. The resulting PSL file was converted to SAM with the Perl script blat_to_sam.pl included in the Trinity software package (Grabherr et al., 2011). FPKM expression values for 454 and Illumina were then computed using a modified version of RSEM (Li and Dewey, 2011) included in Trinity.

Clusters were obtained using our own implementation of KASP, a fast approximate spectral clustering algorithm (Yan et al., 2009). (Spectral clustering procedures are generically $O(n^3)$ in performance, where n is the number of data points. This makes finding exact spectral clusters computationally intractable for $n \sim 10^4$, as in the present case.) The steps we took to cluster the 26,279 points $\boldsymbol{x}_i$ from Figure 3 are as follows:

1. We first performed k-means clustering to separate the points into 262 clusters, associating each $\boldsymbol{x}_i$ with a cluster centroid $\boldsymbol{y}_j$. 200 iterations of the Hartigan-Wong algorithm were used.
2. We ran the spectral clustering algorithm described in (Ng et al., 2001) and implemented in the kernlab R package (Karatazoglou et al., 2011) to separate the 262 $\boldsymbol{y}_j$ into three clusters $\boldsymbol{z}_k$.
3. We assigned cluster memberships for the $\boldsymbol{x}_i$ by associating them with the spectral clusters $\boldsymbol{z}_k$ corresponding to their respective k-means centroids $\boldsymbol{y}_j$.
4. We isolated the spectral cluster for which both Illumina and 454 FPKM values tended to be high, and we merged the other two clusters. In Figure S4D, the merged clusters are in red, and the cluster corresponding to putative protein-coding transcripts is in blue.

Antibodies for Immunocytochemistry and immunohistochemistry (ICC/IHC)

Mouse anti-ASCL1 (BD Pharmigen), goat anti-SOX2 (Santa Cruz Biotechnology), chicken anti-NESTIN (Aves), mouse anti-NESTIN (Millipore), mouse anti-OLIG2 (Millipore), rabbit anti-GFAP (DakoCytomation), chicken anti-GFP (Abcam), mouse anti-Tuj1 (Covance),

mouse anti-PTBP1 (Invitrogen), guinea pig anti-DCX (Millipore), rabbit anti-SATB2 (Abcam), Guinea pig anti-DLX2 is described in (Kuwajima et al., 2006).

shRNA-mediated knockdown of lncRNA candidates

All shRNA sequences were designed using the Dharmacon siDESIGN tool. shRNA oligos were ordered from ELIM Biosystems, annealed, and ligated into the PSICO-R vector (Ventura et al., 2004), which carries a GFP marker. A construct designed against luciferase was used as a negative control. All oligo sequences are described below. After 24 hours infection, cells were passaged 2-3 times before plating on 96-well plates. Cultures were switched to differentiation media, fixed and stained at 7 days. All images were taken using an InCell Analyzer 2200 (GE). For cell counting, at least one field per well was counted. 5-6 wells for control and 2-3 wells for each knockdown construct were counted. At least 600 cells per condition were counted. Student's two-tailed t-test was used to assess statistical significance.

Six3os knockdown efficiency was determined using qPCR on FACS-isolated GFP+ cells. Efficiency was determined with the delta delta Ct method, normalizing to housekeeping gene *Rplp0* and LV-sh-luci-GFP infected cells.

For *Dlx1as* and *Dlx1* contransfection experiments, 293T cells were transfected with 1 ug of each plasmid using JETPRIME transfection reagent, following manufacturers instructions. RNA was extracted after 2 days and RT-qPCR was performed. Relative expression was measured using the delta delta CT method, normalizing to human beta actin and sh-luciferase construct. *Dlx1as* was cloned from RNA isolated from SVZ-NSC cultures using SMARTer RACE cDNA amplification kit, its sequence verified by Sanger sequencing, and then it was subcloned into the pCAG plasmid (Matsuda and Cepko, 2004) pCMV-Dlx1 was purchased from Origene.

For Dlx1as knockdown and qRT-PCR experiment, cells were infected with LV-constructs as described. After 2 days in differentiation media, GFP+ cells were isolated via FACS, and RNA was isolated and reverse transcribed.

RNA Immunoprecipitation Assay (RIP)

RIP assay was performed as described in (Rinn et al., 2007). ~10^7 V-SVZ cells were trypsinized and resuspended in 2 ml PBS, 6 ml water, 2 ml nuclear isolation buffer (1.28 M sucrose, 40 mM Tris pH 7.5, 20 mM MgCl2, 4% Triton X-100) and incubated on ice for 20 mins. Nuclei were then pelleted at 2500 x g for 20 mins at 4°C. Nuclear pellets were resuspended in 1 mL RIP buffer (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 5 mM MgCl2, 0.5% NP-40 0.5 mM DTT, PI tablet (Roche), and 100 u/ml RNAse OUT (Invitrogen)), and sheared with 20 strokes in a dounce homogenizer. Nuclear debris was pelleted by centrifugation at max speed for 10 mins at 4°C. Nuclear lysate was split into two fractions (mock and IP), and 8 ug antibody added. Antibodies used were anti-PTBP1 (Invitrogen) and anti-FLAG (Sigma). IP was performed rotating at 4°C overnight. The next day, 50 ul washed Dynabeads Protein G were added and incubated for an additional 2 hours. Beads were washed on magnetic rack 3x with RIP buffer, and resuspended in Trizol after final wash. Trizol extraction was carried out according to manufacturers' instructions, cDNA was made using the Transcriptor First Strand cDNA synthesis kit (Roche) with both oligo-dT and random hexamer primers. Transcripts were detected with qPCR as described above.

Biotinylated RNA pulldown

Biotinylated RNA pulldown was performed as described in (Hacisuleyman et al., 2014). ~10^7 V-SVZ-NSCs were pelleted and nuclei obtained as described for RIP protocol. To generate biotinylated RNA, *Pnky* sense and antisense were cloned into the PGEM-Teasy vector (Promega), which contains a T7 promoter. Biotinylated RNA was synthesized using the T7 High Yield Synthesis Kit (NEB) and biotinylated UTP (Roche). RNA probes were purified using the RNeasy mini kit (Qiagen). Immediately before use, 30 pmol RNA was resuspended in RNA structure buffer (10 mM HEPES pH 7, 10 mM MgCl2), heated to 65°C, then slow-cooled to 4°C in a thermocycler. RNA was run on a 1% agarose gel to ensure integrity and correct size.

To generate lysate, nuclear pellet was resuspended in RIP buffer with 1% NP-40 and rotated for 30 mins at 4°C, and then debris was cleared by centrifugation at max speed for 30 mins. For preclear, 40 ul/pulldown MyOne T1 beads (Invitrogen) were washed/prepared according to manufacturer's instructions and added to the lysate. Mixture was rotated at 4°C for 1 hour, beads were removed with magnetic rack and discarded. Precleared lysate was diluted 1:2 in RIP buffer without NP-40 (Final concentration of NP-40=0.5%), and probes and yeast tRNA (final concentration= 0.1 ug/ul) (Invitrogen) were added. Binding reaction is carried out overnight at 4°C. The next day, 40 ul washed MyOne T1 beads are added for an additional 1 hour. Tubes are added to the magnetic rack and washed 3x 10 mins at 4°C on the rack with wash buffer (RIP buffer with 1% NP-40 and 300 mM KCl). After final wash, 1X NuPage running buffer is added and beads are boiled for 10 mins to elute protein.

Nuclear fractionation

~10 million V-SVZ NSCs were resuspended in 10 mL PBS, 2 mL nuclear isolation buffer (1.28 M sucrose, 40 mM Tris, pH 7.5, 20 mM $MgCl_2$, 4% Triton X-100), 6 mL of water, and incubated on ice for 20 mins with frequent mixing. Nuclei were then pelleted at 2500xg for 15 mins, resuspended in lysis buffer (150 mM KCl, 25 mM Tris pH 7.5, 5 mM EDTA, 0.5% Igepal, 0.5 mM DTT) and incubated on ice for 30 mins. For whole cell lysate, cell pellet was resuspended in lysis buffer and incubated on ice for 30 mins. RNA was extracted using Trizol LS according to manufacturer's instructions. 1 ug of RNA was used for first-strand synthesis with the Transcriptor First Strand cDNA Synthesis kit (Roche) using oligo-dT primers. cDNA was used for qPCR with Sybr Green master mix (Roche), run on a Light Cycler 480 (Roche).

Mass Spectrometry Analysis

Selected SDS PAGE-separated bands were excised and in-gel digested with trypsin according to the established protocols (Jim nez et al., 2001) LC MS analyses of tryptic peptides

utilized LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, San Jose, CA) equipped with a NanoLC Ultra System (Eksigent, Dublin, CA), as described before (Roan et al., 2014). MS/MS data were interrogated with Mascot 2.2.04 search engine (Matrix Science) following their conversion to the .mgf format with an aid of Mascot Deamon (Matrix Science): Rodentia taxonomy of UniProt database (release 2014_01, 26206 entries) was searched. Dynamic modifications included sulfoxide oxidatation at Met, deamidation at Asn or Gln, and Gln to pyro-Glu conversion at N-terminus. Carbamidomethyl at Cys was included as a static modification. One missed tryptic cleavage was allowed. Precursor and fragment ion mass tolerances were set to 5 ppm and 0.2 Da, respectively. A target-decoy strategy with a 0.045 target false discovery (FDR) rate was used for protein identification (Elias and Gygi, 2007). Relative abundances of proteins in the sense and antisense samples were estimated on the basis of two independent approaches: spectral counting (Liu et al., 2004; Lundgren et al., 2010) and the exponentially modified protein abundance indices (emPAI) . Spectral counting utilized peptides identified with expectation values of 0.05 or lower (Plan A). Protein molar content (mol %) was calculated as described by (Ishihama et al., 2005) using emPAI values generated by the Mascot algorithm: only peptides identified at or above homology threshold are included. (http://www.matrixscience.com/help/quant_empai_help.html).

Time-lapse imaging and quantification

Near-clonal density cultures were established by trypsinization and subsequent mixing of infected cultures (shCtrl or shPnky) with wildtype, uninfected cultures at a ratio of 1:200 to give ~15 GFP+ cells/high powered field and cultured in proliferation media for 8 hours to allow cells to adhere to the plate. Cultures were switched to differentiation medium and imaged on a Leica SP5 inverted confocal microscope fitted with a Life Imaging Services microscope temperature control system. Cultures were maintained at 37°C and 5% CO2, 21% O2 and 8 optical sections were taken every 15 minutes for 3 days. Optical sections were summed and movies assembled

with ImageJ. Cell fate was determined by morphology and representative fields were confirmed by fixing and staining for Tuj1 and GFAP. To determine cell cycle time, the time between the first and second divisions was measured.

In utero electroporation and quantification

In utero electroporation was performed on E13.5 embryos from timed-pregnant wildtype Swiss-Webster mice (Simonsen labs) as described (Saito, 2006). Constructs used were PsicoR-shLuciferase (Shcontrol), and PsicoR-shPnky-2 (shPnky). Embryos were harvested 48 hours later, and heads were fixed in 4% PFA, washed in PBS, and then equilibrated in 20% sucrose before embedding in 1:1 OCT:20% sucrose and sectioning on Cryostat.

1-3 non-adjacent coronal sections per brain were imaged for quantification. Optical sections through the dorso-lateral telencephalon containing GFP+ electroporated cells was acquired at constant separation on a Leica SP5 Upright Confocal microscope. Three to four optical sections through the center were summed using ImageJ. Four animals from three separate surgeries were quantified for each experiment. For quanification of GFP+ cells in each zone, sections were costained with Nestin and Dcx. VZ was defined as NESTIN+DCX-, and CP was NESTIN-DCX+.

Human fetal tissue

Fetal cortical tissue was collected from elective pregnancy termination specimens at San Francisco General Hospital, usually within 2 h of the procedure. Research protocols were approved by the Committee on Human Research (institutional review board) at University of California, San Francisco.

qPCR Primers

Dlx1 primers: F- AGTTTGCAGTTGCAGGCTTT, R- ACTTGGAGCGTTTGTTCTGG

Dlx1as primers: F- GCAGACAGAATTGGGTCGTT, R- CTCAACTACCGCCTGCAAA

Dlx2 primers: F- GGCCTCACCCAA ACTCAG, R- AGGCACAAGGAGGAGAAGC

Dlx5 primers: F- ACTGACGCAAACACAGGTGA, R- CTGGTGACTGTGGCGAGTTA

Ascl1 primers: F- TCTCCTGGGAATGGACTTTG, R- CGTTGGCGAGAAACACTAAAG

Rplp0 primers: F- CCGATCTGCAGACACACACT, R- ACCCTGAAGTGCTCGACATC

Six3os primers: F- CCTCTGAGCCCACCTCCT, R- TGGTGAGCATACGAAGATGG

shRNA Oligo Sequences

Oligo sequences used were (target sequence in **BOLD**):

Six3os_sh1_F (From Rapicavoli et. al. 2011):

T**GACTTCAGTTGCCTCTCAT**TTCAAGAGA**ATGAGAGGCAACTGAAGTC**TTTTTTC

Six3os_sh1_R (From Rapicavoli et. al. 2011):

TCGAGAAAAAA**GACTTCAGTTGCCTCTCAT**TCTCTTGAA**ATGAGAGGCAACTGAAGTC**

Six3os_sh2_F:

T**GGTCAGAGTGAGATTGCTA**TTCAAGAGA**TAGCAATCTCACTCTGACC**TTTTTTC

Six3os_sh2_R:

TCGAGAAAAAA**GGTCAGAGTGAGATTGCTA**TCTCTTGAA**TAGCAATCTCACTCTGACC**

Dlx1as_sh4_F:

T**GAAGCTACATAGATGGTCA**TTCAAGAGA**TGACCATCTATGTAGCTTC**TTTTTTC

Dlx1as_sh4_R:

TCGAGAAAAAA**GAAGCTACATAGATGGTCA**TCTCTTGAA**TGACCATCTATGTAGCTTC**

Dlx1as_sh7_F:

T**GACACTGAAGAAGCTACAT**TTCAAGAGA**ATGTAGCTTCTTCAGTGTC**TTTTTTC

Dlx1as_sh7_R:

TCGAGAAAAAA**GACACTGAAGAAGCTACAT**TCTCTTGAA**ATGTAGCTTCTTCAGTGTC**

Luciferase_F:

T**GAGCTGTTTCTGAGGAGCC**TTCAAGAGA**GGCTCCTCAGAAACAGCTC**TTTTTTC

Luciferase_R:

TCGAGAAAAAA**GAGCTGTTTCTGAGGAGCC**TCTCTTGAA**GGCTCCTCAGAAACAGCTC**

Pnky_sh1_F:

T**GGACAATGGCTGAGAAAGC**TTCAAGAGA**GCTTTCTCAGCCATTGTCC**TTTTTTC

Pnky_sh1_R:

TCGAGAAAAAA**GGACAATGGCTGAGAAAGC**TCTCTTGAA**GCTTTCTCAGCCATTGTCC**

Pnky_sh2_F:

T**GATGACGTGGAGAGGATTT**TTCAAGAGA**AAATCCTCTCCACGTCATC**TTTTTTC

Pnky_sh2_R:

TCGAGAAAAAA**GATGACGTGGAGAGGATTT**TCTCTTGAA**AAATCCTCTCCACGTCATC**

shPTBP1_F:

T**GGGTGAAGATCCTGTTCAA**TTCAAGAGATTGAA**CAGGATCTTCACCCTTTTTTC**

shPTBP1_R:

TCGAGAAAAAA**GGGTGAAGATCCTGTTCAA**TCTCTTGAA**TTGAACAGGATCTTCACCC**

***Pnky* sequences**

>Human *PINKY*

TAAGCAGTGGTATCAACGCAGAGTACATGGGGAGGCGCCAGGGGCCCGGTTGGC
GCGAACGCCGGGTTCCGAGCACCCTGGGCTTCCTTGTCTGCCTCCCAGCGCGGCA
CCTCTTCGGGGCTCCCGAAACCTGAGCTCTCGCTGGTTTTAGGTCCAGACGGGGG
CCTCTCCACCGGTTCCTCCCCCGCCCCGGGCTCTGGGGCCCATTCTTTGGGCTGA
CCCTGTCAGGGCAGAGTCCGCGCGTCTGCCTGCCATTCTCCGCCCGCATAAAAGC
ACGTTGAAGGTGTCTCGGGCAGACACCTCCAGGTTTTGAATCAGTTTATTCCCTTTC
ACTGTTCAAAGCAGCTGTTCAAATACACAGGCTGCTTACGTTGACGTGGAGAGGAT
TTCAAACAACGCTAAAATGCTTTGAACTGACAAGGTGTCTTGATATCTCCCTCACTC
CATCCAGCACAGCTCCTCGAGATCACTCGCTAGGACAATGGCTGAGCAGGCGATT
CGTGCGGGCCTCGCCACCTCGGGGCGCGGACTGCGGGGTGTCCTAAGCCCCTTC
CGCAAGGACAGGATGGAGGCACCTGTAAGGAGATGCTGGCGCCACCCCAGCTTCT
CCCAGGTCCGGAGGAACCTCTACTCAGTCAATACCCTGAGCTGGACTTGTCTGAAG
AAACGGAGCCGACTCCCTCTTGCCGGGGTGCGCTGAGTGGAGGGGAAACATCCTC
GAATAACAGAACTACACCAAAAAGACACCCATGTTATCTCTCACACTTTCACACTCC
TCGAGATAGTGAGCCGGACCTGGGTCTTAGTAGCACCCAGTACCTTGACACAAACC
TCCCAAATTTCCACCTGAGTAACAGTTATGGGGTCAGTCCATGCACTGTAACTTGAA

CTCTAATTTATTAACTATTTCATCTAGTAAACACACTCACACCATATATAAAATAGCA
TTTATTTATTTCTATATACCAGGAGTTGGCAGAAAACCCACCGTGACCACTCCCATA
CATTGAGCTGGAGGCACACAATTACTAAAACAGAGGTGAGATGGTATTCATTTGAT
CTTAATTTTTTCTTATTTATGTAGTCCCAGGATAATAGAAATCAGGAAACAAAAGAAA
ACAAAGAATTTTCTGAGGAGATGGCCATTGGGGGAGTGGAGGTAGCAGCTGGTTTA
AACCTAAGTAAAACTAGAAAAAGAAACTGCTGTTTCCTTTTTCTTATATCCACCTTAG
AGGATCATGTTTGAACGTCCCTACTCCTCCTCCTCTTTTTAAAAAGCCTTGTCTCAG
TCATTCATTCCTGTGCTTCCTGCTCTTCTGCTAGACCCCAGCAGCTGTTTGATTTGG
TGAGGCCCCCCTCCAACCTCTGAGTGGAACTTCTTTTCTAAGGGCCTGCAGAATGT
CAAAACTGAGGCTCTGGCTTCGGAGCTAGAGCTTTGAACAGCCAATCCACACAAAA
AGGCAGCTGGCTGCTTTAATGAAAACTGCTATAAAGCTTCAAGAACTTTAGCCTTGG
GGGATGCATTTATAAGGAACATGGAAAATGCATTTCCAAGTTGCTGGTTCTTGGGA
GAGGCATAATAAACATTTACC

>Mouse *Pinky*

GGGAGAAGCAACTTCCTCTGGTCTTCTGGAGGTGTAACCTACGTGCCCAGTAGGATATACT
GCCGGGTTGTGAAATGTCCACGCCTCTCCCCAACTGTCTTCCTTCCCAGCCCTCCGGGCT
TGGCTTTCTTGCTTCCCCAGGAGTTCAAATCTCCAACCTGCGGAAGAATTCAGCTGCTTGA
AAGGACTTAAGGCAGTGTGCGGAGGACATCTCCTTTCTCCGCCAGTAAAGAGAGCTGTTC
AAAGACCGAGGCTGCTTACGATGACGTGGAGAGGATTTCAAACAACCTTAACAGGCTTTGA
ACTGACAAGAAGCGGTGATATCTGCCTCACTCTGGCACAGCTCCTCCAGTGCACTTGCTAG
GACAATGGCTGAGAAAGCACTTGGTGCTGGCCTcTCTGCCGGGGGCGAGGACTGCCCAAA
CCCCAGAGACCCTAAGGCAGGAGTTGCTGCACTaCAATGGAGAGGTCTTGTcTTAGGCTGC
ACCTCAGGTTCTcTCAGATcTTCTGGAAGGGCGTTATTCAGCGGTACTGTGGTTGCGCTGT
GCCAGCAGCTGCTTGATGGAGACCTCTTCTCCCCaCATCTGAATGGAACGTCTTTGCCCAG
AGTcTACAGAATGTCAAAACTGAGGCTCCGACCTCAGAGCTACAGCTTTAAGGACCACTCC
ACACAAAGAGGCAGCTGGTTGCTTTAATGAAAACTGCCTTTAAGCTTCAAGAACTGAGGCC
TTGGGGAATCCATTTATAAGGAGCCTAGAAAATGCATTTCCAAGTTGTATGTTcTTAGGAGA
TACAAAATAAAATTTACCTGAAAAAATTAATTGT

Chapter 3: Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo

Summary

Long noncoding RNAs (lncRNAs) have been described in cell lines and various whole tissues, but lncRNA analysis of development *in vivo* is limited. Here, we comprehensively analyze lncRNA expression for the adult mouse subventricular zone neural stem cell lineage. We utilize complementary genome-wide techniques including RNA-seq, RNA CaptureSeq, and ChIP-seq to associate specific lncRNAs with neural cell types, developmental processes, and human disease states. By integrating data from chromatin state maps, custom microarrays, and FACS purification of the subventricular zone lineage, we stringently identify lncRNAs with potential roles in adult neurogenesis. shRNA-mediated knockdown of two such lncRNAs, *Six3os* and *Dlx1as*, indicate roles for lncRNAs in the glial-neuronal lineage specification of multipotent adult stem cells. Our data and workflow thus provide a uniquely coherent *in vivo* lncRNA analysis and form the foundation of a user-friendly online resource for the study of lncRNAs in development and disease.

Introduction

The mammalian genome encodes thousands of long noncoding RNAs (lncRNAs), and it is becoming increasingly clear that lncRNAs are key regulators of cellular function and development. Loss-of-function studies performed in cell culture indicate that lncRNAs can regulate gene transcription through the targeting and recruitment of chromatin modifying complexes (Guttman et al., 2011; Khalil et al., 2009; Tsai et al., 2010). While it is now evident that lncRNAs have important cellular and molecular functions, how they participate in

development *in vivo* is poorly understood.

Emerging studies suggest that lncRNAs play critical roles in central nervous system (CNS) development. For instance, in embryonic stem cells (ESCs), specific lncRNAs repress neuroectodermal differentiation (Guttman et al., 2011), and during *in vitro* differentiation of ESC-derived neural progenitor cells (ESC-NPCs), lncRNA expression is dynamic (Mercer et al., 2010). In the mouse brain, some lncRNAs are regionally expressed (Mercer et al., 2008), including among the six layers of the adult cortex (Belgard et al., 2011). *In vivo* functional data is limited, but mice null for the lncRNA *Evf2* have abnormal GABAergic interneuron development and function (Bond et al., 2009), and morpholino inhibition of two CNS-specific lncRNAs in Zebrafish affects brain development (Ulitsky et al., 2011).

The ventricular-subventricular zone (V-SVZ) of the adult mouse brain represents an ideal system for the study of lncRNAs *in vivo*. Throughout life, V-SVZ neural stem cells (V-SVZ-NSCs) generate large numbers of neuroblasts that migrate to the olfactory bulb (OB) where they differentiate into interneurons (**Figure 3.1A**). In addition, V-SVZ-NSCs are multipotent, capable of generating astrocytes and oligodendrocytes, the other major cell types of the CNS. In contrast to the embryonic brain wherein multipotent precursor cells are inherently transient, continually changing their developmental potential and location over time and with organ morphogenesis, the adult V-SVZ retains its NSC population in a stable, spatially restricted niche, producing neurons and glia throughout life (Kriegstein and Alvarez-Buylla, 2009). This enduring population of multipotent stem cells and its well-characterized daughter cell lineages make the V-SVZ a particularly tractable *in vivo* model for molecular-genetic studies of development. The V-SVZ has been used to elucidate key principles of neural development including the role of signaling molecules, transcription factors, microRNAs, and chromatin modifiers (Ihrie and Alvarez-Buylla, 2011). We have previously shown that the *Mixed-lineage leukemia 1 (Mll1)* chromatin modifying factor is required for the V-SVZ neurogenic lineage (Lim et al., 2009), and recent studies indicate that MLL1 protein can be targeted to specific loci by

lncRNAs (Bertani et al., 2011; Wang et al., 2011).

Here, we leveraged the V-SVZ-OB system to develop a greater understanding of lncRNA expression and function. First, we used Illumina-based cDNA deep sequencing (RNA-seq) and *ab initio* reconstruction of the transcriptome to generate a comprehensive lncRNA catalogue inclusive of adult NSCs and their daughter cell lineages. This lncRNA catalogue informed a subsequent RNA Capture-seq approach, which increased the read coverage and read length for our V-SVZ cell analysis, validating the transcript structure and expression of many of these novel lncRNAs. Gene coexpression analysis identified sets of lncRNAs associated with different neural cells types, cellular processes, and neurologic disease states. In our analysis of genome-wide chromatin state maps, we identified lncRNAs that -- like key developmental genes -- demonstrate chromatin-based changes in a neural lineage-specific manner. Using custom lncRNA microarrays, we found that lncRNAs are dynamically regulated in patterns reminiscent of known neurogenic transcription factors. To define lncRNA expression changes throughout the V-SVZ neurogenic lineage *in vivo*, we acutely isolated the major cell types of the V-SVZ with fluorescent activated cell sorting (FACS) and probed lncRNA expression with our custom microarrays. We integrated these diverse experimental approaches to develop an online resource useful for the identification of lncRNAs with potential roles in V-SVZ neurogenesis (http://neurosurgery.ucsf.edu/danlimlab/lncRNA/) Furthermore, expression and shRNA-mediated knockdown experiments confirmed functional roles for lncRNAs identified by our integrative approach. Overall, our study demonstrates a generalizable workflow that assimilates genome-wide bioinformatic strategies with experimental manipulations for the identification of lncRNAs that regulate development.

Results

Cataloging lncRNAs in the adult brain neurogenic zones

Because lncRNAs exhibit tissue-specific expression, previous mouse lncRNA databases were not likely comprehensive for lncRNAs involved in adult neurogenesis. Thus, we identified lncRNAs expressed in the adult brain neurogenic niches by employing an RNA-seq and *ab initio* transcriptome reconstruction approach. First, we generated cDNA libraries of poly-adenylated RNA extracted from microdissected adult V-SVZ tissue, which contains NSCs, transit amplifying cells, and young migratory neuroblasts. To include the transcriptome of later stages of neurogenesis and neuronal function, we also generated cDNA libraries from the OB. Furthermore, we generated cDNA libraries from microdissected adult dentate gyrus (DG), the other major adult neurogenic niche, which locally contains all cell types of an entire neuronal lineage. **Figure 3.1A** shows a schematic of regions used for the cDNA libraries.

We used Illumina-based sequencing to obtain paired-end reads of these cDNA libraries from the V-SVZ (229 million reads), OB (248 million reads), and DG (157 million reads). To broaden our lncRNA catalog, we also included RNA-seq data from embryonic stem cells (ESCs) and ESC-derived neural progenitors cells (ESC-NPCs) (Guttman et al., 2010). With this collection of over 800 million paired end reads, we used Cufflinks (Trapnell et al., 2010) to perfom *ab initio* transcript assembly. This method reconstructed a total of 150,313 multi-exonic transcripts, of which 140,118 (93%) overlapped with known protein-coding genes. Our lncRNA annotation pipeline (see Figure 3.1B and Experimental Procedures) identified 8992 lncRNAs encoded from 5731 loci (**Supplementary File 1**). 6876 (76.5%) were novel compared to RefSeq genes, 5044 (56.1%) were novel compared to UCSC known genes, and 3680 (40.9%) were novel compared to all Ensembl genes. Interestingly, 2108 transcripts (23.4%) were uniquely recovered from our V-SVZ/OB/DG reads.

To substantiate the non-coding nature of our lncRNA candidates, we used the Coding Potential Calculator (Kong et al., 2007) and found that over 80% of these transcripts have essentially no protein coding potential (**Figure S3.1A**). Consistent with previous studies,

lncRNAs were expressed at lower levels than protein-coding genes (2.49 fold difference; Mann-Whitney U, p<0.0001) (**Figure S3.1B**), and their exons were less strongly conserved than protein-coding exons by PhastCons scores (**Figure S3.1C**).

The transcriptional start site (TSS) of some lncRNAs is proximal (<10 kb) to the promoters of protein-coding genes (Cabili et al., 2011; Hung et al., 2011), and we found that the TSS of 2265 lncRNAs (25.2%) in our catalog were located within 5 kb of a protein coding gene promoter (**Figure S3.1D**). Gene onotology (GO) analysis with the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) revealed that these protein-coding neighbors are enriched for homeodomain-containing transcription factors, genes expressed in the brain, and genes that are typically repressed by Polycomb Repressive Complex 2 in ESCs (**Figure S3.1E)**. While some lncRNAs had strongly correlated expression with their protein-coding neighbor, as a group they had no obvious correlation (**Figure S3.1F**), indicating that expression of this subset of lncRNA is not likely related to local transcriptional activity of protein coding genes.

To verify that the cDNA libraries of the V-SVZ and OB together represent a transcriptome enriched for adult neurogenesis, we first analyzed mRNA expression in the RNA-seq data. Differential gene expression identified 1621 genes enriched >2 fold in the V-SVZ cDNA library as compared to the cDNAs from cells in the adjacent non-neurogenic striatum (76.4 million reads). As the primary site where NSCs and transit amplifying cells proliferate, the V-SVZ was enriched for gene ontology (GO) terms related to cell cycle and mitosis (**Figure S3.2A and S3.2B**). Neuroblasts migrate through the V-SVZ and into the OB, and, as expected, transcripts related to this migratory neuroblast stage of neurogenesis were enriched in these regions (Lim et al., 2006). The V-SVZ/OB expression profile included transcription factors known to play key roles in adult neurogenesis, such as *Dlx1*, *Dlx2*, *Ascl1*, and *Pax6* (Hsieh, 2012). Furthermore, *in situ* hybridization (ISH) data from the Allen Brain Atlas (Lein et al., 2007) confirmed the regional expression of many of these V-SVZ/OB-enriched genes (**Figure S3.2C and Figure S3.2D**), and

the V-SVZ/OB transcriptional profile (923 genes) was enriched for GO terms related to cell migration, development, and neurogenesis (**Figure S3.2E**).

lncRNAs have temporally and spatially unique expression patterns

To explore lncRNA expression patterns in multiple adult brain regions and embryonic forebrain development, we analyzed RNA-seq data of the six layers of the adult cortex (Belgard et al., 2011), adult whole prefrontal cortex (PFC), adult preoptic area (POA), whole embryonic day 15 (E15) brain (Gregg et al., 2010), and specific regions of the developing E14.5 cortex (ventricular zone, intermediate zone, and cortical plate) (Ayoub et al., 2011) (**Figure 3.2A and Supplementary File 2**). Unsupervised hierarchical clustering of expression profiles revealed region-specific and temporally related expression of both mRNAs and lncRNAs (**Figure 3.2B and 2C**). We calculated a specificity score for each transcript (Cabili et al., 2011) and found that the mean score was 0.57 (s.d. 0.21) for lncRNAs, while it was 0.45 (s.d. 0.17) for mRNAs (p-value < 10^{-325}, Wilcoxon Rank Sum Test); thus, lncRNAs exhibit greater brain region and temporal specificity than mRNAs, suggesting that they play important roles in the determination and/or function of specific neural cell types.

lncRNAs are associated with specific neural cell types and neurological disease states

To begin to infer functions for lncRNAs, we investigated the relationship between mRNA and lncRNA transcription by using gene co-expression analysis (GCA) to identify groups of transcripts, or 'modules,' whose variation in expression correlate across different brain regions and developmental time points. For mRNAs, module membership distinguishes sets of genes that correspond to specific cell types and biological processes (Oldham et al., 2008), and a similar 'guilt by association' approach has been used to assign putative functions to lncRNAs based on their co-expression with protein-coding genes (Guttman et al., 2009).

Using RNA-seq data from 22 samples (**Figure 3.2A and Supplementary File 2**), we constructed transcript co-expression networks comprised of both mRNAs and lncRNAs. For the 56 modules of co-expressed transcripts, we performed enrichment analysis using gene sets from the Molecular Signatures Database (Subramanian et al., 2005) and other sources (Bult et al., 2008; Cahoy et al., 2008; Thomas et al., 2011; Zhang et al., 2010) to relate modules (described by 'color', **Figure 3.3A-F**) to specific adult neural cell types including cortical neurons (purple), striatal neurons (salmon), ependymal cells (green), and oligodendrocytes (grey60) (**Supplementary File 3**).

The dark red module (**Figure 3.3E**) was enriched for glial markers but also had a large number of known early neurogenic factors as prominent members (**Supplementary File 3**). This module was specifically associated with the ventricular zone of the embryonic brain, which contains radial glia, the stem cells of the developing brain and precursors of the adult V-SVZ-NSCs. We additionally identified a module (red, **Figure 3.3F**) specifically associated with the 'stemness' transcriptional program and the cell cycle.

Interestingly, some modules were also associated with human disease, notably Huntington's disease, Alzheimer's disease, convulsive seizures, major depressive disorder, and various cancers (**Figure S3.3A-F**). For instance, the striatal neuron module (salmon) correlated with a gene expression set misregulated in Huntington's disease mouse models, suggesting a potential role for the 88 lncRNAs in this set in this neurodegenerative condition. Taken together, our co-expression analysis provides an important resource as the first annotation of lncRNAs to specific neural cell types *in vivo* and neurological disease states.

RNA CaptureSeq verifies V-SVZ lncRNA expression and identifies novel splice isoforms

Because many lncRNAs have not been previously annotated and are expressed at low levels, we employed a targeted RNA capture and sequencing strategy (CaptureSeq) to more robustly identify and characterize lncRNAs in the adult V-SVZ. With RNA CaptureSeq, cDNAs

are hybridized to probe libraries tiled against the genomic regions of interest, eluted, and then sequenced (**Figure 3.4A**). Through this enrichment, the read coverage of targeted transcripts is dramatically increased (Mercer et al., 2012). Furthermore, by using a 454 GS-FLEX Titanium instrument for sequencing, we obtained longer reads, which improves the delineation of rare splice isoforms.

For our RNA CaptureSeq probe library, we tiled across 100 MB of putative lncRNA loci and 30 MB of protein-coding regions as a control. We used this library to capture V-SVZ cDNA for sequencing (5,882,293 reads, median length of 356 bases per read). As expected, *de novo* assembly of sequences accurately reconstructed protein-coding transcripts and previously annotated lncRNAs (median identity of 90% for protein-coding RefSeq genes and median identity of 95% for annotated non-coding RefSeq RNA). As an example, *Evf1* and *Evf2*, lncRNAs with roles in neural development, have overlapping genomic structures (Feng et al., 2006), and RNA CaptureSeq identified and distinguished both transcripts in the V-SVZ (**Figure S3.4A**). RNA CaptureSeq also eliminated sequencing bias related to transcript abundance (**Figure S3.4B and S3.4C**), and measured expression values were well correlated between CaptureSeq and conventional RNA-seq strategies (**Figure S3.4D**).

The enrichment and longer reads provided by RNA CaptureSeq enabled the identification of rare lncRNAs as well as uncommon splice isoforms in the V-SVZ transcriptome, yielding more than 3,500 novel lncRNAs that could not be detected by the short-read sequencing technology. For example, CaptureSeq identified a lncRNA transcript with an intron overlapping *Pou3f3*, a known neurogenic transcription factor (**Figure 3.4B**). In addition to this discovery of a novel lncRNA locus downstream of *Pou3f3*, RNA CaptureSeq also identified splice isoforms that includes exons of a previously annotated lncRNA *(2620017I09Rik)* that lies upstream of the *Pou3f3* locus. Some lncRNAs are transcribed from multiple transcriptional start sites (TSS), which can be a challenge for transcript assembly. Adjacent to the locus of neurogenic transcription factor *Nr2f1*, CaptureSeq identified a series of lncRNAs originating

from 4 unique TSS's. This organization of protein-coding gene and multiple lncRNAs is conserved in humans, hinting at an evolutionarily conserved functional significance (**Figure 3.4C**). Thus, RNA CaptureSeq, in addition to providing a genome-wide validation of our V-SVZ lncRNA analysis, demonstrated previously underappreciated complexity to the structure of lncRNA loci. A complete annotation of Capture-Seq derived transcripts is available at http://neurosurgery.ucsf.edu/danlimlab/lncRNA/.

Correlation between histone modifications and lncRNA expression.

Methylation of histone lysine residues is a critical determinant of transcriptional activity(Jenuwein and Allis, 2001). In previous work, lncRNA loci have been identified in part by the presence of H3K4me3 at the TSS (Guttman et al., 2009). For protein-coding genes, H3K4me3 enrichment at the TSS correlates with active transcription whereas H3K27me3 is associated with a repressed state. Genes that are "bivalent" for both H3K4me3 and H3K27me3 are generally silenced but remain transcriptionally "poised" for activation or repression (Bernstein et al., 2006). To investigate whether lncRNA loci exhibit a similar correlation between promoter histone modifications and transcription, we performed ChIP-seq for H3K4me3 and H3K27me3 in V-SVZ-NSC cultures and included sequencing data from ChIP-seq and RNA-seq studies of mouse ESCs, ESC-NPCs, and mouse embryonic fibroblasts (MEFs) (Mikkelsen et al., 2007).

In V-SVZ-NSCs, 3671 (40.8%) lncRNAs were marked by either H3K4me3 or H3K27me3 and 928 (10.3%) were bivalent (**Figure S3.5A**). As has been described for protein-coding genes, these TSS chromatin modifications correlated strongly with lncRNA expression levels: lncRNAs monovalent for H3K4me3 exhibited higher expression levels than those by marked by only H3K27me3 or bivalent chromatin domains ($p<0.0001$, Mann-Whitney U, **Figure S3.5B**). These data suggest that transcription of both lncRNAs and protein coding genes utilizes similar chromatin-based regulatory mechanisms.

A subset of bivalent lncRNAs in ESCs become resolved to monovalent H3K4me3 in V-SVZ-NSCs

In ESCs, bivalent domains identify key developmental genes. As ESCs differentiate into lineage-specific cell populations, many of these bivalent genes become activated (monovalent H3K4me3) or repressed (monovalent H3K27me3), reflecting the lineage specification and restriction of developmental potential (Bernstein et al., 2006). Thus, genes that are more likely to play roles in the neural identity of V-SVZ NSCs would be those that are bivalent in ESCs, activated in V-SVZ-NSCs, and also repressed (H3K27me3 monovalent or bivalent) in a non-neural cell type (MEFs). We found 302 protein-coding genes that meet these criteria, and analysis revealed that the most statistically significant GO terms for these activated genes pertain to early brain development (**Figure S3.5C**). For example, proneural *Ascl1*, *Pou3f3*, and *Pou3f2* were bivalent in ESCs, H3K4me3-monovalent in V-SVZ-NSCs, and H3K27me3-repressed in MEFs (**Figure S3.5D**).

100 lncRNAs have a similar pattern of chromatin-based changes (**Figure 3.5A**). Furthermore, 76% of this set of lncRNAs was also monovalent for H3K4me3 in ESC-NPCs, suggesting that these lncRNAs are common to an early neural development transcriptional program. An example is *lnc-pou3f2*: this novel lncRNA is 2 kb upstream of known neurogenic transcription factor *Pou3f2*. Both the *lnc-pou3f2* and the *Pou3f2* loci were bivalent in ESCs, monovalent for H3K4me3 in NSCs, and H3K27me3 monovalent in MEFs (**Figure 3.5B**). Given the known relationship between chromatin modifications and the expression of key developmental regulators, we propose that this set of lncRNAs is enriched for those that play roles in early neural commitment in the adult V-SVZ.

lncRNAs can retain bivalency in an adult stem cell population

Tissue-specific stem cells also retain bivalency at key loci, possibly reflecting retained

gene expression plasticity (Cui et al., 2009; Lien et al., 2011). Protein coding genes that are bivalent in ESCs and NSCs were highly enriched for GO terms related to neurogenesis (*e.g.,* neuron differentiation, axonogenesis, **Figure S3.5E**). For instance, in adult V-SVZ NSCs, *Dlx1* and *Dlx2* are transcription factors required for interneuron development, and these were bivalent in NSCs (**Figure S3.5F**). Thus, the identification of lncRNAs that are bivalent in both ESCs and NSCs might enrich for those involved in neuronal differentiation. 583 lncRNAs met these criteria (**Figure 3.5C**), such as three splice variants encoded from a novel lncRNA locus located ~50 kb upstream of protein-coding gene *Odf3l1* (**Figure 3.5D**). We propose that lncRNAs bivalent in V-SVZ-NSCs are enriched for those that function in neuronal lineage specification.

lncRNAs are dynamically regulated during neuronal differentiation

We next sought to define the dynamic changes in lncRNA expression during V-SVZ neurogenesis. V-SVZ-NSCs cultured as a monolayer can efficiently recapitulate key aspects of *in vivo* neurogenesis as assessed by immunocytochemistry (ICC, **Figure 3.6A and S3.6A**) (Scheffler et al., 2005). We generated cDNA libraries from V-SVZ-NSC cultures in self-renewal conditions and after 1,2, and 4 days (1,2, and 4d) of differentiation and hybridized to both custom lncRNA (see **Methods**) and standard gene expression arrays. Included in the set of upregulated transcripts were genes related to V-SVZ neurogenesis (*e.g., Dlx1/2* and *Dlx5/6*) **(Supplementary File 4).** Also as expected, genes expressed at higher levels early in the V-SVZ lineage (*e.g. Egfr* and *Nestin*) were in the set of downregulated transcripts (**Supplementary File 4**). Like mRNA transcripts, lncRNAs also exhibited similar patterns of induction and repression (**Figure 3.6B and S3.6B**) over this 4-day differentiation timecourse.

***In vivo* V-SVZ lineage analysis of lncRNA expression**

The adult V-SVZ contains three major cell types that represent a developmental continuum: (1) activated NSCs, which express glial fibrillary acidic protein (GFAP) and the

epidermal growth factor receptor (EGFR), (2) transit amplifying cells, which are EGFR-positive but GFAP-negative, and (3) migratory neuroblasts, which are CD24-positive (**Figure 3.6C**). In addition, the V-SVZ contains GFAP+ cells that do not express EGFR, and these have been termed 'niche' astrocytes (Pastrana et al., 2009). We used these cell-specific characteristics to perform fluorescent-activated cell sorting (FACS) to acutely isolate cell populations representing each stage of this neurogenic lineage and the niche astrocytes (**Figure 3.6D**). cDNA libraries for each V-SVZ cell type were generated and hybridized to our custom lncRNA and standard gene expression microarrays. Expression levels of both protein-coding genes and lncRNAs were visualized by heat maps organized by k-means clustering (transcripts) and unsupervised hierarcical clustering (cell types) (**Figure 3.6E and S3.6C**).

To confirm the separation of V-SVZ cells, we examined differential mRNA expression. 12,812 protein-coding probesets were differentially expressed (>2-fold) in a comparison between activated V-SVZ-NSCs and migratory neuroblasts (**Supplementary file 4**). As V-SVZ-NSCs become activated and differentiate into transit-amplifying cells, *Dlx1/2* and *Ascl1* become upregulated (Doetsch et al., 2002; Kim et al., 2007) and this was reflected in our transcriptional profiles (**Supplementary file 4**). As expected, the transcriptome of migratory neuroblasts was enriched for *Dlx1/2* downstream targets, including *Dlx5* and *Arx*, as well as markers of young neurons, including *Tubb3* (**Supplementary File 4**). Thus, these transcriptomes represent distinct stages of the V-SVZ neurogenesis and also distinguish niche astrocytes from NSCs.

Similar to the cell culture data (**Figure 3.6B**), we found sets of lncRNAs that showed transient increases in TA cells, repression throughout differentiation, and significant induction in the terminally differentiated neuroblast population (**Figure 3.6E**). By integrating our chromatin state maps with this microarray expression data, we were able to begin to define a 'lncRNA signature' for each stage of neurogenesis *in vivo* (**Supplementary File 5**).

Identification of lncRNAs with roles in adult V-SVZ neurogenesis

To facilitate the identification of lncRNAs with potential roles in V-SVZ neurogenesis, we constructed an online resource that allows the user to easily filter the lncRNA catalogue for multiple variables including locus chromatin-state, expression in FACS-isolated V-SVZ cells, and regulation during in vitro neurogenesis (http://neurosurgery.ucsf.edu/danlimlab/lncRNA/). Using this resource, we filtered for those lncRNAs that were bivalent in ESCs and H3K27me3-repressed in MEFs. Of this set, which includes lnc-pou3f2, 100 lncRNAs were monovalent for H3K4me3 in V-SVZ-NSCs (**Figure 3.5A and 3.5B**), which would predict expression in the adult V-SVZ; indeed, in situ hybridization (ISH, **Figure S3.7A and S3.7B**) revealed lnc-pou3f2 expression in the V-SVZ, and, as predicted by the FACS microarray data, this transcript was not detected in the OB (**Figure S3.7B**).

Like *lnc-pou3f2*, lncRNA *Six3os* was also monovalent for H3K4me3 in V-SVZ-NSCs and downregulated in neuroblasts. Consistent with these observations, *Six3os* transcripts were detected in the V-SVZ but not the OB core (**Figure 3.7A**). To further investigate the role of *Six3os* in V-SVZ NSCs, we used lentiviruses to separately introduce two different short hairpin RNA (shRNA) sequences to knockdown *Six3os* (LV-sh-*Six3os*-GFP) in monolayers of V-SVZ-NSCs. After confirming *Six3os* knockdown in proliferating NSCs (**Figure S3.7C**), we assessed neuronal and glial lineages from LV-sh-*Six3os*-GFP infected cells in comparison to controls infected with LV-sh-luc-GFP. After 7 d of differentiation, there were 2-fold fewer Tuj1-positive cells and 2.5-fold fewer cells expressing OLIG2, a marker of the oligodendrocyte lineage (Zhou and Anderson, 2002). These decreases were accompanied by an increase in the number of GFAP+ cells (**Figure 3.7B and 3.7C**).

The *Dlx1/2* bigene cluster encodes lncRNA *Dlx1as* , and this locus was also bivalent in ESCs and H3K27me3 monovalent in MEFs. We used V-SVZ-NSC monolayer cultures to further investigate the chromatin state of the *Dlx1as* TSS. In self-renewal conditions, *Dlx1as* was bivalent, and after 30 hours (h) of differentiation, H3K27me3 decreased, correlating with the

start of *Dlx1as* upregulation (**Figure S3.7D and S3.7E)**. Interestingly, we also found enrichement of the H3K27me3-specific demethylase JMJD3 (Agger et al., 2007) at the *Dlx1as* TSS during differentiation (**Figure S3.7F**), suggesting that this chromatin-modifying factor plays a role in the activation of this lncRNA. Consistent with the transcriptional upregulation of *Dlx1as* during V-SVZ neurogenesis *in vitro*, we observed robust *Dlx1as* expression in V-SVZ regions with migratory neuroblasts and the OB core **(Figure 3.7D)**. We designed two knockdown constructs targeting the splice junction of *Dlx1as*, and verified that these constructs target *Dlx1as* and not full-length *Dlx1* transcript (**Figure S3.7G**). Knockdown of *Dlx1as* caused a decrease in expression of *Dlx1* and *Dlx2* after two days of differentiation compared to control (**Figure S3.7H**), suggesting this lncRNA can regulate expression of its protein-coding gene neighbors. After 7 d of differentiation, we found a nearly 3-fold decrease in Tuj1+ neuroblasts, and a ~60% increase in the number of GFAP+ astrocytes. In contrast to knockdown of *Six3os*, the production of OLIG2+ cells was unaffected by *Dlx1as* knockdown (**Figure 3.7B and 3.7E**).

Discussion

We performed an in-depth analysis of lncRNA expression of adult V-SVZ-OB neurogenesis, an excellent *in vivo* model system for the study of multipotent stem cells and neural development. Our use of two high-throughput sequencing-based approaches for the study of the lncRNA transcriptome (RNA-seq and RNA CaptureSeq) provided complementary datasets that together allowed the identification of thousands of novel lncRNAs, confirmation of rare transcripts, and resolution of previously unappreciated complexity of lncRNA loci.

Like the loci of genes encoding key developmental transcription factors, a subset of lncRNA loci showed changes of chromatin state during lineage specification. By integrating these chromatin state maps with data from custom microarrays and FACS purification of the V-SVZ lineage, our online resource (http://neurosurgery.ucsf.edu/danlimlab/lncRNA/) and files

(**Supplementary File 5**) facilitate the identification lncRNAs with potential roles in adult NSCs as well as neural development. Interestingly, we found that *Dlx1as* is required selectively for the V-SVZ neuronal lineage, whereas *Six3os* appears to play a role in both neuronal and oligodendrocyte differentiation. These data indicate that lncRNAs can play key roles in the glial-neuronal lineage specification of multipotent adult stem cells.

2,265 lncRNAs had proximal protein-coding gene neighbors (**Figure S3.1D)**, and this gene set was enriched for homeodomain-containing genes. For instance, *Six3os* is proximal to the *Six3* homeobox gene, and *Dlx1as* is encoded from the *Dlx1/2* bigene cluster (Liu et al., 1997; Dinger et al., 2008). Interestingly, *Dlx1as* knockdown caused a decrease in *Dlx1/2* expression (**Figure S3.7H**), suggesting that this lncRNA plays a role in neuronal differentiation by regulating expression of its homeobox gene neighbors. In developing retina, knockdown of *Six3os* results in deficits in lineage specification through modulation of SIX3 activity (Rapicavoli et al., 2011); it will be interesting to investigate whether the defect in V-SVZ neurogenesis with *Six3os* knockdown (**Figure 3.7C**) similarly involves a change in SIX3 activity. Taken together, our genome wide analysis and functional data further support the notion that lncRNAs and homeobox-gene neighbors function cooperatively (Rapicavoli and Blackshaw, 2009).

A recent model of lncRNA action suggests that lineage-specific lncRNAs become activated during differentiation and guide histone modifications that create cell-type specific transcriptional programs (Guttman et al., 2011). MLL1 is a trithorax group (trxG) chromatin-modifying factor that is enriched at *Dlx2* during V-SVZ NSC differentiation and is required for proper *Dlx2* expression (Lim et al., 2009); however, how MLL1 is targeted to *Dlx2* is not known. Interestingly, in mouse ESCs, lncRNA *Mistral* directly binds MLL1 and recruits it to *Hoxa6* and *Hoxa7 (*Bertani et al., 2011) and lncRNA HOTTIP recruits MLL1 through an interaction with WDR5 to distal *HOXA* genes in human fibroblasts (Wang et al., 2011). Our work provides a useful resource for the identification of such lncRNAs. For instance, lncRNAs that immunoprecipitate with chromatin modifiers could be identified by hybridization to the lncRNA

microarray and then filtered online by multiple other criteria (*e.g.,* enrichment in neuroblasts, upregulation during neurogenesis, bivalency in ESCs, repression in MEFs).

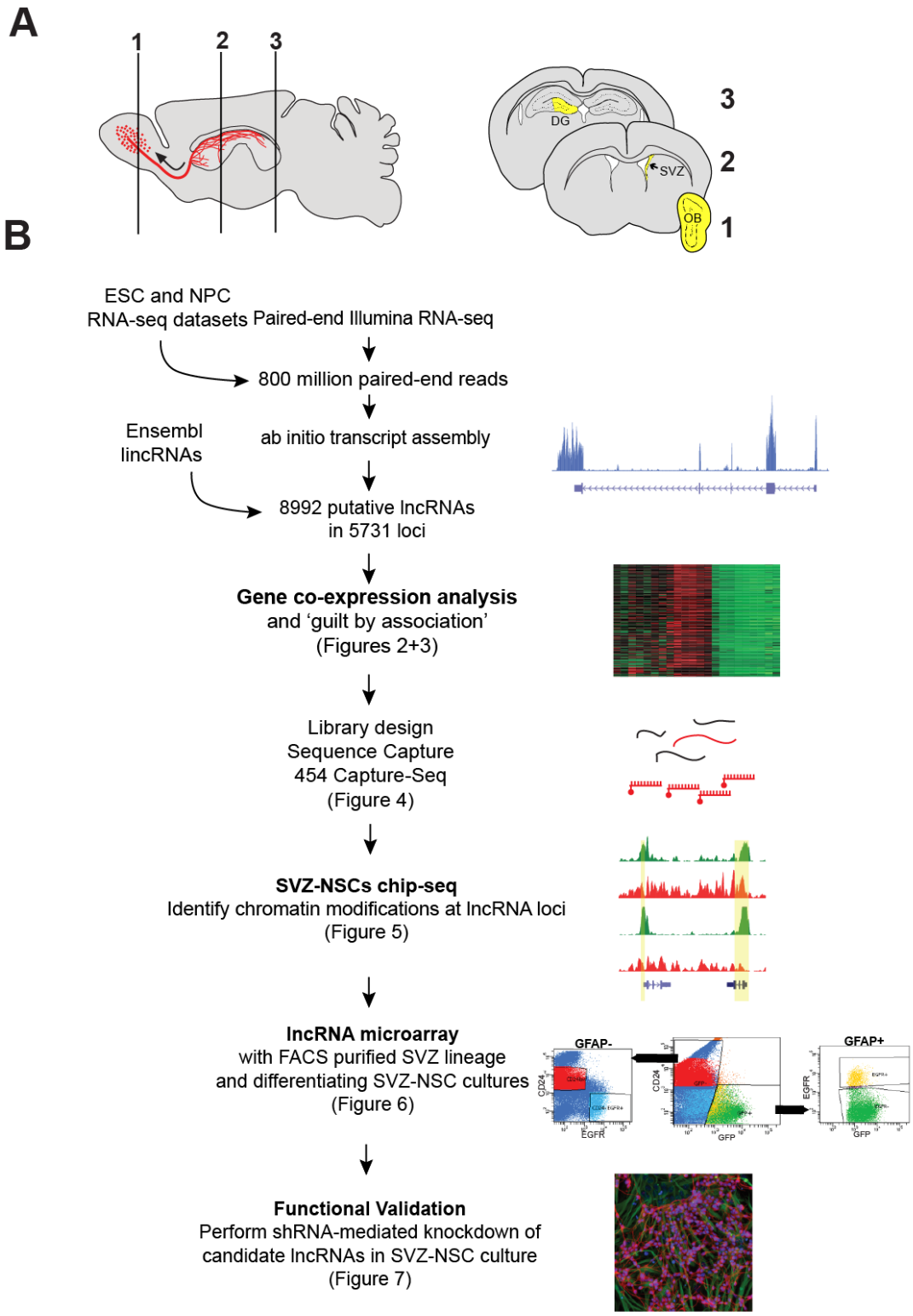
Our analysis of chromatin state maps and transcript expression suggest that histone modifications correlate with lncRNA expression in a manner similar to that of protein coding genes. Some lncRNA loci were bivalent in both ESCs and V-SVZ-NSCs, and many of these lncRNA loci became transcriptionally active in V-SVZ neuroblasts, supporting their candidacy as key determinates of neurogenesis. In V-SVZ-NSC monolayer cultures, *Dlx1as* was bivalent and H3K27me3 repression decreased during neuronal differentiation (**Figure S3.7D**), correlating with the upregulation of *Dlx1as* transcription (**Figure S3.7E**). Interestingly, we also found enrichment of the H3K27me3-specific demethylase JMJD3 at the *Dlx1as* locus (**Figure S3.7F**), suggesting that active removal of repressive histone modifications plays a role in the expression of lncRNAs. Overall, our data raise the possibility that lncRNA loci, like protein coding genes, are targeted by chromatin modifying factors that have critical roles in development.

While this study attempted to be as comprehensive as possible, it is possible that some lncRNAs important for V-SVZ neurogenesis were not identified. The initial sequencing experiments were performed on microdissected tissues that contain several cell types. Even at our sequencing depth, transcripts that are expressed at low copy number in a small number of cells might not be detected. Despite this potential shortcoming, we were still able to identify thousands of novel lncRNA transcripts. Furthermore, our initial catalog proved sufficient for our primary objective, which was to integrate complementary data analysis strategies and experimental methods to identify lncRNA expression patterns coherent to an *in vivo* experimental model system.

The role of lncRNAs in development and disease is in the early states of investigation, and our analysis of the V-SVZ lineage provides a resource for the movement of this research into *in vivo* studies. More broadly, this work presents a generalizable workflow for the identification and categorization of novel transcripts, both coding and noncoding.

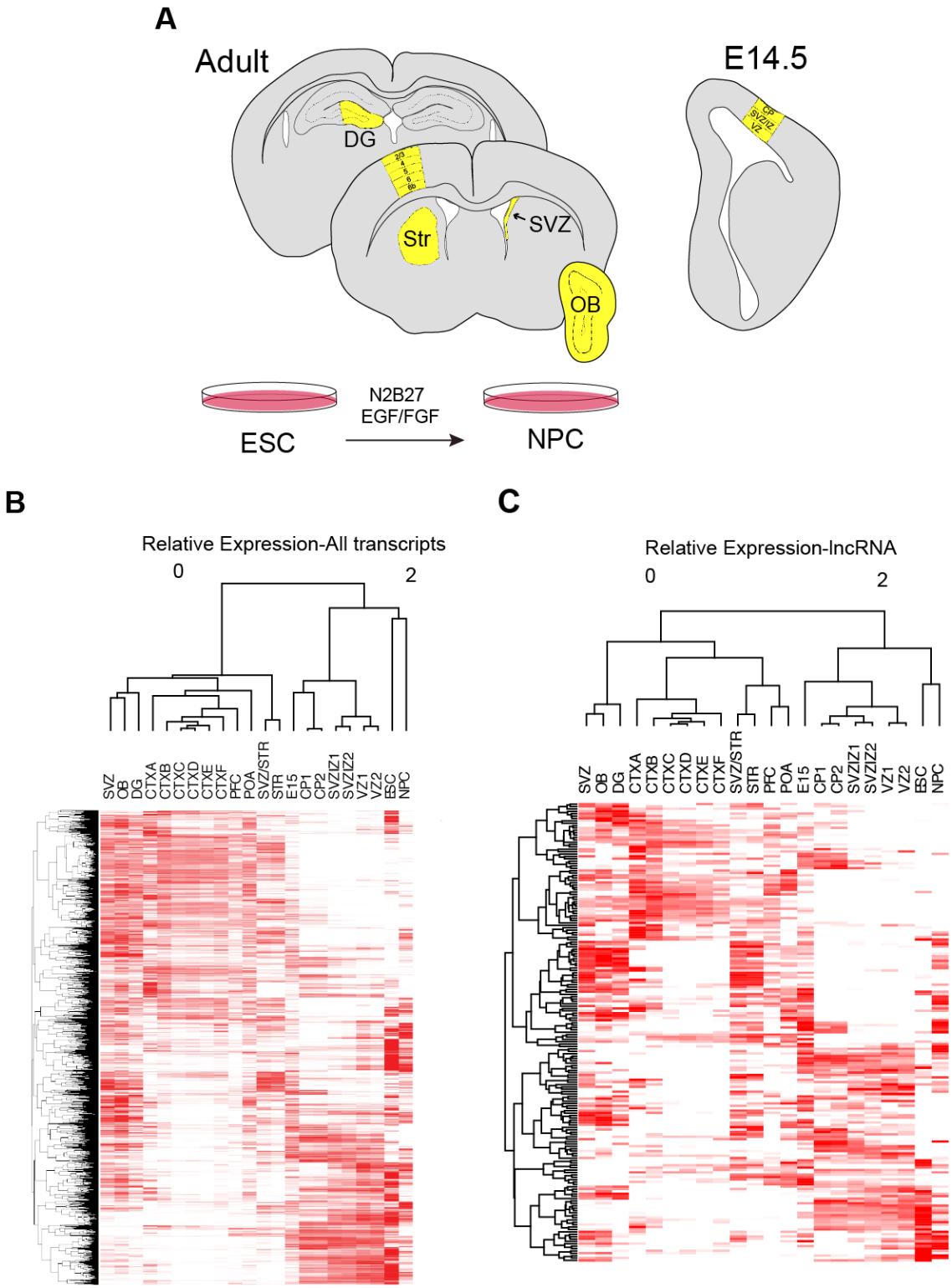
Figures

Figure 3.1



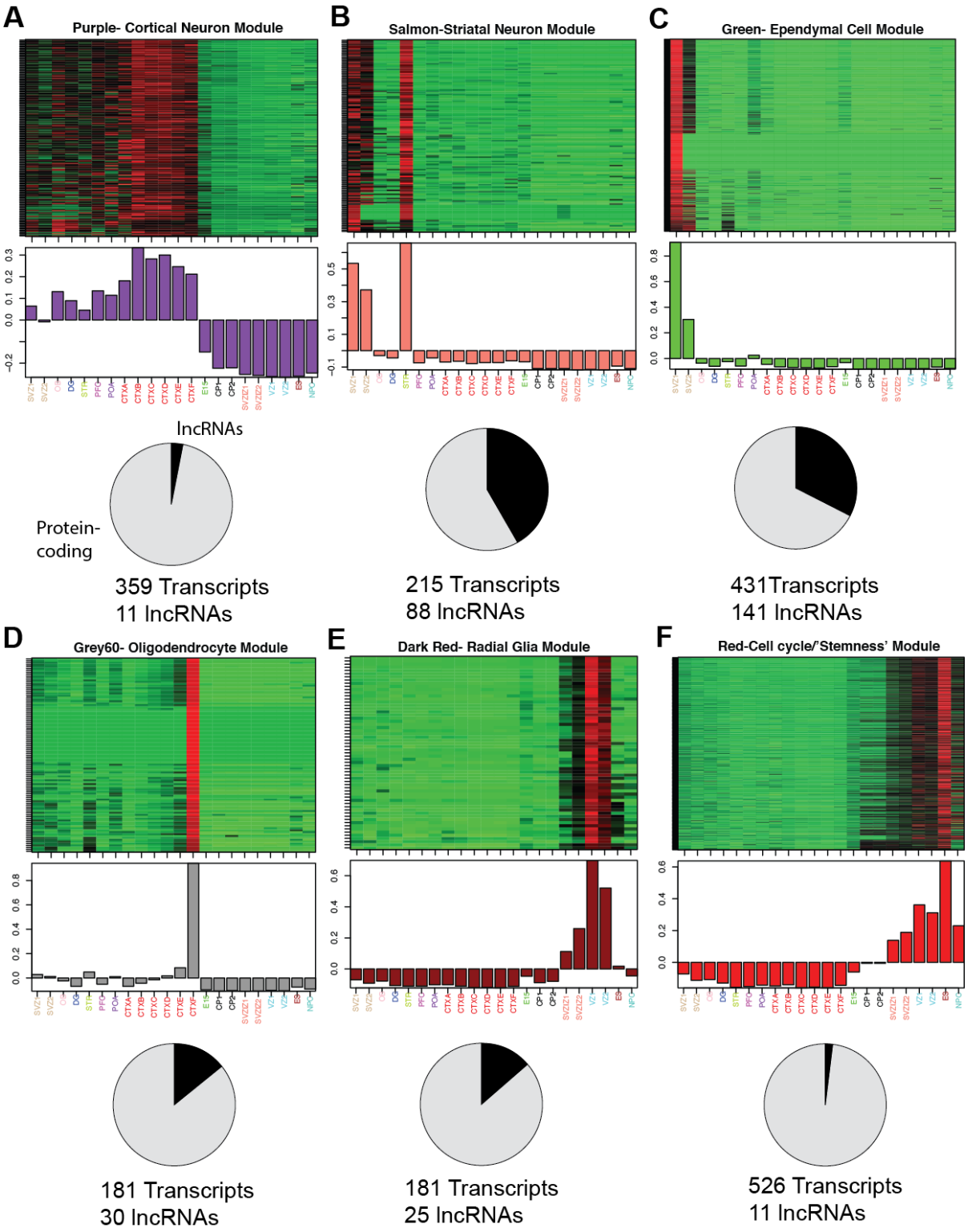
Outline of lncRNA catalog generation (A) Schematic of sagittal section of adult mouse brain. V-SVZ neural stem cells give rise to migratory neuroblasts (red). These neuroblasts travel along the rostral migratory stream (curved arrow) before terminally differentiating and integrating into olfactory bulb (OB) neuronal circuits. Numbered schematics correspond to coronal brain sections highlighting dissected regions (yellow) used for RNA collection. (B) Workflow for lncRNA catalog construction and characterization.

Figure 3.2


A
Adult
E14.5
DG
Str
SVZ
OB
CP
SVZ/IZ
VZ
N2B27
EGF/FGF
ESC
NPC
B
Relative Expression-All transcripts
0
2
SVZ
OB
DG
CTXA
CTXB
CTXC
CTXD
CTXE
CTXF
PFC
POA
SVZ/STR
STR
E15
CP1
CP2
SVZIZ1
SVZIZ2
VZ1
VZ2
ESC
NPC
C
Relative Expression-lncRNA
0
2
SVZ
OB
DG
CTXA
CTXB
CTXC
CTXD
CTXE
CTXF
SVZ/STR
STR
PFC
POA
E15
CP1
CP2
SVZIZ1
SVZIZ2
VZ1
VZ2
ESC
NPC

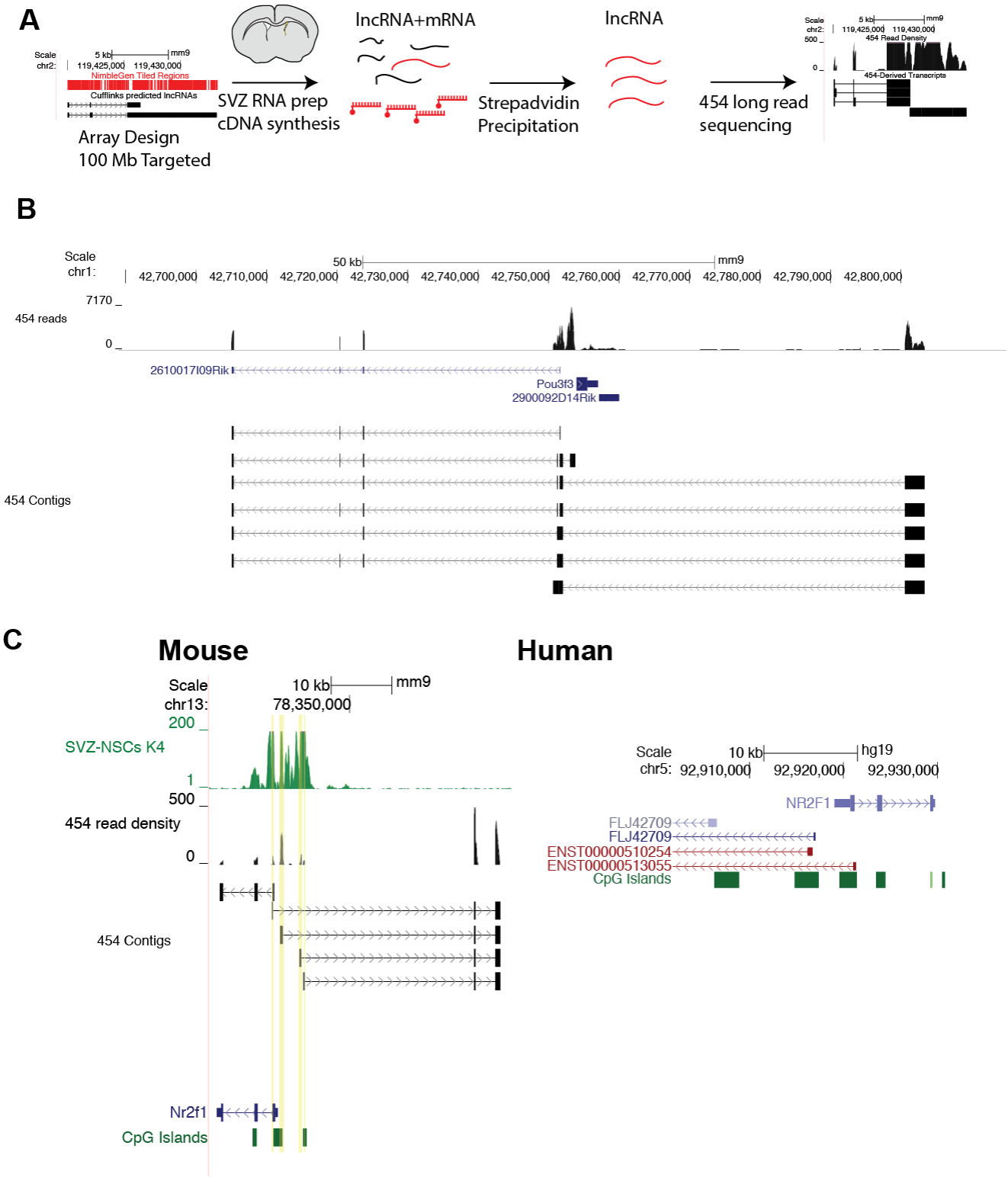
mRNAs and lncRNAs have temporally and spatially unique expression patterns (A) Schematic summarizing regions used for this analysis, colored in yellow. (B) Hierarchical clustering results of all transcripts expressed across all samples. (C) Hierarchical clustering results of lncRNAs expressed across all samples. The Pearson correlation coefficient was used as the distance metric. DG=dentate gyrus, STR=striatum, SVZ=subventricular zone, STR/SVZ=mixed dissection including both SVZ and striatal regions, OB= olfactory bulb, CTXA= cortical dissection, layer 2/3, CTXB= cortical dissection, layer 4, CTXC=cortical dissection, layer 5, CTXD=cortical dissection, layer 5, CTXE= cortical dissection, layer 6, CTXF= cortical dissection, layer 6b. POA= preoptic area, PFC= prefrontal cortex, E15= whole embryonic day 15 brain, VZ= ventricular zone of E 14.5 cortex, SVZ/IZ= subventricular zone/ intermediate zone of E14.5 cortex, CP= cortical plate of E14.5 cortex, ESC= cultured embryonic stem cells, NPCs= ESC-derived neural progenitor cells.

Figure 3.3


A
Purple- Cortical Neuron Module
B
Salmon-Striatal Neuron Module
C
Green- Ependymal Cell Module
lncRNAs
Protein-coding
359 Transcripts
11 lncRNAs
215 Transcripts
88 lncRNAs
431Transcripts
141 lncRNAs
D
Grey60- Oligodendrocyte Module
E
Dark Red- Radial Glia Module
F
Red-Cell cycle/'Stemness' Module
181 Transcripts
30 lncRNAs
181 Transcripts
25 lncRNAs
526 Transcripts
11 lncRNAs

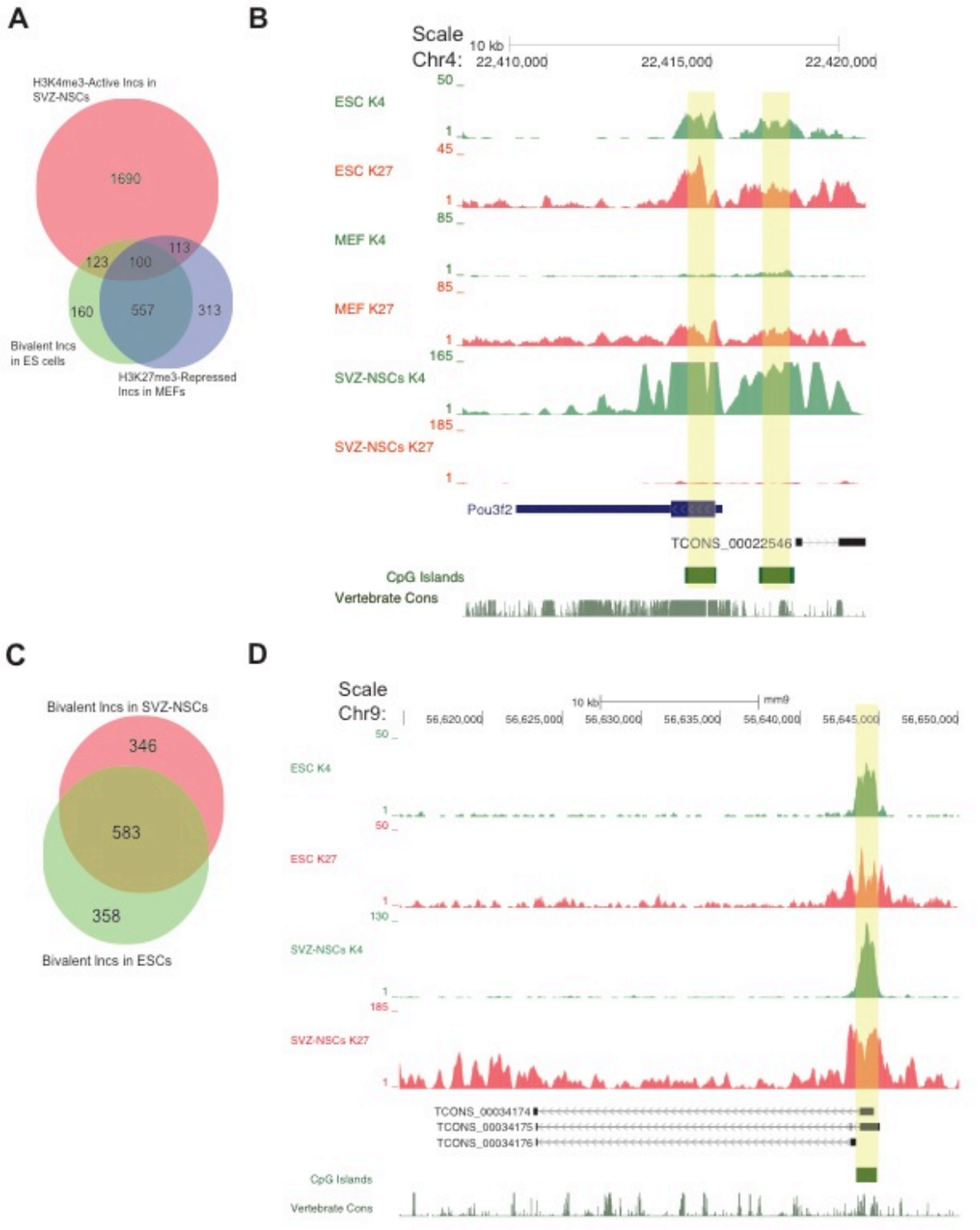
lncRNAs are associated with specific neural cell types and cellular processes (A-F) Top: heat maps depicting expression levels for six modules of co-expressed transcripts (rows) in 22 samples (columns) representing various brain regions and cell lines. Samples are labeled as in Fig. 3.2. Red, increased expression; black, neutral expression; green, decreased expression. Middle: barplots of the values of the module eigengenes (Horvath and Dong, 2008), which correspond to the first principal component obtained by singular value decomposition of each module. Modules were characterized by performing enrichment analysis with known gene sets (See Supplemental File 3). Bottom: pie charts indicating the abundance of lncRNAs within each module. Module members are defined as all transcripts that were positively correlated with the module eigengene at $P < 2.61e-08$.

Figure 3.4


A
Scale
chr2:
5 kb
mm9
119,425,000
119,430,000
NimbleGen Tiled Regions
Cufflinks predicted lncRNAs
Array Design
100 Mb Targeted
SVZ RNA prep
cDNA synthesis
lncRNA+mRNA
Strepadvidin
Precipitation
lncRNA
454 long read
sequencing
454 Read Density
454-Derived Tranecripts
B
Scale
chr1:
50 kb
mm9
42,700,000
42,710,000
42,720,000
42,730,000
42,740,000
42,750,000
42,760,000
42,770,000
42,780,000
42,790,000
42,800,000
454 reads
7170
0
2610017I09Rik
Pou3f3
2900092D14Rik
454 Contigs
C
Mouse
Human
Scale
chr13:
10 kb
mm9
78,350,000
200
SVZ-NSCs K4
1
500
454 read density
0
454 Contigs
Nr2f1
CpG Islands
Scale
chr5:
10 kb
hg19
92,910,000
92,920,000
92,930,000
NR2F1
FLJ42709
FLJ42709
ENST00000510254
ENST00000513055
CpG Islands

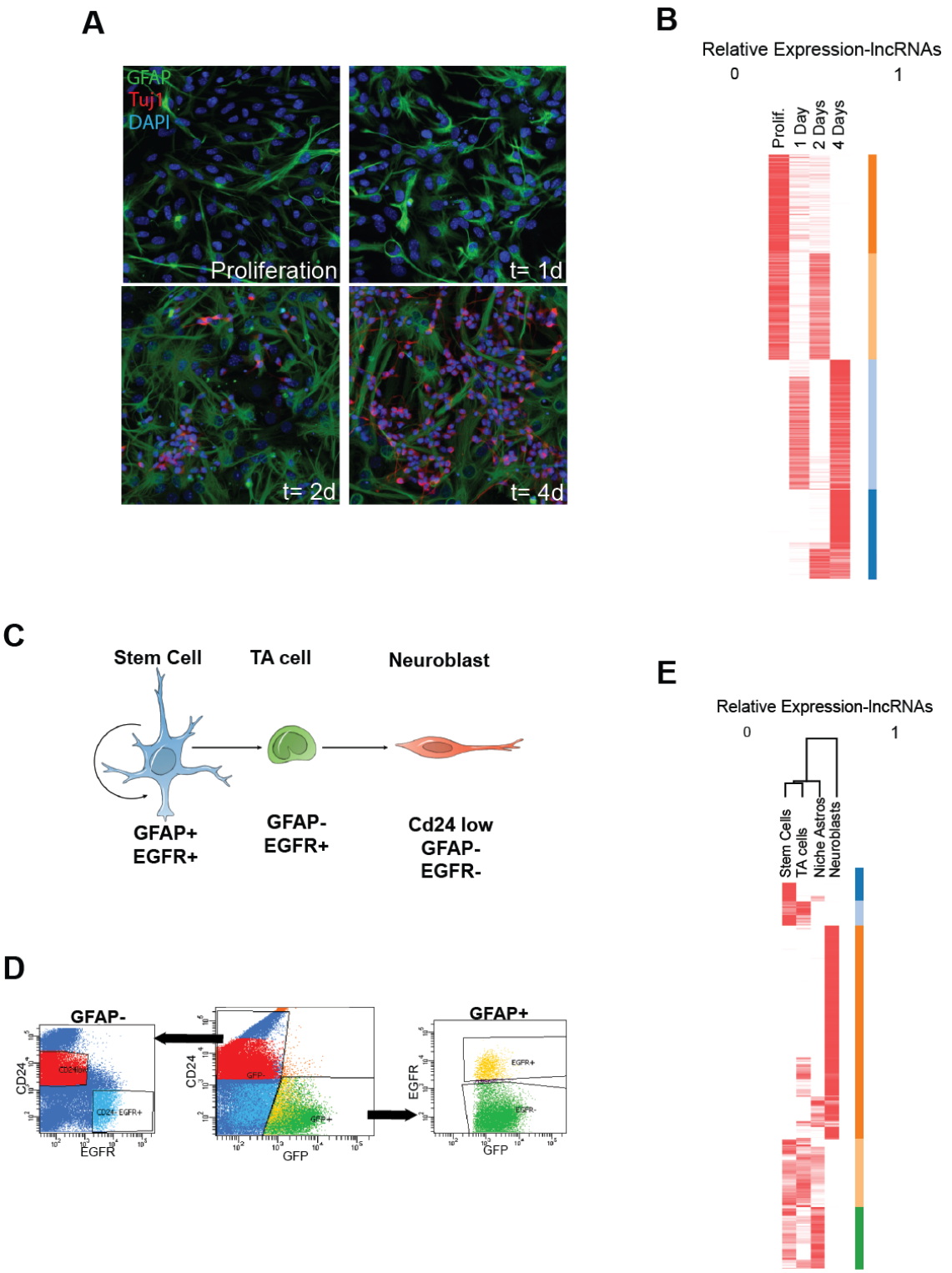
RNA CaptureSeq validates V-SVZ lncRNA expression and reveals multiple isoforms and complex locus structures (A) Schematic of RNA-Capture seq procedure. We used Cufflinlks' lncRNA assembly to define putative lncRNA loci and designed tiled probe libraries against these loci. The cDNA library was then hybridized to this biotin-labeled probe library, and after purification by streptavidin, the enriched population of lncRNAs was sequenced by 454 (Roche) long-read chemistry. (B) Isotigs assembled at the Pou3f3 locus revealed a distal transcriptional start site for a transcript that can be spliced into known noncoding RNA 2610017I09Rik. (C) CaptureSeq-derived reads correctly assembled known protein-coding gene Nr2f1 and identified 4 distinct TSS's for a lncRNA transcribed divergently from the Nr2f1 promoter. The syntentic region in human reveals a similar organization of CpG islands and divergent transcriptional start sites for non-coding transcripts. Genes derived from RefSeq are colored purple, genes from Ensembl are red.

Figure 3.5


A
H3K4me3-Active lncs in SVZ-NSCs
1690
113
123
100
160
557
313
Bivalent lncs in ES cells
H3K27me3-Repressed lncs in MEFs
B
Scale
Chr4:
10 kb
22,410,000
22,415,000
22,420,000
ESC K4
ESC K27
MEF K4
MEF K27
SVZ-NSCs K4
SVZ-NSCs K27
Pou3f2
TCONS_00022546
CpG Islands
Vertebrate Cons
C
Bivalent lncs in SVZ-NSCs
346
583
358
Bivalent lncs in ESCs
D
Scale
Chr9:
10 kb
mm9
56,620,000
56,625,000
56,630,000
56,635,000
56,640,000
56,645,000
56,650,000
ESC K4
ESC K27
SVZ-NSCs K4
SVZ-NSCs K27
TCONS_00034174
TCONS_00034175
TCONS_00034176
CpG Islands
Vertebrate Cons

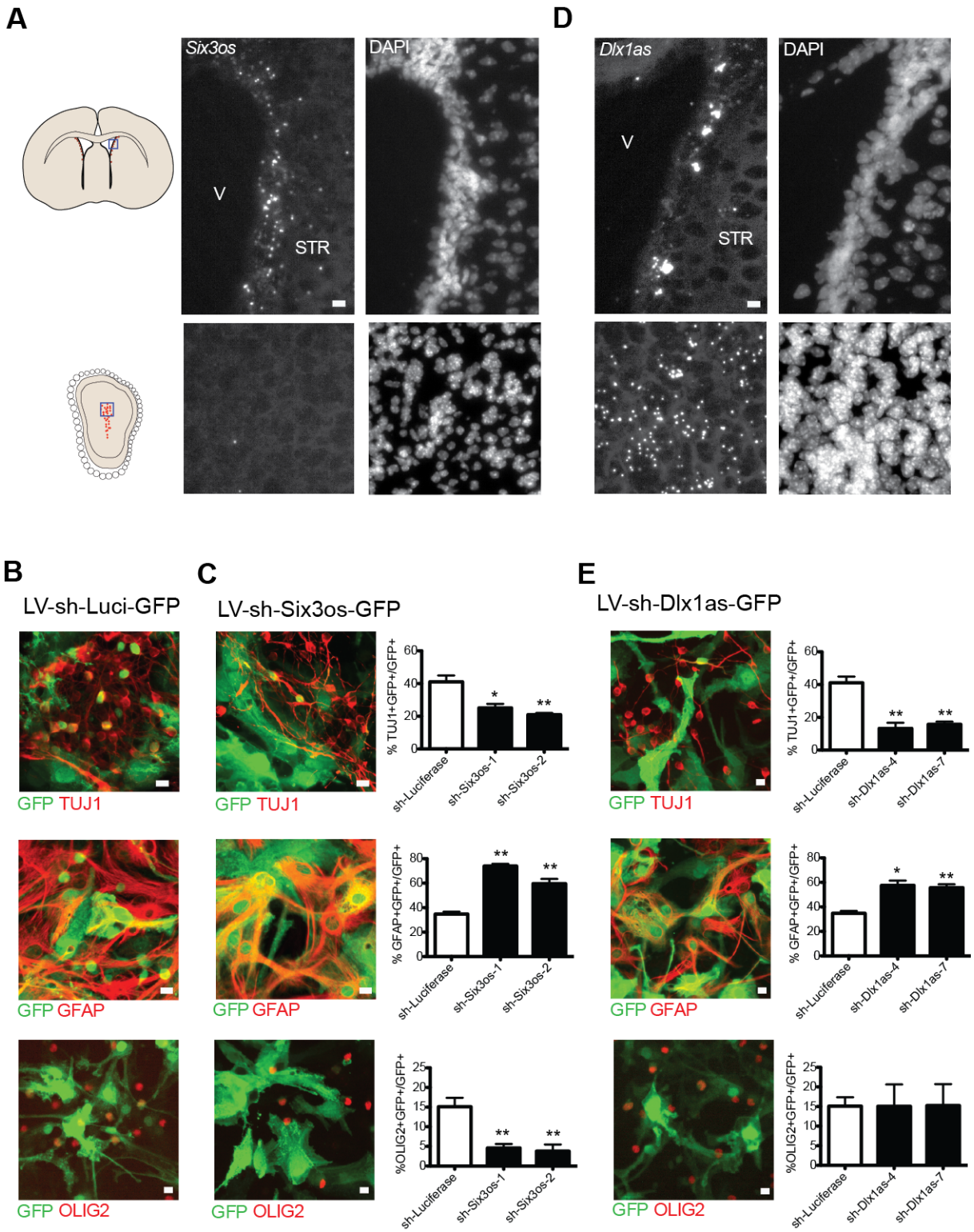
lncRNA loci can be bivalent in stem cell populations (A) Venn diagram highlighting lncRNAs that were bivalent in ESCs, monovalent H3K4me3 in SVZ-NSCs, and H3K27me3-repressed (monovalent or bivalent) in MEFs. (B) The *Pou3f2* promoter and the promoter (yellow boxes) of a nearby lncRNA demonstrated a similar pattern of histone modifications (bivalent in ESCs, repressed in MEFs, and activated in SVZ-NSCs). (C) Venn diagram demonstrating the number of lncRNAs that were bivalent in both ESCs and SVZ-NSCs. (D) A novel lncRNA locus ~50 kb downstream of protein-coding gene *Odf3l1*. The promoter was bivalent in both SVZ-NSCs and ESCs.

Figure 3.6


A
GFAP
Tuj1
DAPI
Proliferation
t= 1d
t= 2d
t= 4d
B
Relative Expression-lncRNAs
0
1
Prolif.
1 Day
2 Days
4 Days
C
Stem Cell
TA cell
Neuroblast
GFAP+
EGFR+
GFAP-
EGFR+
Cd24 low
GFAP-
EGFR-
D
GFAP-
GFAP+
CD24
EGFR
GFP
E
Relative Expression-lncRNAs
0
1
Stem Cells
TA cells
Niche Astros
Neuroblasts

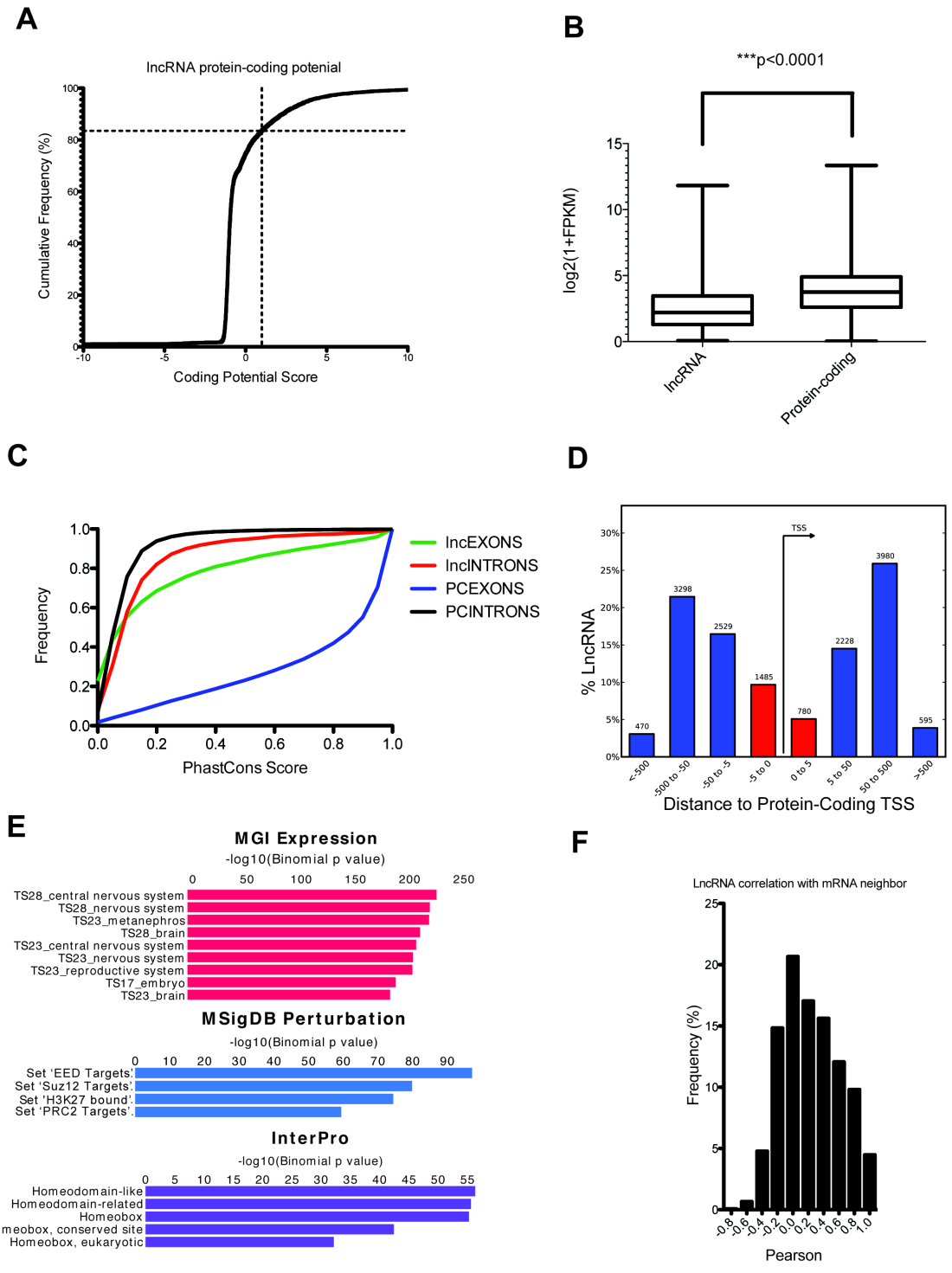
Analysis of lncRNA expression in the V-SVZ lineage *in vitro* and *in vivo* (A) Immunocytochemistry (ICC) of V-SVZ-NSC differentiation *in vitro*. In proliferation conditions, the culture is composed of neural precursor cells including GFAP+ (green) NSCs. After growth factor withdrawal, cells in these cultures differentiate into Tuj1+ neuroblasts (red, increasing numbers at 2d and 4d). (B) Heat map representing expression of lncRNAs that were changed > 4 fold from proliferation conditions to 4 d of differentiation. Color bars (orange, peach, light blue, dark blue) at the right represent gene clusters resulting from k-means clustering, k=4, Pearson distance metric. (C) Schematic of the V-SVZ lineage. GFAP+, EGFR+ stem cells (blue) give rise to transit amplifying (TA, green) cells. These TA cells give rise to Cd24+ low migratory neuroblasts (red). (D) FACS plots for isolation of the V-SVZ lineage. Cells were dissociated from freshly dissected V-SVZ tissue from the hGFAP-GFP mouse and stained with EGF conjugated to the A667 fluorophore and a CD24 antibody conjugated to PE. (E) Heatmap of lncRNAs differentially expressed throughout the V-SVZ lineage *in vivo*. Genes differentially expressed >2 fold between activated NSCs and neuroblasts were k-means clustered using the Pearson correlation metric, k=5. Color bars at the right (dark blue, light blue, orange, peach, green) represent gene clusters resulting from k-means clustering.

Figure 3.7


A
Six3os
DAPI
V
STR
D
Dlx1as
DAPI
V
STR
B
LV-sh-Luci-GFP
GFP TUJ1
GFP GFAP
GFP OLIG2
C
LV-sh-Six3os-GFP
GFP TUJ1
GFP GFAP
GFP OLIG2
% TUJ1+GFP+/GFP+
% GFAP+GFP+/GFP+
%OLIG2+GFP+/GFP+
sh-Luciferase
sh-Six3os-1
sh-Six3os-2
E
LV-sh-Dlx1as-GFP
GFP TUJ1
GFP GFAP
GFP OLIG2
% TUJ1+GFP+/GFP+
% GFAP+GFP+/GFP+
%OLIG2+GFP+/GFP+
sh-Luciferase
sh-Dlx1as-4
sh-Dlx1as-7

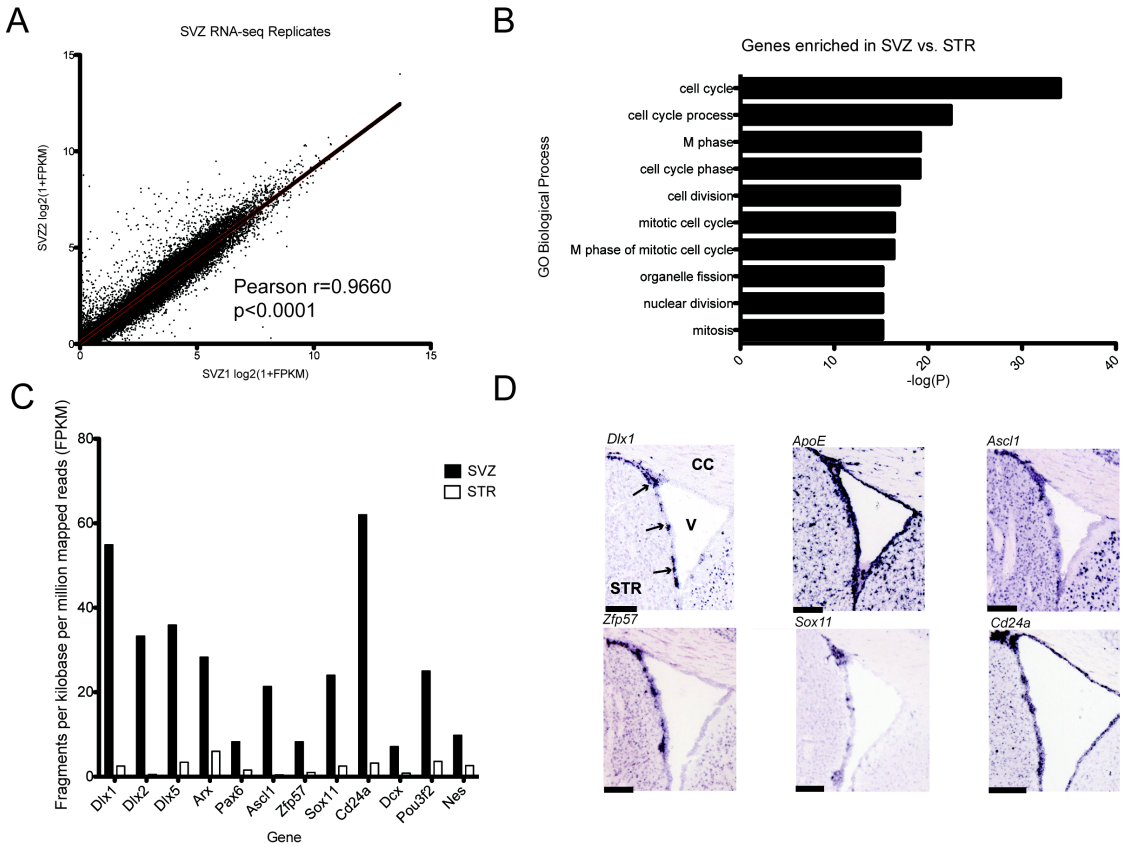
Functional validation of lncRNA candidates (A) *In situ* hybridization (ISH) for *Six3os* using branched DNA probes. Positive signal is revealed by Fast Red alkaline phosphatase substrates, which appears as highly fluorescent, punctate deposits (left panels); DAPI nuclear counterstain is shown at the right. Blue box in SVZ and OB coronal schematics at left indicate regions shown at right. Scale bars= 10 μm**.** V=ventricle, STR= striatum. (B) Control (LV-sh-Luci-GFP) lentiviral infections in V-SVZ-NSC cultures after 7 days of differentiation. Top, immunocytochemistry (ICC) for Tuj1 (red) and GFP (green), Middle, ICC for GFAP (red) and GFP (green), Bottom, ICC for OLIG2 (red) and GFP (green). (C) Analysis of *Six3os* knockdown in V-SVZ-NSCs after 7 days of differentiation. Two different constructs were used (sh-Six3os-1, sh-Six3os-2). Top, immunocytochemistry (ICC) for Tuj1 (red) and GFP (green), Middle, ICC for GFAP (red) and GFP (green), Bottom, ICC for OLIG2 (red) and GFP (green) after infection with control vector expressing shRNAs targeting *Six3os* (LV-sh-*Six3os*-GFP). Quantification of data is presented at right. Scale bars= 10 μm. Error bars= SEM, 5-6 replicates for control group, 2-3 per experimental group. *p<.05, **p<0.01, compared to sh-Luci, two-tailed t-test. (D) ISH with branched DNA probes for *Dlx1as* in the V-SVZ (top) and OB (bottom). Scale bars= 10 μm**.** V=ventricle, STR= striatum. (E) Analysis of *Dlx1as* knockdown after 7 days of differentiation. Two unique targeting sequences were used (sh-Dlx1as-4, sh-Dlx1as-7). Top, immunocytochemistry (ICC) for Tuj1 (red) and GFP (green), Middle, ICC for GFAP (red) and GFP (green), Bottom, ICC for OLIG2 (red) and GFP (green). Quantification of of data is presented at right. Scale bars= 10 μm. Error bars= SEM, 5-6 replicates for control group, 3 per experimental group. *p<.05, **p<0.01, compared to sh-Luci, two-tailed t-test.

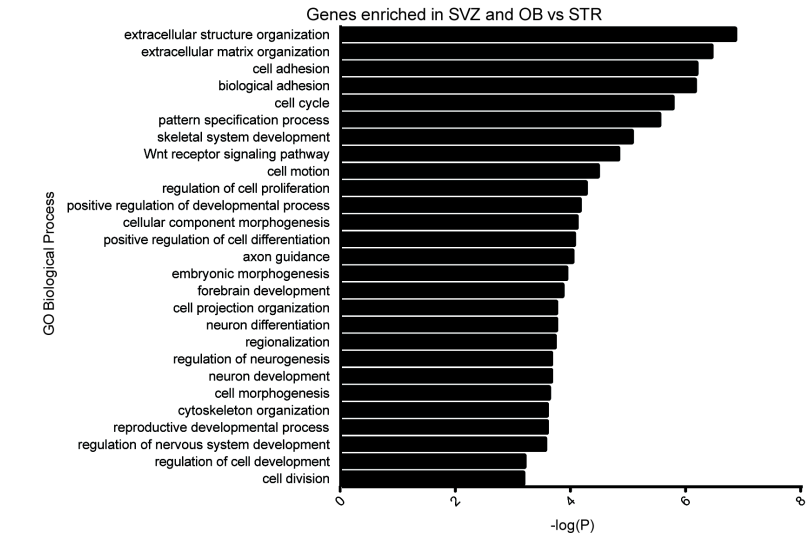
Figure S3.1


A
lncRNA protein-coding potenial
Cumulative Frequency (%)
Coding Potential Score
B
***p<0.0001
log2(1+FPKM)
lncRNA
Protein-coding
C
Frequency
PhastCons Score
lncEXONS
lncINTRONS
PCEXONS
PCINTRONS
D
% LncRNA
TSS
470
3298
2529
1485
780
2228
3980
595
<-500
-500 to -50
-50 to -5
-5 to 0
0 to 5
5 to 50
50 to 500
>500
Distance to Protein-Coding TSS
E
MGI Expression
-log10(Binomial p value)
TS28_central nervous system
TS28_nervous system
TS23_metanephros
TS28_brain
TS23_central nervous system
TS23_nervous system
TS23_reproductive system
TS17_embryo
TS23_brain
MSigDB Perturbation
-log10(Binomial p value)
Set 'EED Targets'.
Set 'Suz12 Targets'.
Set 'H3K27 bound'.
Set 'PRC2 Targets'.
InterPro
-log10(Binomial p value)
Homeodomain-like
Homeodomain-related
Homeobox
meobox, conserved site
Homeobox, eukaryotic
F
LncRNA correlation with mRNA neighbor
Frequency (%)
Pearson

Characteristics of catalogued lncRNAs (A) Plot of coding potential calculator scores of lncRNAs. 82% (horizontal dotted line) of our lncRNA catalogue were noncoding, or have weak coding potential (vertical dotted line, CPC=1). (B) Box and whisker plot of lncRNA and protein-coding gene FPKM values. lncRNAs had lower expression levels than protein-coding genes. P<0.0001, Mann Whitney U. (C) Chart of cumulative frequency of phastcons scores for lncRNA exons (green), lncRNA introns (red), protein-coding exons (blue), and protein-coding introns (black). (D) Histogram of distance of lncRNA TSS to protein-coding TSS. In red are those lncRNAs within 5kb of a protein-coding TSS. (E) Gene ontology of protein-coding genes that have lncRNAs transcribed proximal (<5kb) to their promoter. (F) Frequency distribution histogram of Pearson correlation coefficients of lncRNA-mRNA expression levels across 22 samples (defined in **3.2**).

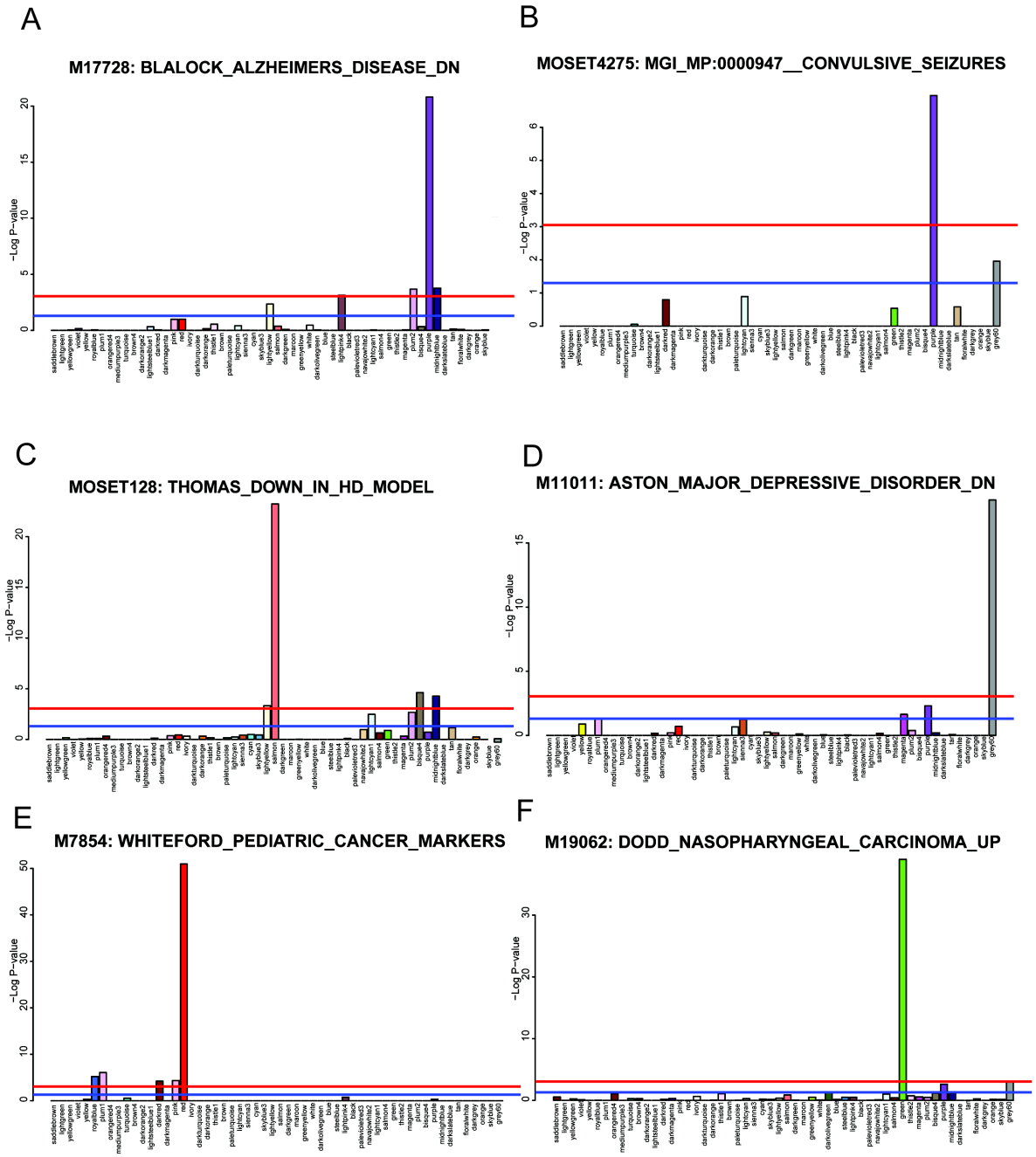
Figure S3.2


A
SVZ RNA-seq Replicates
SVZ2 log2(1+FPKM)
SVZ1 log2(1+FPKM)
Pearson r=0.9660
p<0.0001
B
Genes enriched in SVZ vs. STR
GO Biological Process
cell cycle
cell cycle process
M phase
cell cycle phase
cell division
mitotic cell cycle
M phase of mitotic cell cycle
organelle fission
nuclear division
mitosis
-log(P)
C
Fragments per kilobase per million mapped reads (FPKM)
SVZ
STR
Dlx1
Dlx2
Dlx5
Arx
Pax6
Ascl1
Zfp57
Sox11
Cd24a
Dcx
Pou3f2
Nes
Gene
D
Dlx1
CC
V
STR
ApoE
Ascl1
Zfp57
Sox11
Cd24a


E
Genes enriched in SVZ and OB vs STR
GO Biological Process
extracellular structure organization
extracellular matrix organization
cell adhesion
biological adhesion
cell cycle
pattern specification process
skeletal system development
Wnt receptor signaling pathway
cell motion
regulation of cell proliferation
positive regulation of developmental process
cellular component morphogenesis
positive regulation of cell differentiation
axon guidance
embryonic morphogenesis
forebrain development
cell projection organization
neuron differentiation
regionalization
regulation of neurogenesis
neuron development
cell morphogenesis
cytoskeleton organization
reproductive developmental process
regulation of nervous system development
regulation of cell development
cell division
-log(P)

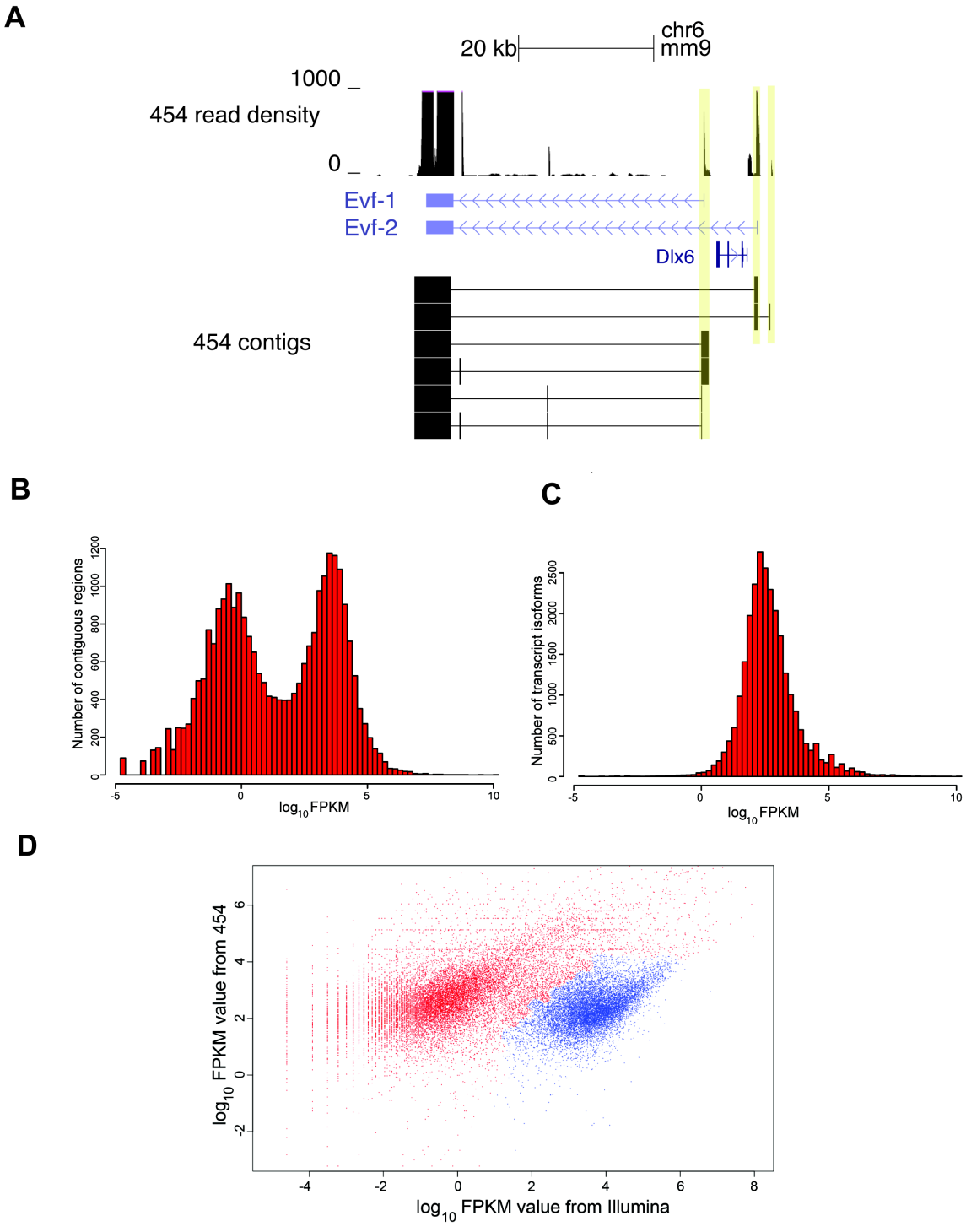
The V-SVZ and OB have gene expression profiles distinct form the neighboring non-neurogenic striatum (A) Scatter plot of SVZ RNA seq replicates. SVZ1 and SVZ2 were each dissected from 5 mice at different sessions. (B) Gene ontology analysis of protein coding genes differentially expressed >2 fold in SVZ vs. striatum (STR). (C) Selection of FPKM values for genes significantly enriched in SVZ vs. striatum and known to have a role in neurogenesis and/or enriched expression in the SVZ. (D) *In situ* hybridizations from the Allen Brain Atlas for a selection of genes that were enriched in the SVZ/OB vs. STR. SVZ region is indicated by arrows in the first panel. V, ventricle; CC, corpus callosum. Scale bar = 163 μm for *Dlx1,* 105 μm for *ApoE, Ascl1, Zfp57*, *Sox11*, and *ApoE*, and *Cd24a*. (E) Gene ontology analysis of protein coding genes that were enriched (>2 fold) in both the SVZ and OB as compared to the STR.

Figure S3.3


A
M17728: BLALOCK_ALZHEIMERS_DISEASE_DN
−Log P−value
B
MOSET4275: MGI_MP:0000947__CONVULSIVE_SEIZURES
−Log P−value
C
MOSET128: THOMAS_DOWN_IN_HD_MODEL
−Log P−value
D
M11011: ASTON_MAJOR_DEPRESSIVE_DISORDER_DN
−Log P−value
E
M7854: WHITEFORD_PEDIATRIC_CANCER_MARKERS
−Log P−value
F
M19062: DODD_NASOPHARYNGEAL_CARCINOMA_UP
−Log P−value

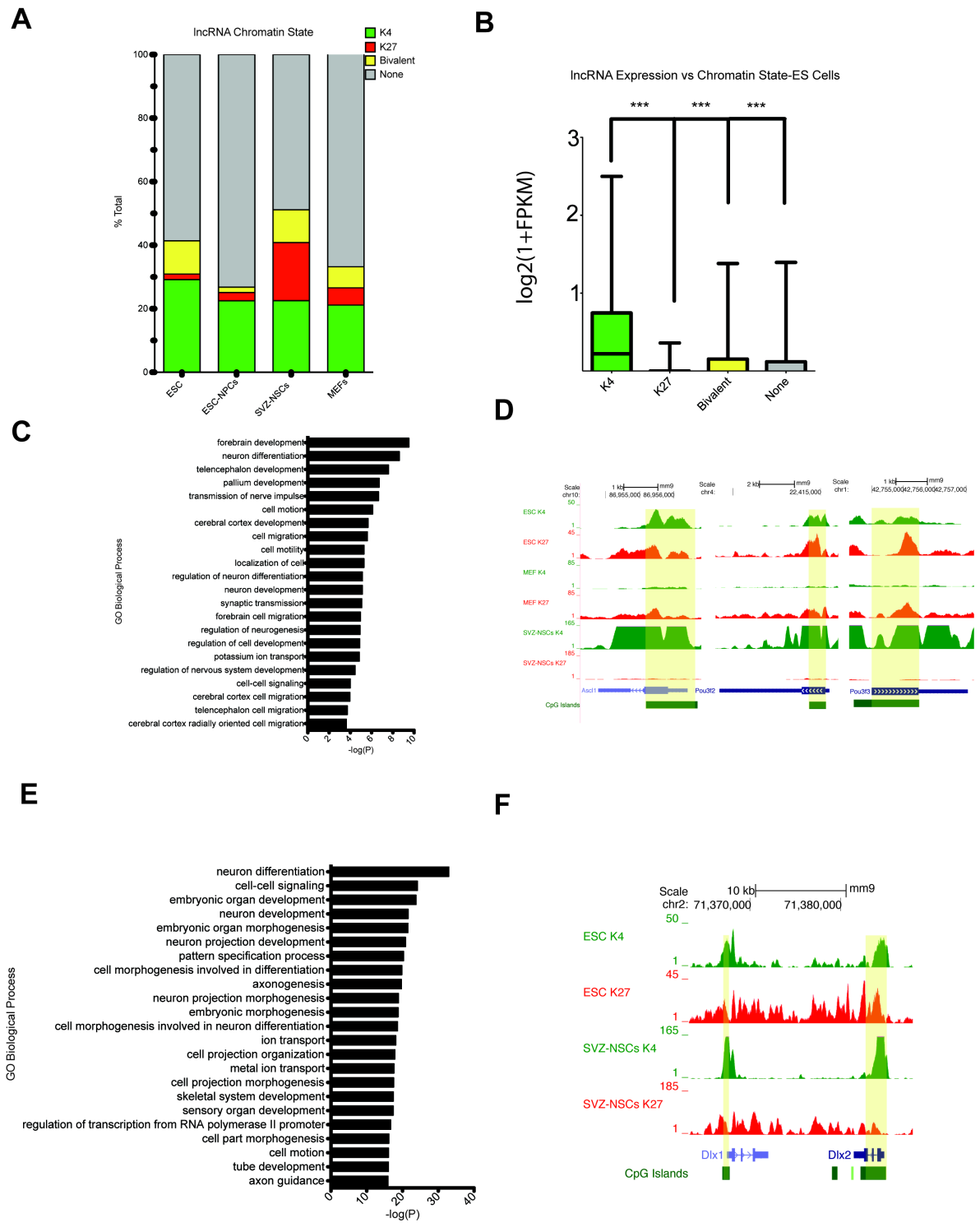
Specific gene co-expression modules associate with with distinct disease states Module composition was cross-referenced with curated gene sets from the Molecular Signatures Database (Subramanian et al., 2005) (A,D,E,F); the Mouse Genome Informatics database (Bult et al., 2008; Zhang et al., 2010) (B), a study of an *in vivo* mouse model of Huntington's disease (Thomas et al., 2011) (C); and others (**Supplementary File 3**). Bars indicate the significance of enrichment of the gene set in each module (Fisher's exact test). Each module was defined as consisting of all transcripts that were positively correlated with the module eigengene at P < 2.61e-08 (Experimental Procedures). Blue line corresponds to P = .05 and red line corresponds to a Bonferroni-corrected P-value based on the number of modules (n = 56).

Figure S3.4


A
20 kb
chr6
mm9
1000
454 read density
0
Evf-1
Evf-2
Dlx6
454 contigs
B
Number of contiguous regions
log10FPKM
C
Number of transcript isoforms
log10FPKM
D
log10 FPKM value from 454
log10 FPKM value from Illumina

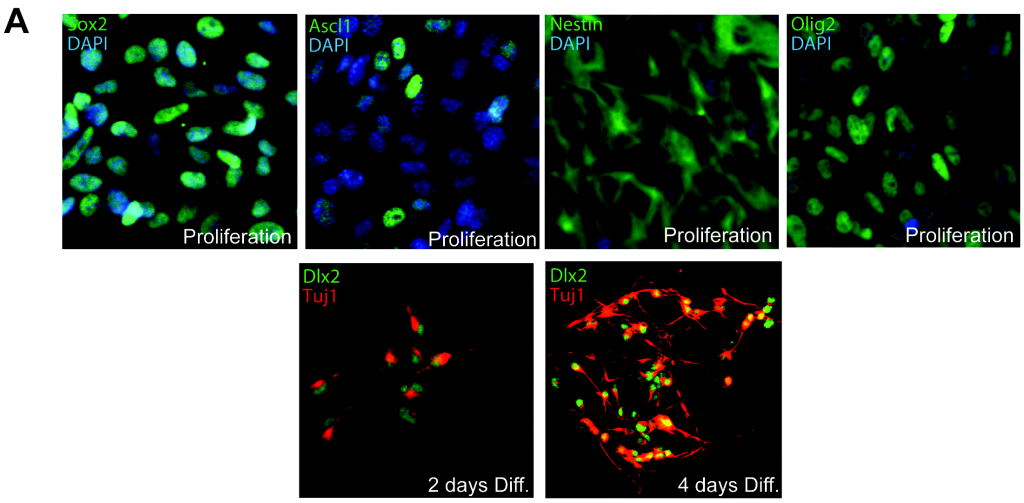
RNA-capture seq analysis (A) Example of RNA CaptureSeq reconstruction of known and novel lncRNA isoforms. Shown are the 454 read density plots and assembled contigs of *Evf*-1 and *Evf-2*. Yellow boxes highlight the TSS. (B) Histogram of $\log_{10}$ FPKM values for the Illumina reads mapped to Capture-Seq derived contigs. (C) Histogram of $\log_{10}$ FPKM values for the 454 reads mapped to Capture-Seq derived contigs. (D) Graph of ordered pairs of the 26,279 transcript isoforms with nonzero FPKM values for both the Illumina and 454 data. Here, the two peaks in (B) take the form of two ellipsoidal clusters of points. Putative lncRNA transcripts are mostly associated with the red points. The Pearson correlation coefficient between the 454 and Illumina FPKM values for reads associated with putative lncRNA transcripts (red points) is 0.613. For reads associated with putative protein-coding transcripts (blue points), the analogous correlation coefficient is 0.523. See also Chapter 2, **454 Alignment and Comparison to Illumina.**

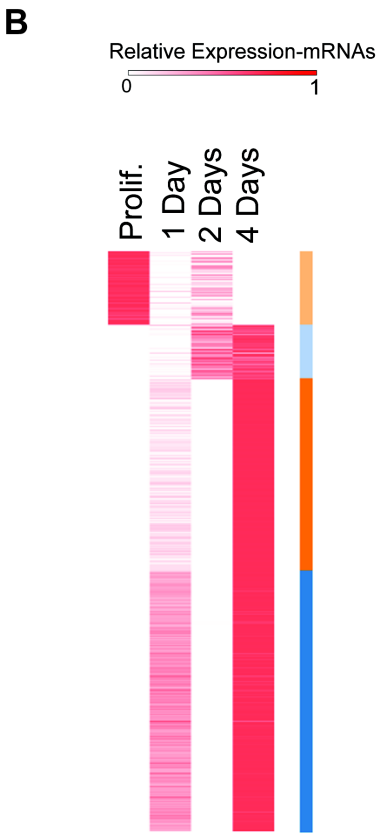
Figure S3.5

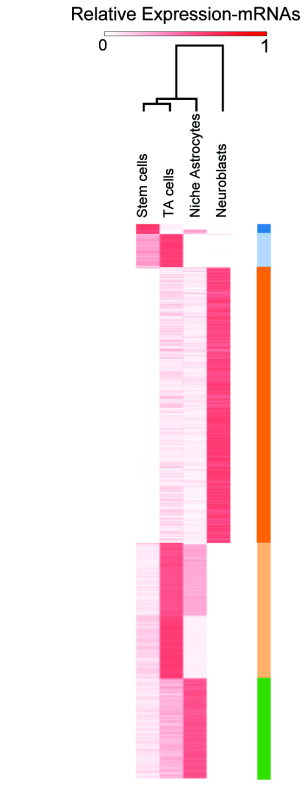

A
lncRNA Chromatin State
K4
K27
Bivalent
None
% Total
ESC
ESC-NPCs
SVZ-NSCs
MEFs
B
lncRNA Expression vs Chromatin State-ES Cells
***
log2(1+FPKM)
K4
K27
Bivalent
None
C
GO Biological Process
forebrain development
neuron differentiation
telencephalon development
pallium development
transmission of nerve impulse
cell motion
cerebral cortex development
cell migration
cell motility
localization of cell
regulation of neuron differentiation
neuron development
synaptic transmission
forebrain cell migration
regulation of neurogenesis
regulation of cell development
potassium ion transport
regulation of nervous system development
cell-cell signaling
cerebral cortex cell migration
telencephalon cell migration
cerebral cortex radially oriented cell migration
-log(P)
D
ESC K4
ESC K27
MEF K4
MEF K27
SVZ-NSCs K4
SVZ-NSCs K27
CpG Islands
E
GO Biological Process
neuron differentiation
cell-cell signaling
embryonic organ development
neuron development
embryonic organ morphogenesis
neuron projection development
pattern specification process
cell morphogenesis involved in differentiation
axonogenesis
neuron projection morphogenesis
embryonic morphogenesis
cell morphogenesis involved in neuron differentiation
ion transport
cell projection organization
metal ion transport
cell projection morphogenesis
skeletal system development
sensory organ development
regulation of transcription from RNA polymerase II promoter
cell part morphogenesis
cell motion
tube development
axon guidance
-log(P)
F
Scale
10 kb
mm9
chr2: 71,370,000
71,380,000
ESC K4
ESC K27
SVZ-NSCs K4
SVZ-NSCs K27
Dlx1
Dlx2
CpG Islands

Chromatin signatures of lncRNA and mRNA loci (A) Summary of the chromatin modifications of lncRNAs across 4 cell types: embryonic stem cells (ESCs), ESC-derived neural progenitors (NPCs), SVZ neural stem cells (SVZ-NSCs), and mouse embryonic fibroblasts (MEFs). (B) Box and whisker plot showing lncRNA expression levels in ESCs related to the chromatin state of the lncRNA promoter region (***p<.0001, Mann-Whitney U). (C) Gene ontology analysis of 302 protein-coding genes that are bivalent in ESCs, H3K4me3-monovalent in SVZ-NSCs, and H3K27me3-repressed in MEFs. (D) Example of 3 neurogenic transcription factors that were bivalent in ESCs, become repressed in MEFs and activated in SVZ-NSCs. The yellow boxes highlight the promoter regions. (E) Gene ontology analysis of 3109 protein-coding genes that are bivalent in ESCs, and SVZ-NSCs (E) The neurogenic *Dlx1/2* locus was bivalent in both ESCs and SVZ-NSCs.

Figure S3.6

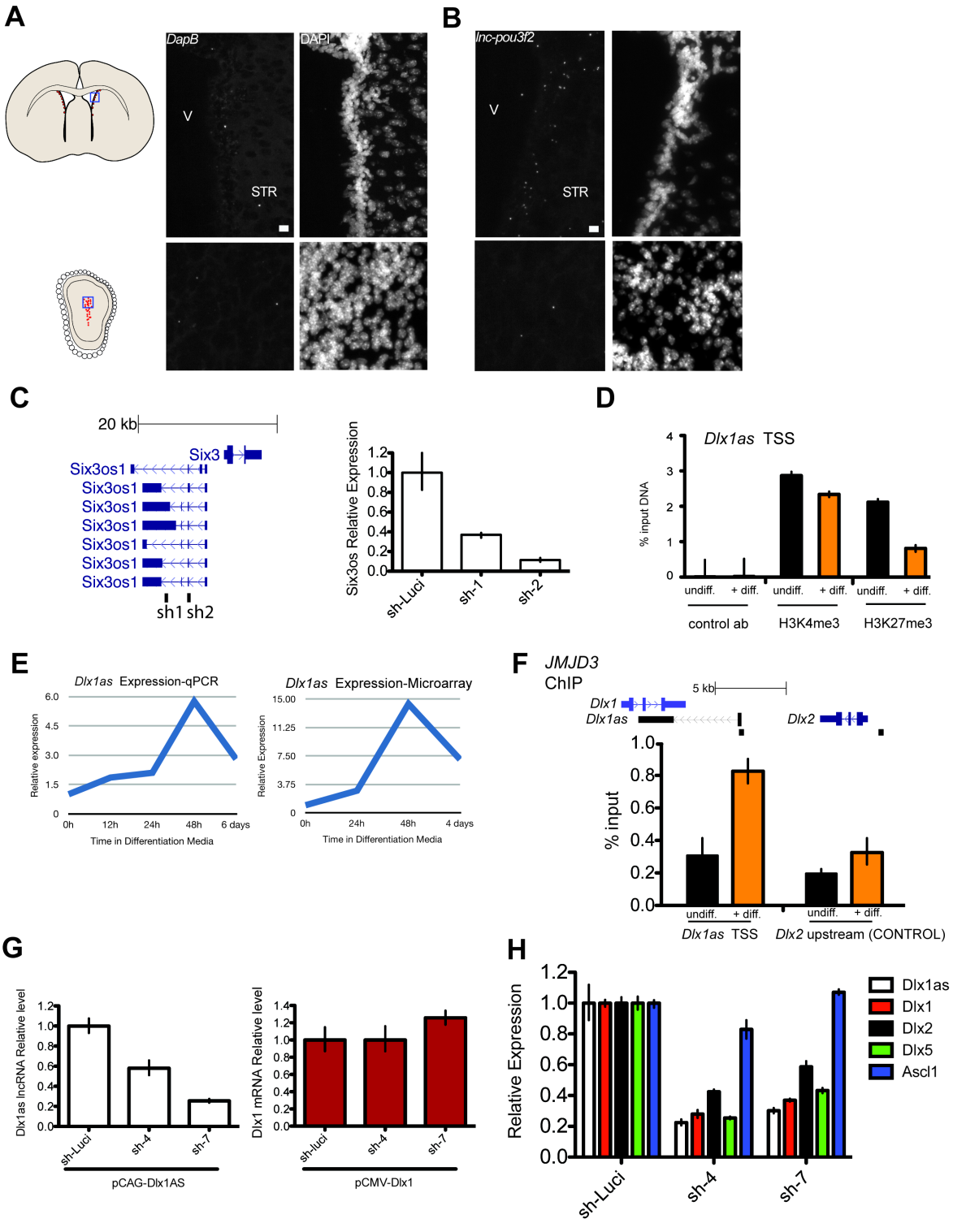

A
Sox2
DAPI
Proliferation
Ascl1
DAPI
Proliferation
Nestin
DAPI
Proliferation
Olig2
DAPI
Proliferation
Dlx2
Tuj1
2 days Diff.
Dlx2
Tuj1
4 days Diff.


B
Relative Expression-mRNAs
0
1
Prolif.
1 Day
2 Days
4 Days


C
Relative Expression-mRNAs
0
1
Stem cells
TA cells
Niche Astrocytes
Neuroblasts

Dynamic changes in mRNA expression during differentiation *in vitro* and *in vivo* (A) Immunocytochemistry for SOX2, ASCL1, NESTIN, OLIG2 (green) in proliferating cultures, and DLX2 (green) and Tuj1 (red) over a differentiation time course. (B) Heat map representing expression of mRNAs that were changed > 4 fold from proliferation conditions to 4 d of differentiation. Color bars (orange, peach, light blue, dark blue) at the right represent gene clusters resulting from k-means clustering, k=4, Pearson distance metric. (C) Heatmap of protein-coding genes differentially expressed throughout the V-SVZ lineage *in vivo*. Genes differentially expressed >2 fold between activated NSCs and neuroblasts are shown. Color bars (dark blue, light blue, orange, peach, green) at the right represent gene clusters resulting from k-means clustering, k=5, Pearson distance metric. Cell types were hierarchically clustered using Pearson correlation metric and complete linkage.

Figure S3.7


A
DapB
DAPI
V
STR
B
lnc-pou3f2
V
STR
C
20 kb
Six3
Six3os1
sh1 sh2
Six3os Relative Expression
sh-Luci
sh-1
sh-2
D
Dlx1as TSS
% input DNA
undiff.
+ diff.
control ab
H3K4me3
H3K27me3
E
Dlx1as Expression-qPCR
Relative expression
Time in Differentiation Media
Dlx1as Expression-Microarray
Relative Expression
Time in Differentiation Media
F
JMJD3
ChIP
5 kb
Dlx1
Dlx1as
Dlx2
% input
Dlx1as TSS
Dlx2 upstream (CONTROL)
G
Dlx1as lncRNA Relative level
sh-Luci
sh-4
sh-7
pCAG-Dlx1AS
Dlx1 mRNA Relative level
pCMV-Dlx1
H
Relative Expression
Dlx1as
Dlx1
Dlx2
Dlx5
Ascl1

Characterization of the expression and function of candidate lncRNAs (A) *In situ* hybridization (ISH) for negative control bacterial transcript DapB using branched DNA probes (left panels); DAPI nuclear counterstain is shown at the right. Blue box in V-SVZ and OB coronal schematics at left indicate regions shown at right. (B) ISH with branched DNA probes for *lnc-pou3f2* in the V-SVZ (top) and OB (bottom). (C) Left: Genome browser schematic of the *Six3os* locus. Black boxes indicate approximate target site of shRNAs. Right: knockdown efficiency of *Six3os* in lentiviral-infected V-SVZ-NSCs in proliferation conditions, measured by RT-qPCR. V-SVZ-NSCs were infected with *Six3os* knockdown constructs or control luciferase knockdown construct and GFP+ cells were isolated by FACS. Error bars are standard deviation propagated by the least-squares method. (D) ChIP-qPCR analysis of H3K4me3 and H3K27me3 levels at the *Dlx1as* TSS. V-SVZ-NSC cultures in both proliferating (undiff.) conditions and after 30 hours differentiation (+diff.) were assayed. Anti-Gal4 was the negative control antibody. Error bars represent standard deviation from three replicates. (E) Relative Expression of *Dlx1as* as measured by RT-qPCR (left) and expression microarray (right). All values are relative to expression in proliferation media. (F) Top: Schematic of the *Dlx1/2* locus indicating the location of *Dlx1as* TSS and *Dlx2* upstream (control) primers (Black boxes). Bottom: ChIP-qPCR analysis of JMJD3 levels at both the *Dlx1as* TSS and a control site 1 kb upstream of *Dlx2*. V-SVZ-NSC cultures were assayed in proliferating (undiff.) conditions and after 30 hours differentiation (+diff). Error bars represent standard deviation from three replicates. (G) Knockdown efficiencies for *Dlx1as* and *Dlx1* by *Dlx1as*-specific shRNAs. 293T cells were co-transfected with a *Dlx1as* overexpression construct (pCAG-*Dlx1as*) + shRNA constructs or *Dlx1* overexpression construct (pCMV-*Dlx1*) + shRNA constructs, and expression levels of *Dlx1as* or *Dlx1* were measured by RT-qPCR. Error bars are standard deviation propagated by the least-squares method. (H) Gene expression changes of neurogenic transcription factors upon *Dlx1as* knockdown. V-SVZ-NSCs infected with *Dlx1as* knockdown construct or control luciferase knockdown construct were switched to differentiation media for two days. GFP+ cells were

isolated by FACS, and RNA extracted for RT-qPCR. Error bars are standard deviation propagated by the least-squares method.

SUPPLEMENTARY FILES

All supplementary files are available from:
http://www.sciencedirect.com/science/article/pii/S1934590913000982

Supplementary File 1: UCSC Genome Browser BED file of identified lncRNAs, related to Figure 3.1: Allows for visualization of all lncRNAs catalogued for this study. Coordinates are indicated in genome build mm9.

Supplementary File 2: Summary of sequencing data used in this paper, related to 3.2: Excel spreadsheet that describes reads generated in this study as well as reads downloaded from other studies, used for (1) RNA-seq and (2) ChIP-seq.

Supplementary File 3: Module analysis and gene set enrichment analysis, related to Figure 3.3: Excel workbook contains three spreadsheets: (1) table of kme values for coding and noncoding transcripts, module membership, and associated p values (2) Enrichment analysis of known genesets curated from multiple published sources. (3) Enrichment analysis of genesets from the MolSigDB. See **Supplementary Methods** and **Figure 3.3.**

Supplementary File 4: Microarray expression data, related to figure 6: An Excel workbook with four spreadsheets: (1) Normalized expression values and fold changes across the *in vitro* differentiation assay timepoints described in **Figure 6** run on the Nimblegen Mouse Gene Expression 12x135K Array (2) Normalized expression values and fold changes across the *in vitro* differentiation assay timepoints described in **Figure 7** run on our custom lncRNA array. P-vaues are reported from Student's t-test, using the Benjamini-Hochberg correction for multiple

testing. (3) Normalized expression values and fold changes across FACS isolated cell populations run on the Nimblegen Mouse Gene Expression 12x135K Array, see also **Figure S6** (4) Normalized expression values and fold changes across FACS isolated cell populations run on our custom lncRNA array. P-vaues are reported from Student's t-test, using the Benjamini-Hochberg correction for multiple testing.

Supplementary File 5: V-SVZ lineage cell type specific lncRNAs, related to figure 7: An Excel workbook with 3 spreadsheets: (1) lncRNAs defined to be "Activated NSCs" specific. These were derived by filtering for lncRNA transcripts bivalent in ESCs, repressed in MEFs, H3K4me3-marked in V-SVZ-NSCs, and downregulated during differentiation into neuroblasts *in vivo*. (2) lncRNAs defined to be Transit Amplifying cell specific. These are transcripts bivalent in ESCs, repressed in MEFs, H3K4me3-marked in V-SVZ-NSCs, and upregulated in TA cells *in vivo*. (3) lncRNAs defined to be Neuroblast specific. These are bivalent in both ESCs and V-SVZ-NSCs, and are upregulated in neuroblasts *in vivo*.

Chapter 4: The long noncoding RNA Pinky regulates neurogenesis from embryonic and postnatal brain neural stem cells

Summary

During neurogenesis, neural stem cells (NSCs) in both the adult and embryonic brain give rise to intermediate progenitor cells, a transit-amplifying population that divides once or more to produce neurons. Here, we identify *Pinky* (*Pnky*) as a long non-coding RNA (lncRNA) that regulates the transition between NSCs and neurogenic progenitors. *Pnky* is expressed in NSC populations and becomes downregulated during neurogenesis. shRNA-mediated knockdown of *Pnky* expands the pool of neurogenic progenitors, increasing neurogenesis. *Pinky* is conserved and expressed in NSCs of the developing mouse and human cortex. In embryonic mouse brain, *Pnky* knockdown accelerates neurogenesis and depletes this embryonic NSC population. *Pnky* physically interacts with PTBP1, a regulator of mRNA splicing, and *Pnky* depletion leads to the differential splicing of PTBP1 targets. These data thus identify *Pinky* as a conserved lncRNA that regulates a critical stage of neurogenesis from NSCs in the adult and embryonic brain.

Introduction

The mammalian genome transcribes many thousands of lncRNAs – transcripts >200 nucleotides long that have no evidence of protein coding potential – and it is now clear that lncRNAs can have critical biological functions (Batista and Chang, 2013; Lee, 2012; Mercer and Mattick, 2013; Rinn and Chang, 2012). A number of lncRNAs have been described that play a role in neurogenesis: Deletion of *lnc-brn1b* leads to a loss of outer-layer neurons and loss of progenitors in the developing cortex (Sauvageau et al., 2013). Deletion of the *lncRNA Evf-2*

causes a mis-regulation of the key *Dlx5/6* locus and a subsequent loss of hippocampal interneurons at embyronic timepoints (Bond et al., 2009). Knockdown of the lncRNAs TUNA and RMST both lead to a significant reduction in the number of neurons generated from embryonic stem cell-derived neural stem cells (Lin et al., 2014; Ng et al., 2013). In addition, knockdown of the TUNA homologue *Megamind* in the developing zebrafish leads to a dramatic central-nervous system phenotyope, including reduced brain size and locomotor deficits (Lin et al., 2014; Ulitsky et al., 2011). While these studies have all provided evidence that a lncRNA can control neurogenic differentiation, few studies exist that pinpoint the stage of the lineage at which neurogenesis is affected.

In both the embryonic cortical ventricular zone (VZ) and adult ventricular-subventricular zone (V-SVZ), NSCs are glial cells that can both self-renew and yield intermediate progenitors that divide once or more before producing migratory young neurons (Kriegstein and Alvarez-Buylla, 2009). The production of transit-amplifying progenitors and their proliferation is a key aspect of brain development, and it has been suggested that defects at this stage of the neurogenic lineage underlie a number of human developmental disorders (Lui et al., 2011). Although transcription factors, microRNAs, and signaling pathways that control the transition between NSCs and transit-amplifying progenitors have been studied intensively (Ihrie and Alvarez-Buylla, 2011; Kriegstein and Alvarez-Buylla, 2009; Lui et al., 2011), lncRNAs that regulate this critical point in lineage progression have not been identified.

Pnky (previously called *lnc-pou3f2*) is a lncRNA that we initially identified as being expressed in the adult V-SVZ and specifically down-regulated during neurogenesis (Ramos et al., 2013). Unlike other lncRNAs that have been described in neurogenesis, we show that *Pnky* depletion leads to a dramatic increase in the number of young neurons generated from V-SVZ NSCs. This increase occurs at the transit-amplifying stage; *Pnky*-depleted NSCs generate more DLX2+ transit-amplifying cells. Taking advantage of near-clonal culture conditions and time lapse imaging, we were able to directly observe changes in cell behavior and differentiation.

Pnky depletion causes both an increase in NSCs that become neurogenic, as well as enhances the expansion of intermediate progenitor cells. *Pnky* is evolutionarily conserved to humans and is additionaly expressed in neural stem cells of the embryonic mouse cortex. In these embryonic stem cells depletion of *Pnky* causes an apparent accelerated differentiation and enhanced neurogenesis, mirroring the findings in adult NSCs. Finally, we show that *Pnky* binds known regulator of neurogenesis and splicing regulator PTBP1. We use RNA-seq to show that knockdown of *Pnky* results in alternative exon usage, suggesting a mechansim by which *Pnky* can enhance neurogenesis. Overall, our results describe a previously unappreciated role for a lncRNA in the regulation of neurogenic progenitor expansion across two unique neural stem cell pools.

Results

Pinky is a nuclear, neural lineage-specific lncRNA

RACE cloning followed by Sanger sequencing demonstrated *Pnky* to be an 825 nucleotides (nt) polyadenylated RNA encoded from three exons (**Fig. 4.1A**). Analysis with the Coding Potential Calculator (CPC) (Kong et al., 2007), PhyloCSF (Lin et al., 2011), and the Coding-Potential Assessment Tool (CPAT) (Wang et al., 2013a) indicated the *Pnky* transcript has no protein coding potential (**Fig. S4.1A**). Analysis of available RNA-seq datasets indicated that *Pnky* is specifically expressed in neural tissues and lineages (**Fig. 4.1B**). Furthermore, like many key developmental genes, the promoter of *Pnky* was "bivalent" with histone 3 lysine 27-trimethylation (H3K27me3) and histone 3 lysine 4-trimethylation (H3K4me3) in embryonic stem cells (ESCs), coherent with its repressed but "poised" transcriptional state (**Fig. 4.1B**). In contrast, in ESC-derived NSCs (ESC-NSCs) and V-SVZ NSCs, *Pnky* was monovalent with H3K4me3, consistent with its active transcription (**Fig. 4.1B and S4.1B**).

Nuclear fractionation of V-SVZ NSC cultures followed by RT-qPCR analysis demonstrated *Pnky* to be enriched in the nucleus as compared to coding mRNAs, which were enriched in cell lysates containing cytoplasm (**Fig. 4.1C**). Consistent with the nuclear fractionation studies, *in situ* hybridization (ISH) for *Pnky* demonstrated predominantly nuclear localization of the transcript (**Fig. 4.1D and S4.1C**).

ISH of adult mouse brain tissue revealed prominent expression of *Pnky* in the V-SVZ (**Fig. 4.1E**). To investigate whether *Pnky* expression is dynamic in the neurogenic lineage, we analyzed gene expression of V-SVZ cells acutely isolated from the brain with fluorescent activated cell sorting (FACS). *Pnky* expression was highest in NSCs, and decreased by 5.6-fold in migratory neuroblasts (**Fig. 4.1F**). Thus, *Pnky* is normally downregulated during lineage progression *in vivo.*

Pnky depletion enhances neurogenesis in V-SVZ NSCs

To investigate the function of *Pnky* in neurogenesis, we used V-SVZ NSC monolayer cultures that recapitulate key features of neurogenesis, including the production of transit-amplifying progenitors (**Fig. 4.1G**). After lentiviral transduction with control (shLuci-GFP) or *Pnky* (sh-1-GFP and sh-2-GFP) knockdown constructs, GFP+ NSCs were isolated with FACS and cultured (**Fig. S4.2A)**. *Pnky* knockdown was efficient in both self-renewing NSCs and after their differentiation (**Fig. S4.2B and S4.2C**). *Pnky*-deficient NSCs incorporated the thymidine analog 5-ethynyl-2'-deoxyuridine (EdU) at the same rate as control NSCs, suggesting that proliferation in self-renewal conditions was not affected (**Fig. 4.1H and S4.2D**). However, after differentiation, *Pnky*-deficient NSCs generated 3-fold more Tuj1+ neuroblasts after 7 d of differentiation (**Fig. 4.1I and 4.1J**). V-SVZ transit-amplifying cells express DLX2+ and normally divide several times before giving rise to neuroblasts (Doetsch et al., 2002; Ponti et al., 2013). With *Pnky* knockdown, there were 1.8 to 3.2 fold more EdU+, DLX2+ cells at 2 d of

differentiation, suggesting that *Pnky* plays a role in neurogenesis at the transit-amplifying stage of the lineage (**Fig. 4.1H and S4.2E**).

To further investigate how *Pnky* knockdown results in enhanced neurogenesis, we plated NSCs infected with *Pnky* knockdown or control vector with large numbers of uninfected NSCs (ratio of infected, GFP+ NSCs to uninfected NSCs was approximately 1:5000, **Fig 4.2A**). After 4 d of differentiation, these cultures produced well-isolated colonies of GFP+ cells (**Fig. 4.2B**). With *Pnky* knockdown, the proportion of colonies containing Tuj1+, GFP+ neuroblasts was increased (**Fig. 4.2C)**. Furthermore, these neuronal colonies also contained greater numbers of Tuj1+, GFP+ cells **(Fig. 4.2D)**. These data suggested that the increased neurogenesis resulted from both a shift towards neuronal lineage commitment and an increase in cell amplification within the neurogenic lineage.

Time-lapse microscopy demonstrates *Pnky* regulates progenitor expansion

To directly observe the behavior of individual NSCs with *Pnky* knockdown, we used time-lapse video microscopy to image the genesis of GFP+ daughter cells every 15 min for 3 d. We followed the fate of 316 GFP+ NSCs with Pnky-knockdown and 531 GFP+ control NSCs. After differentiation, we could identify both Tuj1+ neurogenic clones (**Fig. S4.3A**) and GFAP+ glial clones (**Fig. S4.3B**). Clones arising from NSCs with Pnky-knockdown were 48% more likely to be neurogenic (**Fig. S4.3C**). Most non-neurogenic NSCs did not divide or divided infrequently (average number of divisions per cell=0.20, **Fig. S4.3D**) and produced GFAP+ cells. Thus, *Pnky*-knockdown increased the likelihood that NSCs produce neurogenic progenitors.

Neurogenic clones with *Pnky*-knockdown contained on average 2.74 fold more neuroblasts than control (**Fig. S4.3E**). This increase in neurogenesis in part related to increased progenitor cell proliferation before differentiation into neuroblasts: while control progenitors divided 3.93 (SD=4.68, n=44) times and gave rise to 2.05 (SD=1.42) generations of

daughter cells, *Pnky* knockdown resulted in 8.94 (SD=7.64, n=33) divisions (Fig. 2F) and 3.45 (SD=1.41) generations (**Fig. 4.2F and 4.2G**). The average cell cycle time was not different with *Pnky* knockdown (19.4 hours vs. 19.1 hours, p=0.89) (**Fig. 4.2H)**, indicating that the increased number of divisions relates to the maintenance of a proliferative cell state and not an accelerated cell cycle. Time-lapse imaging also enabled direct observation of apoptosis, and the number of GFP+ neurogenic progeny that underwent cell death was reduced 57% by *Pnky* knockdown (**Fig. 4.2I)**. Thus, *Pnky*-deficiency promotes neurogenic lineage progression from NSCs, increases the number of divisions of transit-amplifying neurogenic progenitors, and reduces their cell death (**Fig 4.2J)**.

***Pnky* is expressed in the embryonic cortex of mouse and human**

Pnky transcripts were detected in embryonic mouse brain (**Fig. 4.1B**), and ISH revealed its expression in the VZ, where embryonic NSCs reside, at E14.5 and E16.5 in the cortex (**Fig. 4.3A, S4.4A, and S4.4B**). *Pnky* has two regions of high conservation among vertebrates (**Fig. 4.3B and S4.4C**), and using strand-specific RNA-seq of gestational week 16 (GW16) human cortical samples, we detected a non-coding transcript divergent to *BRN2* (*POU3F2*) that included the conserved sequences. RACE cloning identified human *PINKY* (*PNKY*) as a polyadenylated 1592 nt transcript containing two conserved elements expressed from a conserved promoter region (**Fig. 4.3B**). As in the developing mouse brain, ISH of GW14.5 human cortex demonstrated *PNKY* expression in the VZ, with decreased levels in the subventricular zone and intermediate zone (SVZ/IZ), where young neurons begin to differentiate (**Fig. 4.3C and S4.4D**).

To investigate the role of *Pnky* in the embryonic VZ, we electroporated *Pnky* knockdown construct (sh-2-GFP) or control (shCtrl-GFP) into VZ cells at embryonic day 13.5 (E13.5) (**Fig. 4.3D)** and analyzed GFP+ cells in the VZ, SVZ/IZ, and cortical plate (CP), where neurons begin to mature (**Fig. 4.3E)**. With *Pnky* knockdown, there was a 35% decrease in GFP+ cells in the

VZ and a 26% increase in GFP+ cells in the SVZ/IZ (**Fig. 4.3F**), suggesting that *Pinky*-deficiency accelerated the progression of NSCs into transit-amplifying cells. *Sox2* is expressed in embryonic brain NSCs, and *Pnky*-deficiency reduced the proportion of SOX2+, GFP+ VZ cells (**Fig. 4.3G**). Furthermore, *Pnky*-deficiency increased the proportion of GFP+ cells that expressed SATB2, a transcription factor expressed in young neurons at this time point (**Fig. 4.3H**). These data indicate that *Pnky* regulates neurogenesis in both the embryonic and adult brain.

Pnky binds PTBP1 and regulates alternative splicing

Many lncRNAs regulate gene expression through interactions with specific protein partners. To identify *Pnky*-interacting proteins, we incubated *in vitro* transcribed biotinylated *Pnky* or antisense *Pnky* control RNA with V-SVZ NSC nuclear extract and used mass spectrometry for protein identification (**Fig. S4.5A and S4.5B**). Polypyrimidine tract-binding protein 1 (PTBP1) was identified as a binding partner of *Pnky*, and this interaction was confirmed by Western blot analysis (**Fig. 4.4A**). In contrast, both *Pnky* and control RNA bound HnRNP K, an RNA-binding protein previously shown to functionally interact with lncRNAs (Huarte et al., 2010) (**Fig. 4.4A)**. Furthermore, RNA immunoprecipitation with PTBP1 antibodies enriched for *Pnky* transcript but not U1 snRNA or beta-actin mRNA (**Fig. 4.4B**).

PTBP1 regulates pre-mRNA splicing during neuronal development (Keppetipola et al., 2012), and can also regulate mRNA steady-state levels (Yap et al., 2012). In the embryonic brain, loss of *Ptbp1* results in precocious neuronal differentiation (Shibasaki et al., 2013), and in fibroblasts, PTBP1 knockdown leads to direct neuronal trans-differentiation (Xue et al., 2013). PTBP1 knockdown in V-SVZ NSCs did not affect *Pnky* RNA levels (**Fig. S4.5C and S4.5D)**, suggesting PTBP1 is not required for *Pnky* RNA maturation or stability. Additionally, analysis of PTBP1 knockdown demonstrated larger neuronal colonies in a near-clonal assay, reminiscent of the effect seen with *Pnky* knockdown (**Fig. S4.5E)**

RNA-seq of V-SVZ-NSC culturers in proliferating conditions revealed very few differentially expressed genes upon *Pnky* knockdown. We reasoned that *Pnky* may be interacting with PTBP1 to affect the differential splicing of downstream targets, and so we used SpliceTrap (Wu et al., 2011) to identify transcripts that demonstrate differential exon usage. 60 genes were differentially spliced in response to *Pnky* depletion with both knockdown constructs (sh-1, sh-2) compared to control (**Fig 4.4C**)**,** of which a significant number (51/59, p = 3.738614e-06, Fisher's exact test) have been shown to bind PTBP1 . Taken together, our genome-wide analysis demonstrates *Pnky* depletion causes splicing changes of PTBP1-bound mRNAs, suggesting *Pnky* and PTBP can cooperate to regulate isoform expression essential for transit amplification and neurogenesis.

Discussion

This work characterizes the expression and function of the lncRNA *Pnky* for the first time. Because of the generally low expression levels of the transcript and its specific expression in neural stem cells, this transcript was unannotated until our lncRNA catalogue generation effort. These studies thus validate our approach and establish the V-SVZ lineage as an invaluable model system in the study of lncRNAs.

The *Pnky* phenotype differs from previously studied lncRNAs because it *increases* neurogenesis, suggesting *Pnky* is required for the maintanence of NSC identity rather than progression down the neuronal lineage. Our use of near-clonal assays and time-lapse imaging allowed us to carefully dissect the effects of *Pnky* knockdown on cell cycle and differentiation dynamics that would be impossible using bulk-culture assays.

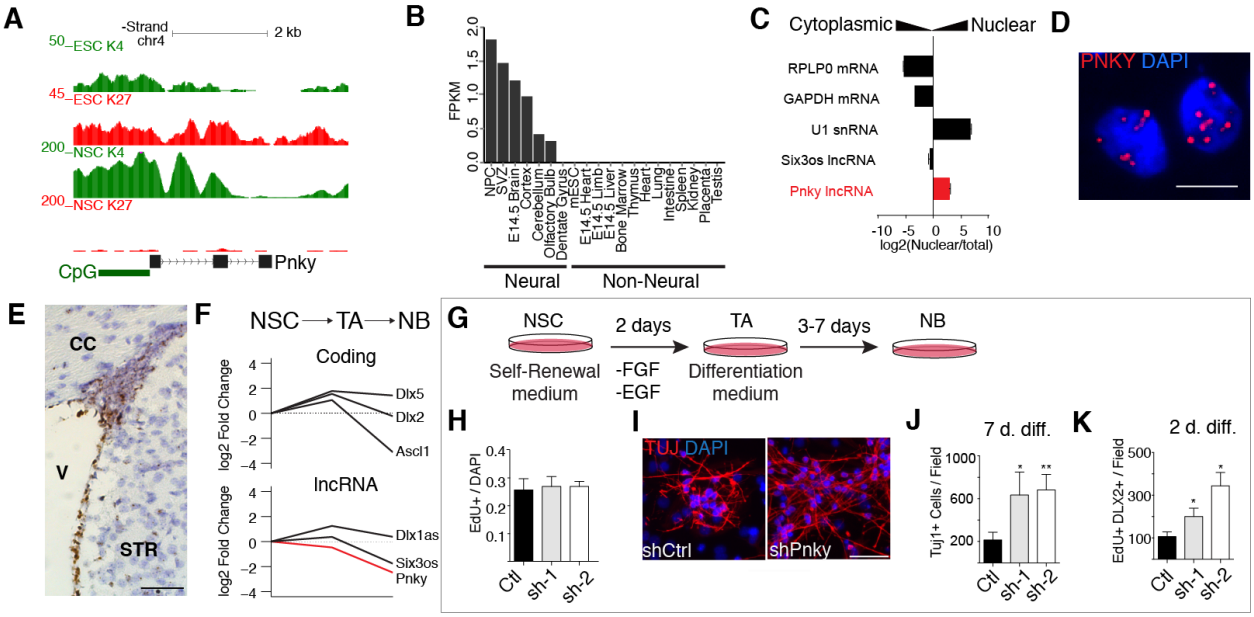
Pnky has two short regions of extremely high conservation, as has been described for other lncRNAs (Ulitsky et al., 2011), and it is expressed in the human cortex in the same cellular populations as in the mouse cortex. This suggests the intriguing possibility that *Pnky* has a conserved function across different neural stem cell systems and even across species. The

differentiation of neural stem cells and subsequent expansion of progenitors is thought to be important for the increased size and complexity of the human brain (Lui et al., 2011), and is also thought to underlie many developmental and psychiatric diseases (LaMonica et al., 2012; Mao et al., 2009). Could *Pnky* be disrupted or disregulated in disease states? Is *Pnky* differentially regulated in the human brain to allow for enhanced progenitor expansion?

Finally, our studies implicate control of alternative splicing as the molecular mechanism of action of *Pnky*. Interestingly, one of the differentially spliced genes, *Map3k7* (**Fig. 4.4D)**, has been previously described as undergoing alternative splicing and isoform switching during ESC differentiation (Salomonis et al., 2010), and this kinase can directly influence the *Wnt* signaling pathway (Ishitani et al., 1999; Salomonis et al., 2010). These data suggest an intriguing model whereby a complex of *Pnky* and PTBP1 can influence the differential splicing of key signaling components. Differential isoform expression ultimately determines whether a NSC will continue to differentiate (**Fig. 4.5**). These studies thus identify a novel mode of action for a lncRNA in neurogenesis.

Figures

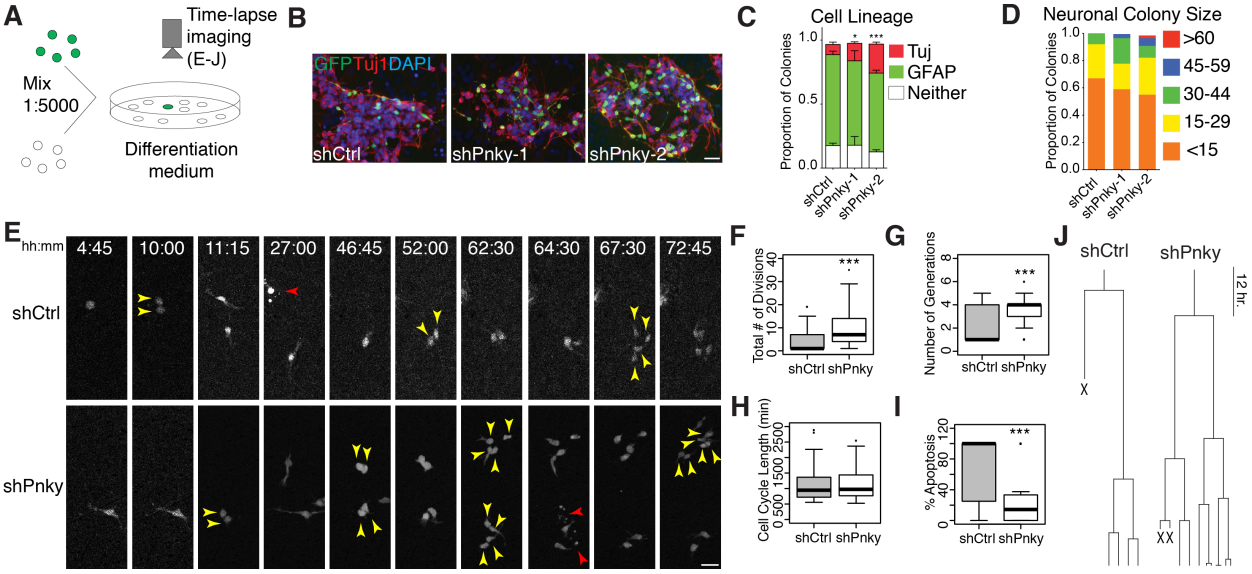
FIGURE 4.1



The lncRNA *Pinky* is expressed in V-SVZ-NSCs and regulates neuronal differentiation. A) UCSC genome browser view of the Pnky locus. Also shown are ChIP-seq tracks for H3K27me3 and H3K4me3 in ESCs and V-SVZ NSCs B) Fragments per kilobase per million mapped reads (FPKM) values for Pnky in indicated tissues C) Subcellular Fractionation followed by RT-qPCR for indicated lncRNAs and mRNAs. Error bars are propagated standard deviation from technical triplicate wells D) Branched-DNA in situ hybridization for Pnky in V-SVZ NSC cultures. Nuclei are counter-stained with 4',6-diamidino-2-phenylindole (DAPI). Scale bar= 10 um. E) Branched-DNA in situ hybridization for Pnky in adult mouse coronal tissue section. Pnky stains brown, nuclei are counter-stained with hematoxylin. V= ventricle, CC= corpus collosum, STR=striatum. Scale bar= 50 um F) Results of microarray expression analysis from FACS-isolated neural stem cells (NSCs), transit-amplifying cells (TAs), and neuroblasts (NBs). Value in NSCs set to 0. G) Schematic of V-SVZ NSC culture system. H) Quantification of EdU labeling counted from pure cultures of V-SVZ NSCs infected with control or *Pnky*-knockdown constructs. I)

Immunocytochemistry for Tuj1 (red) after 7 d. of differentiation in control-infected or *Pnky*-knockdown pure cultures. Scale bar= 50 um. J) Quantification of Tuj1+ NBs produced after 7 d. differentiation. K) Quantification of the number of TA cells, defined as being co-labeled with DLX2 and EdU, after 2 d. of differentiation. Error bars= standard deviation from triplicate wells, * p< .05, ***p<.001, student's t-test.

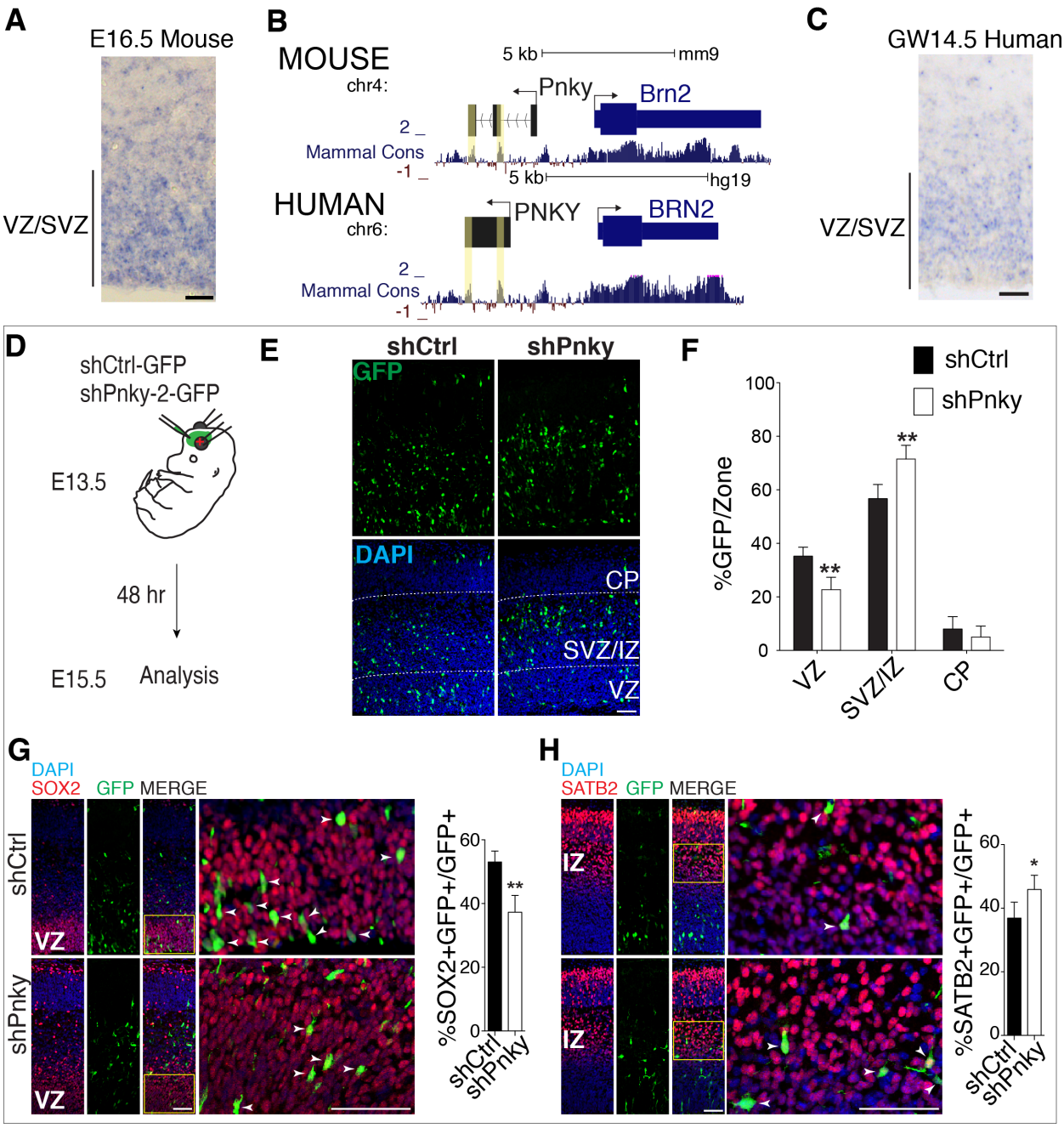
FIGURE 4.2



***Pinky* depletion leads to an expansion of neurogenic transit-amplifying progenitors** A) Schematic of experimental design. B) Representative images of isolated colonies after 4 d differentiation immunostained for Tuj1 (red) and GFP (green). Scale bar= 50 um C) Quantification of the fate of GFP+ cells within isolated colonies. Error bars are standard deviation from triplicate experiments. D) Quantification of number of Tuj1+ neuroblasts found in individual neurogenic colonies. E) Representative frames from time-lapse video of control-infected (top) or shPnky (bottom)-infected single cells. Time in differentiation medium is indicated. Yellow arrows indicate daughter cells resulting from a recent division, red arrows indicate apoptosis. Scale bar= 10 um. F) Box-and-whisker plots of total number of divisions undergone by a single initial progenitor and all of its daughter cells. N= 44 shCtrl and 33 shPnky progenitors. G) Box-and-whisker plots of number of generations per initial neurogenic progenitor for shCtrl and shPnky. N=44 shCtrl and 33 shPnky progenitors. H) Box-and-whisker plots of cell-cycle length, measured as the time between the first and second division. N= 17 shCtrl and 28 shPnky progenitors.I) Box-and-whisker plots of % of progeny per single neurogenic

progenitor that underwent apoptosis. N= 44 shCtrl and 33 shPnky progenitors. J) Tree diagram for the frames shown in E and corresponding movies. X indicates cell underwent apoptosis. *p < .05 ***p<.001, student's t-test.

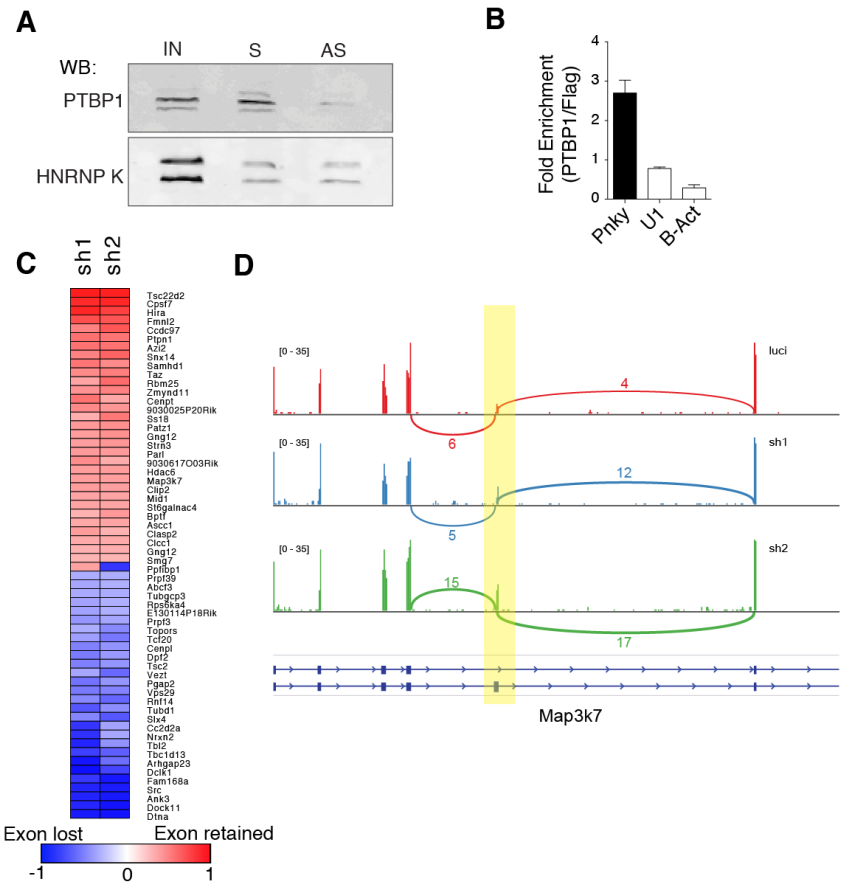
FIGURE 4.3



Pinky is expressed in the developing mouse and human cortex and regulates the differentiation of mouse cortical progenitors *in vivo* A) *In situ* hybridization of embryonic day 15.5 (E15.5) mouse brain for *Pnky*. B) Genome browser track demonstrating that mouse *Pnky* transcripts has two regions of high conservation, shown by PhyloP score (yellow box). Human PNKY genomic region is shown below with conservation indicated. For clarity, the mouse –

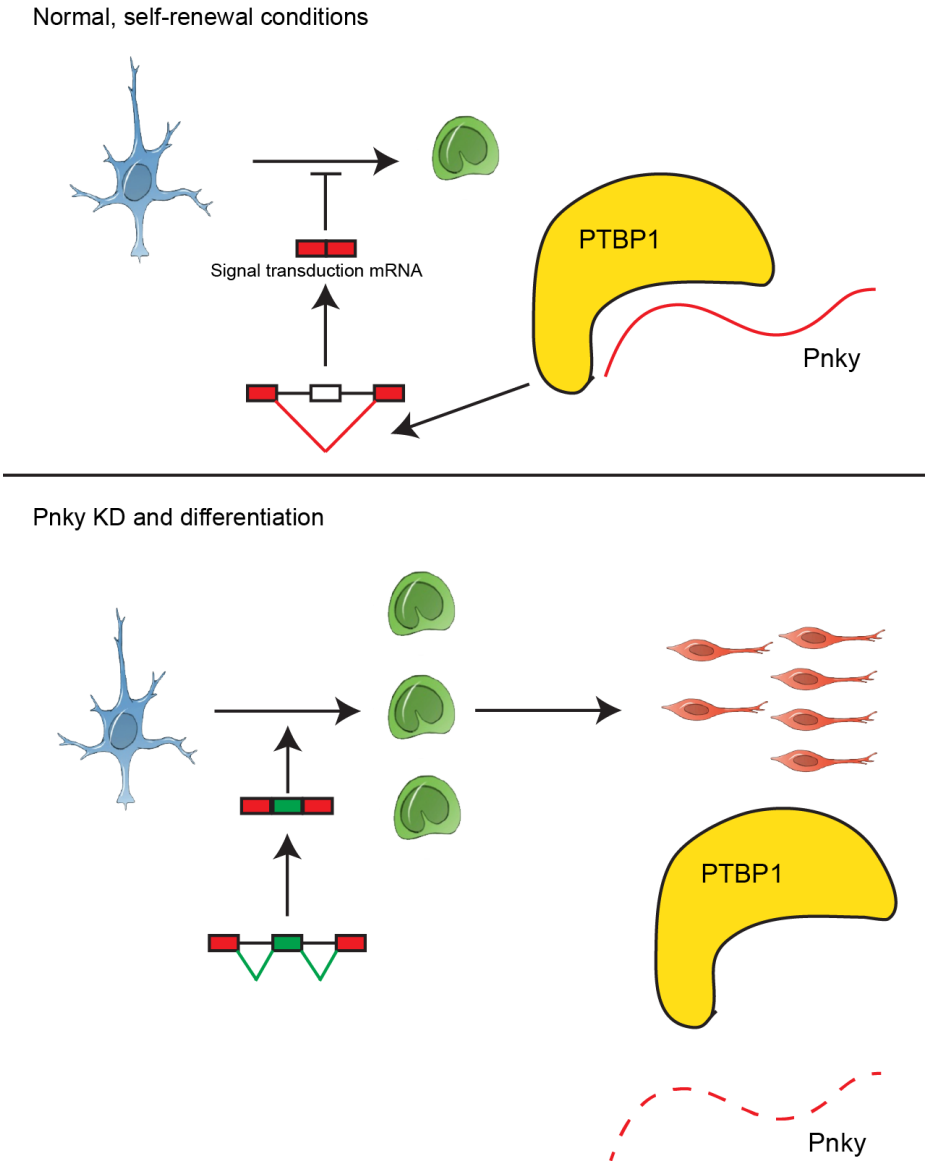
strand is shown left to right. C) *In situ* hybridization of gestational week 14.5 (GW14.5) human brain for *PNKY*. D) Schematic of electroporation of mouse embryonic brain in utero. E) Cortical sections at E16.5 electroporated with shCtrl (left) or shPnky (right), stained for GFP (top) and with DAPI counterstain (bottom). VZ= ventricular zone, SVZ/IZ= subventricular zone/intermediate zone, CP= cortical plate. F) Quantification of GFP+ cell distribution in indicated zones. G) Cortical sections at E16.5 electroporated with shCtrl or shPnky stained for GFP (Green), Sox2 (Red), with DAPI nuclear counterstain (Left). Yellow box indicates region expanded in the fourth panel. Arrowheads indicate co-labeled cells. Right: quantification of Sox2+ GFP+ cells. H) Left: Cortical sections stained for GFP (Green), SATB2 (Red), with DAPI nuclear counterstain (Left). Yellow box indicates region expanded in the fourth panel. Right: Quantification of SATB2+ GFP+ cells. All error bars are standard deviation, n=4 brains. *p<.05, **p<.01. All scale bars = 50um.

FIGURE 4.4



***Pinky* interacts with PTBP1 and regulates alternative splicing** A) Immunoblot for RNA-pulldown experiment with biotin-labeled Sense (S) or Anti-sense (AS) *Pnky* RNA with V-SVZ NSC nuclear extract. B) RT-qPCR detection for indicated RNAs recovered by PTBP1-specific antibody compared to control FLAG antibody. Error bars are propagated standard deviation from technical triplicate wells. C) Heat-map representation of differential exon usage determined using SpliceTrap in pure cultures of V-SVZ NSCs infected with *Pnky* sh-1 or *Pnky* sh-2, compared to control infection. Transcripts with differential splicing in both sh-1 and sh-2 samples are shown. D) Plot of junction-spanning reads for the Map3k7 gene from RNA-seq of cultures infected with of shCtrl (red), shPnky-1 (blue), shPnky-2 (green). Yellow box indicates an exon that is preferentially gained in upon *Pnky* knockdown.

FIGURE 4.5



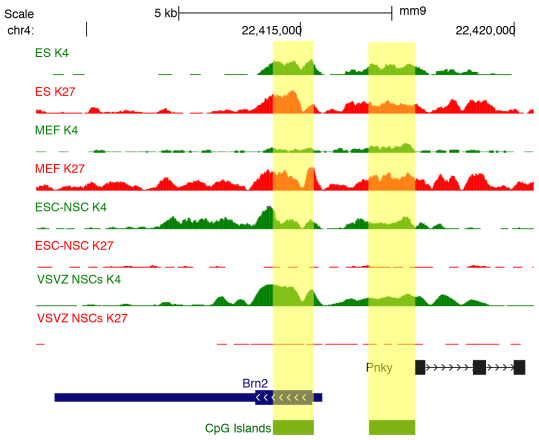
A model for *Pnky*-PTBP regulation of neurogenesis: Under differentiation conditions, *Pnky* binding PTBP influences target splicing, resulting in isoforms that inhibit differentiation and transit-amplification. Upon differentiation or Pnky knockdown, PTBP1 cannot influence the splicing of its targets, and pro-differentiation isoforms are produced.

FIGURE S4.1

A

Method	Pnky Value	Protein-Coding Cutoff
CPAT	0.0262	> 0.44
CPC	-0.729778	>0
PhyloCSF	-89.5917	>0

B



C

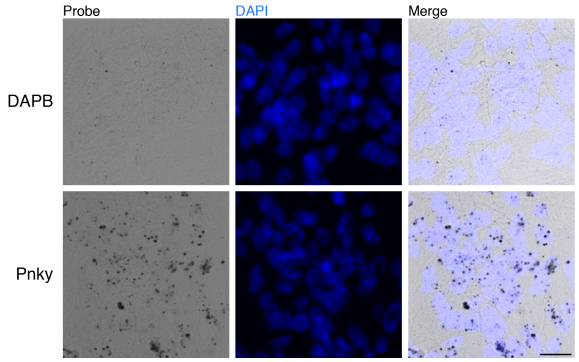
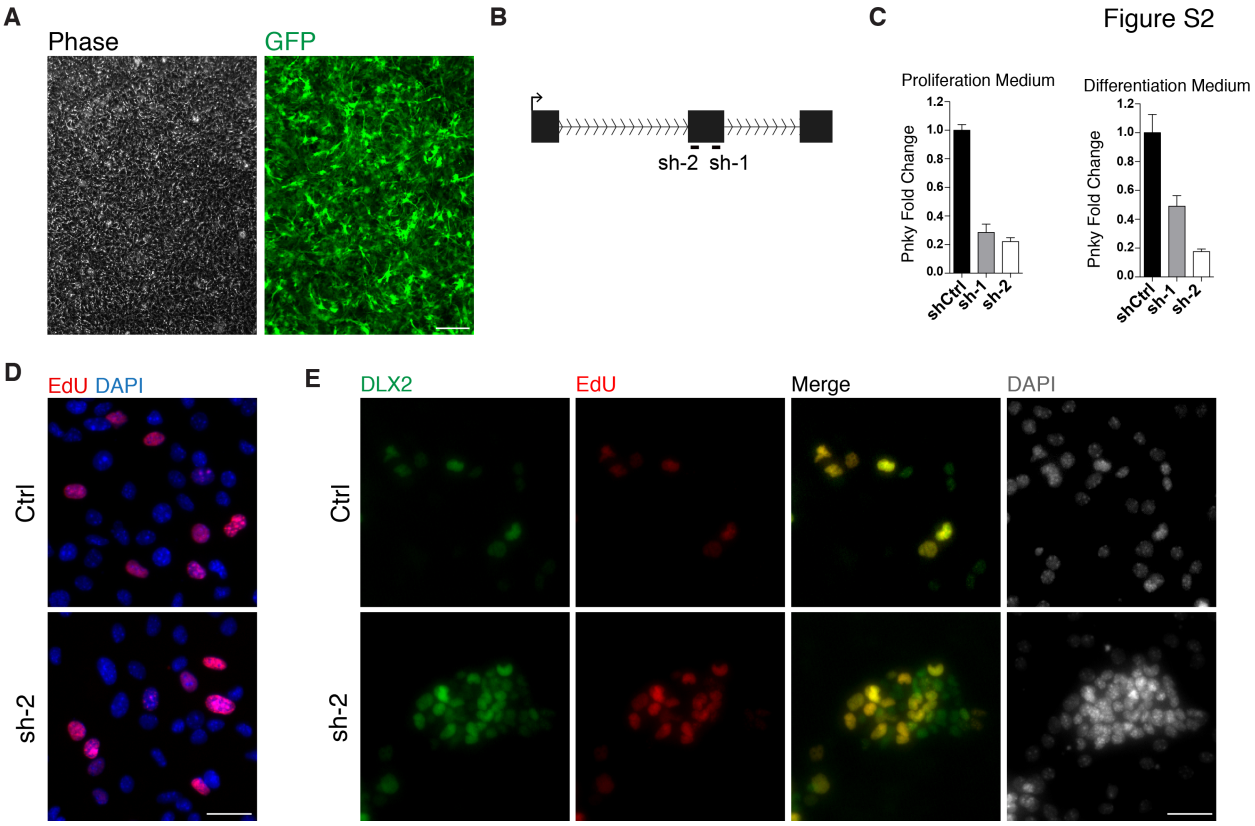


Figure S1

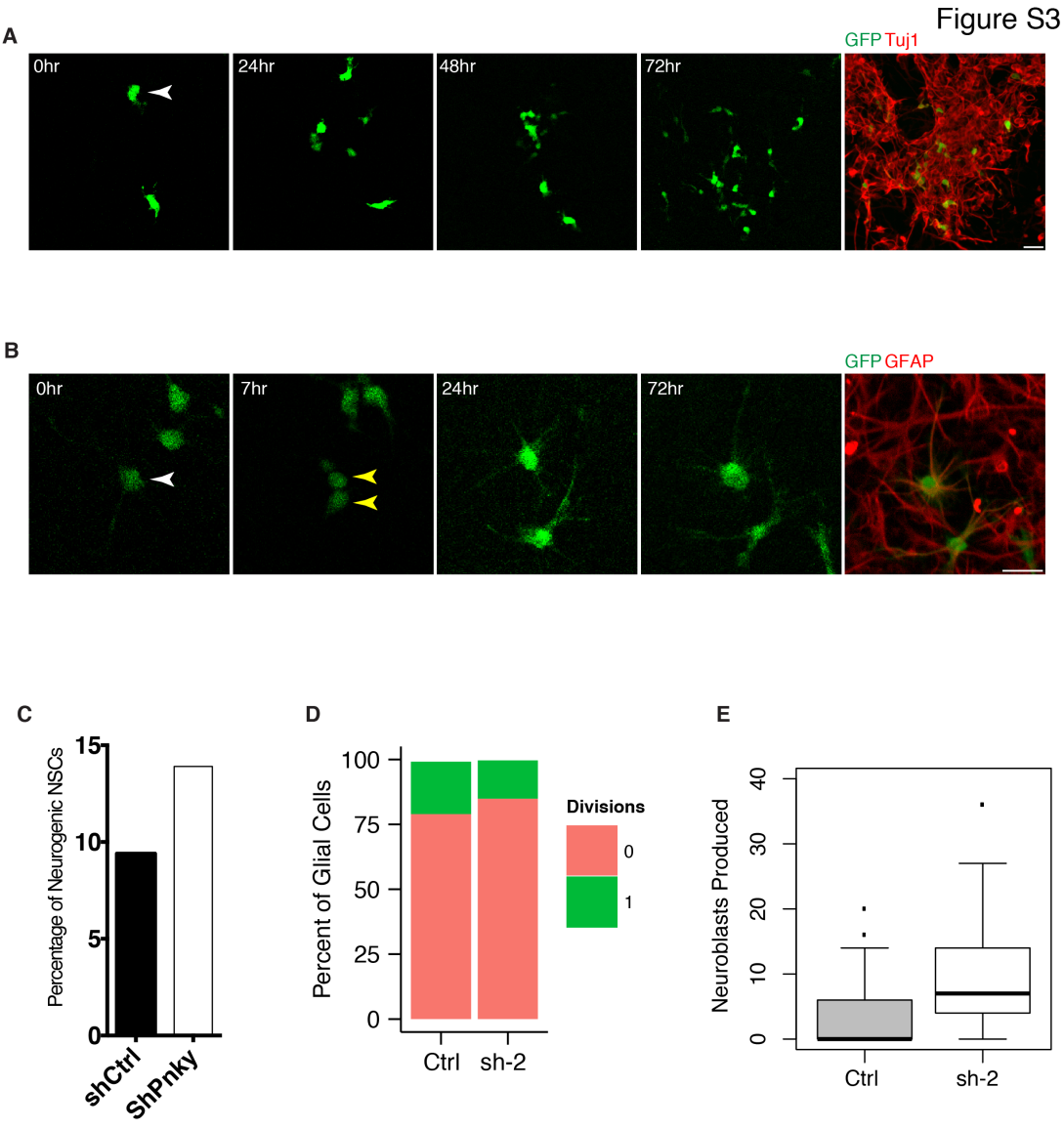
Characterization of the long noncoding RNA *Pnky*. A) Table of protein-coding scores derived from CPAT, CPC, and PhyloCSF, and the cut-off for determining a transcript is likely to encode a protein. B) Genome browser view of the Brn2 and Pnky loci with H3K4me3 and H3K27me3 ChIP-seq tracks for embryonic stem cells (ES), mouse embryonic fibroblasts (MEF), ESC-derived neural stem cells (ESC-NSCs), and V-SVZ NSCs. Yellow boxes indicate promoter regions defined by CpG islands. C) In situ hybridization using branched DNA probes targeting Pnky and negative control bacterial transcript DAPB in V-SVZ NSC cultures. Nuclei are counter-stained with DAPI. Scale bar= 20 um

FIGURE S4.2



Characterization of Pnky knockdown cultures A) Representative image of purified GFP+ cultures photographed with phase (left) and GFP filter (right). Scale bar= 100 um B) Schematic of *Pnky* RNA and the regions targeted by sh-1 and sh-2. C) *Pnky* knockdown efficiency in purified cultures in proliferation medium (left) or differentiation medium (right). Fold change was calculated with the delta-delta Ct method, normalized to shCtrl and Rplp0 housekeeping gene. Error bars are standard deviation propagated by least-squares method. D) Pure cultures of VSVZ-NSCs infected with shCtrl (top) or Pnky-sh-2 (bottom) labeled with EdU for 1 h and stained for incorporated EdU with Click-iT chemistry. Nuclei counterstained with DAPI. Scale bar= 20 um E) Pure cultures immunostained with DLX2 (green) and stained for EdU (red) after 3 hours of labeling in differentiation medium. Nuclei counterstained with DAPI. Scale bar= 20 um

FIGURE S4.3

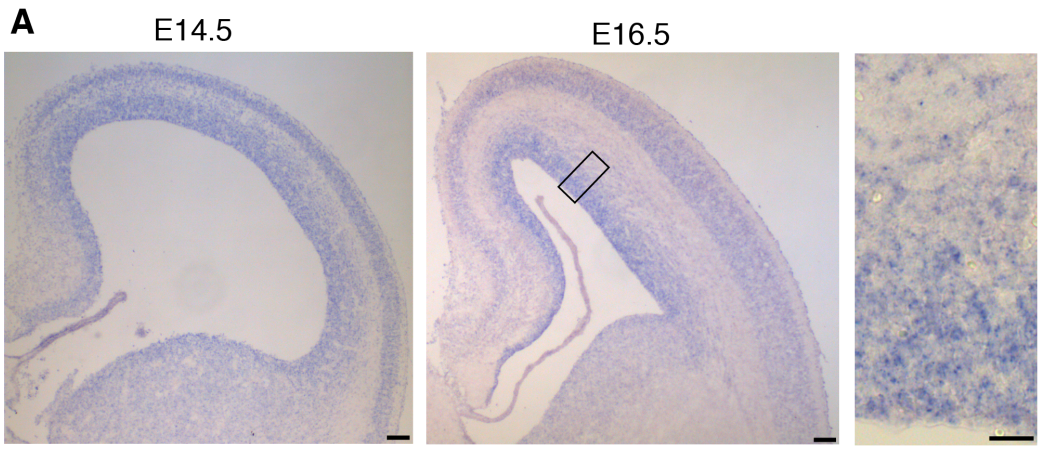


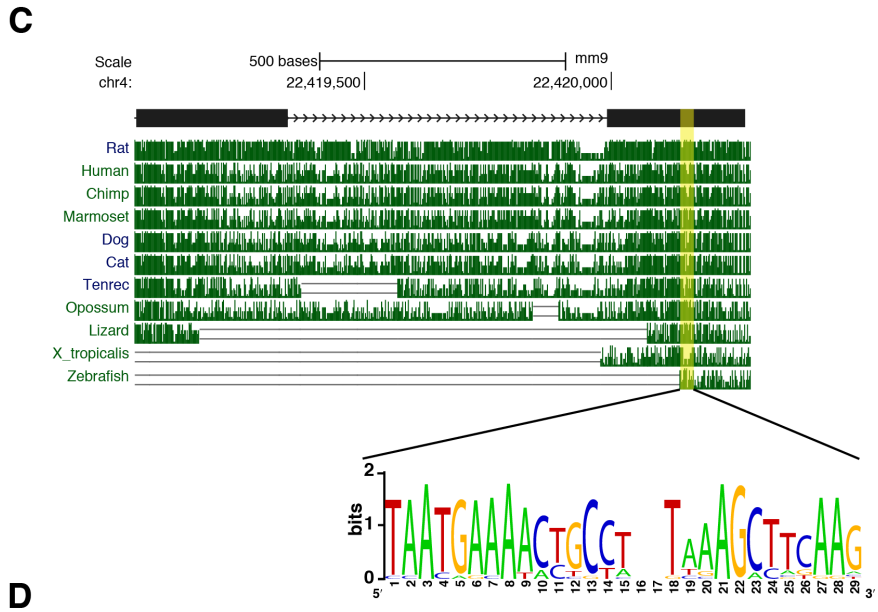
Analysis of time-lapse imaging of V-SVZ cultures A) Frames from time-lapse movies, demonstrating a neurogenic clone. Immediately after imaging, plate was fixed and immunostained for GFP (green) and Tuj1 (red). White arrow indicates initial tracked NSC. Scale bar= 20 um B) Frames from time-lapse movies, demonstrating a glial clone undergoing one division. White arrow indicates initial tracked NSC and yellow arrows indicate daughter cells resulting from a division. Immediately after imaging, plate was fixed and immunostained for GFP (green) and GFAP (red). Scale bar=20 um C) Bar graph representing the percentage of initial

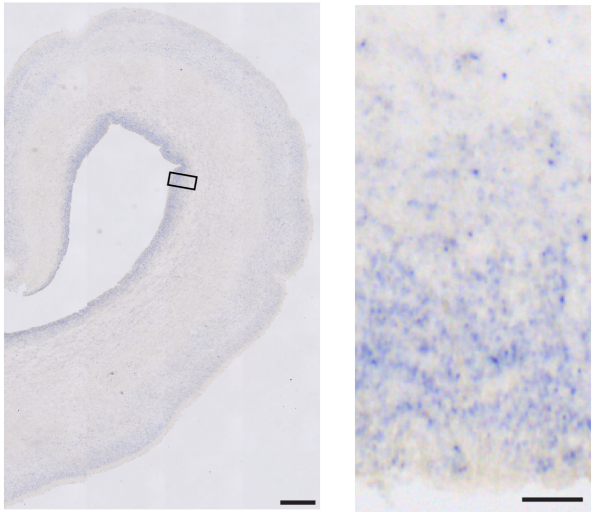
tracked progenitors that gave rise to neuroblasts. N= 316 GFP+ NSCs with Pnky-knockdown and 531 GFP+ control NSCs D) Bar graph representing the proportion of glial clones that did not divide, or divided once. N= 481 shCtrl cells and 272 shPnky cells. E) Box-and-whisker plot representing the total number of progeny produced per initial neurogenic NSC. N= 44 shCtrl and 33 shPnky progenitors.

FIGURE S4.4

Figure S4



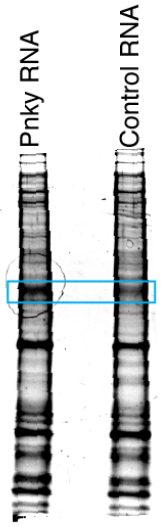




Evolutionary conservation of *Pnky* and expression in mouse and human embryonic cortex A) In situ hybridization of *Pnky* in embryonic day 14.5 (E14.5) mouse brains. Scale bar= 100 um. B) In situ hybridization for *Pnky* in E16.5 mouse brain. Scale bar= 100 um. Black box indicates area of detail shown at right and in Figure 4.3A. Scale bar for low magnification images= 20 um. Scale bar for high magnification images= 20 um C) Genome browser view of exons 2 and 3 of mouse *Pnky* and PhastCons scores across indicated vertebrates. The blue box highlights a sequence conserved to zebrafish. In the sequence logo, a score of 2 bits indicates bases are perfectly conserved across all genomes. D) In situ hybridization of *Pnky* in human gestational week 14.5 (GW14.5) brain. Scale bar= 500 um. Black box indicates area of detail shown at right and in Figure 4.3C. Scale bar=20 um. E) In situ hybridization using negative control sense probe for human *PNKY* in embryonic human brain. Scale bar= 20 um

FIGURE S4.5

A



B

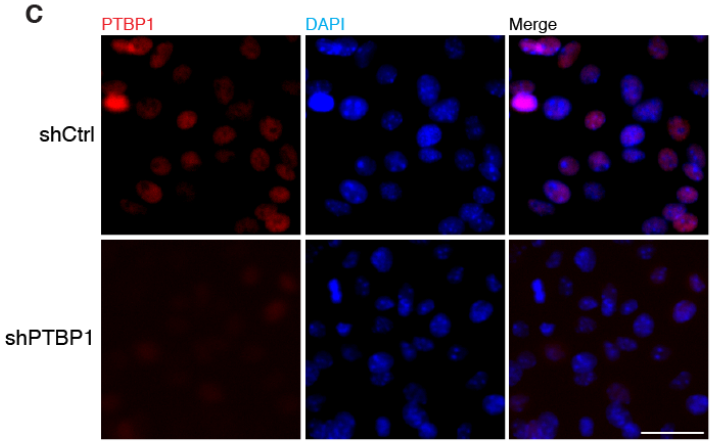
sense

Protein	# Peptides Matching	Protein Coverage (%)
VIM	29	27
DDX5	32	20.8
PTBP1	12	21.4
SGPL1	14	13.2
DHE3	5	9.7

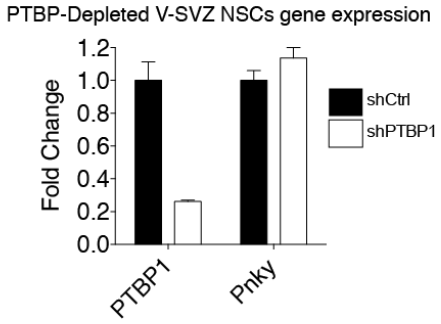
antisense

Protein	# Peptides Matching	Protein Coverage (%)
VIM	38	37.8
U2AF1	19	14.1
DDX5	27	13.7
DHE3	10	11.5
NONO	6	9.5

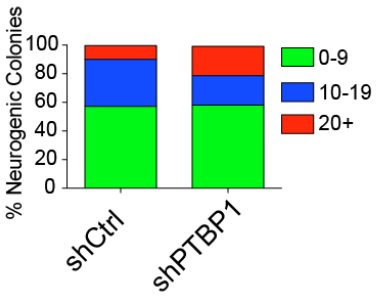
C



D



E



PTBP1 binds *Pnky* and is expressed in V-SVZ cultures A) Sypro Ruby-stained SDS-PAGE protein gel of proteins retrieved with biotinylated Pnky RNA or control RNA. Blue box indicates area excised and sent for mass spectrometry. B) Top 5 proteins identified by mass spectrometry from excised bands, excluding keratin contaminant peptides. C) Immunostaining for PTBP1 (red) in pure V-SVZ NSC cultures infected with control or shPTBP1. Nuclei are counterstained with DAPI. Scale bar= 20 um D) qRT-PCR for *Ptbp1* mRNA or *Pnky* lncRNA in pure V-SVZ NSC cultures infected with control or shPTBP1. Fold change was calculated with the delta-delta Ct method, normalized to shCtrl and GAPDH housekeeping gene. Error bars are standard deviation propagated by least-squares method. E) Quantification of number of Tuj1+ neuroblasts found in individual neurogenic colonies in control or shPTBP-infected colonies.

Chapter 5: Conclusions and Future Directions

Implications for disease

The fact that lncRNAs are more brain-region specific than mRNAs makes them attractive targets for molecular diagnosis and potentially treatment of disease. LncRNAs have been implicated in a wide variety of neurodegenerative and psychiatric disorders, including Huntington's disease (Lin et al., 2014), Alzheimer's disease (Faghihi et al., 2008), schitzophrenia (Barry et al., 2013), developmental delay (Talkowski et al., 2012), and autism (Ziats and Rennert, 2013). Our own studies found lncRNAs associated with protein-coding gene expression groups that correspond to various diseases including Huntington's and major depressive disorder (**Figure S3.3**). Our characterization and annotation of lncRNAs in the neural lineage provides the basis for studies into the role of lncRNAs in neurological disease.

LncRNAs are most well-characterized in cancer (Schmitt and Chang, 2013), and the expression of lncRNAs have been associated with tumors of the central nervous system (Han et al., 2012; Zhang et al., 2013; 2012b). Our custom microarrays and annotation of NSC lncRNAs have already been put to use in preliminary studies of a mouse brain tumor model, which demonstrates clear differential expression of lncRNAs during tumor progression. Because lncRNAs seem to be more cell-type specific than mRNAs, they represent ideal therapeutic targets and can potentially have high diagnostic and prognostic value. Interestingly, PTBP1 overexpression can drive the invasiveness of glioma cell lines (Cheung et al., 2009). This raises the possibility that a set of oncogenic lncRNAs can target PTBP1 in glioma for enhanced invasiveness. Further studies could be carried out to characterize both the expression of lncRNAs in glioma as well as the unique repertoire of lncRNAs bound to key proteins, including PTBP1.

Implications for human development

The expansion of the human cortex is thought to be a result of an increased number of transit-amplifying progenitors (Lui et al., 2011), and so the players involved in the transition between stem and progenitor cell and subsequent expansion are of great interest for the evolution and development of the human brain. *Pnky* is expressed in the human VZ and becomes downregulated in the SVZ. Does *Pnky* restrict the transit-amplifying state in humans as it does in mice? Could differential regulation of *Pnky* in the human brain contribute to an expanded transit-amplifying population? With the cloning of human *Pnky* and identification of its conserved domains, these questions can begin to be answered. *Pnky* knockdown experiments could be performed in human ESC-derived NSCs and even human cortex slice cultures. Characterization of the conserved stretches of *Pnky* sequence with mass-spectroscopy based methods from human cell lysate could establish new or conserved interacting partners for *Pnky* in humans.

In an analysis of regions of DNA that rapidly evolved from chimpanzee to human, Pollard et al. identified a putative noncoding RNA transcript that was expressed during cortical development (Pollard et al., 2006). This study illustrates that lncRNAs and other noncoding regions have the potential to be a driving force in brain expansion and evolution. Our lncRNA pipeline is can be adapted to any system, and efforts should now be focused on defining a 'lncRNA signature' for neurogenesis in the embryonic human brain. A comparison of these lncRNAs to the lncRNAs implicated in mouse neurogenesis has the potential to identify species-specific functions for lncRNAs, and could provide mechanistic insight into the evolution of the human cortex.

Chromatin-based control of neurogenesis

Both polycomb and trithorax group proteins have been shown to play essential roles in embryonic and adult neurogenesis (Gonzales-Roybal and Lim, 2013). PRC1 component *Bmi1* is

essential for the self-renewal of stem cells cultured from the V-SVZ (Molofsky et al., 2003) and from stem cells in the developing cortex through its repression of cell cycle inhibitors (Fasano et al., 2009). Deletion of PRC2 components EED or EZH2 in the embryonic cortex led to a prolonged neurogenic phase, in part due to the derepression of proneural transcription factor *neurogenin1* (Hirabayashi and Gotoh, 2010). In V-SVZ NSCs, EZH2 regulates both the proliferation and self-renewal of stem cells as well as a distinct role in proper neuronal differentiation through the repression of *Olig2* (Hwang et al., 2014). H3K4 methyltransferase MLL1 is required for activation of the *Dlx2* promoter in the V-SVZ NSCs: MLL-null V-SVZ-NSCs cells do not properly express *Dlx2*, and there is a dramatic decrease in neurogenesis, both *in vivo* and *in vitro*. Surprisingly, these cells maintain normal levels of H3K4me3 at the *Dlx2* promoter, but have aberrant enrichment of H3K27me3, indicating that MLL is required to resolve this 'bivalent' chromatin structure to allow for the expression of *Dlx2* and subsequent neuronal differentiation.

Studies of epigenetic modifiers in neurogenesis have clearly demonstrated that both polycomb and trithorax group proteins need to be dynamically targeted to loci genome-wide to execute a self-renewal and/or differentiation transcriptional program, however the mechanism through which these modifiers can be targeted in a cell-type and lineage-specific manner remains an open question. LncRNAs' cell-type specificity and documented ability to recruit and direct histone modifiers make them intriguing candidates for future study of the epigenetic control of neurogenesis in neural stem cells.

RNA-splicing and microRNA control of neurogenesis

miRNA-124 is abundantly expressed in the mammalian brain, and *in vitro* studies in ESCs and other cells lines implicated it in the repression of non-neuronal transcripts(Makeyev et al., 2007). The function of miR-124 has also been investigated in V-SVZ NSCs, where its expression is specifically upegulated during the transit-amplifying cell to neuroblast transition

(Cheng et al., 2009). Both *in vivo* and *in vitro*, knockdown of miR-124 blocked neurogenesis and promoted the maintenance of precursor divisions; in contrast, ectopic expression resulted in precocious differentiation into neuroblasts. The long noncoding RNA *Rncr3* is the major source of mir-124; *Rncr3* knockout mice displayed small brain size, impaired axon development and function in the dentate gyrus, and increased cell death in the developing retina (Sanuki et al., 2011). While in cannot be ruled out that the *Rncr3* transcript itself may have function beyond serving as a source for miR-124, these data suggest that miR-124 is an essential component of proper neuronal development. Are there lncRNAs (including *Rncr3*) that could serve as 'sponges' for miR-124?

Interestingly, miR-124 can indirectly regulate alternative splicing during neural development through its down-regulation of PTBP1 (Makeyev et al., 2007). PTBP1 is ubiquitously expressed throughout the body, except for in mature neurons, where it is replaced by PTBP2 (sometimes called neuronal PTB, nPTB). PTBP1 post-transcriptionally regulates PTBP2 levels; exon-skipping induced by PTBP1 generates a *Ptbp2* transcript that is targeted for non-sense mediated decay (NMD) and destruction (Boutz et al., 2007). Further studies into PTBP1 function demonstrates it can post-transcriptionally regulate the steady-state mRNA levels of hundreds of targets through NMD-independent mechanisms (Yap et al., 2012), and can bind the 3' UTR of target genes and effectively block or enhance the actions of microRNAs, including miR-124 (Xue et al., 2013). These data therefore place PTBP1 at the center of a post-transcriptional gene regulatory network that incorporates alternative splicing, mRNA decay pathways, and microRNA targeting to maintain a non-neuronal state.

Interestingly, the neuron and ESC-expressed TUNA is found in a PTBP1-containing complex, raising the intriguing possibility that targeting or proper functioning of PTBP1 can controlled by different lncRNAs. Perhaps TUNA targets PTBP to a set of targets in ESCs and then upon neural differentiation *Pnky* is induced and retargets PTBP1 to a unique set of mRNAs. Follow up studies should address both the differential binding of lncRNAs, mRNAs,

and other proteins to PTBP1. LncRNAs could serve as a scaffold that integrates protein-protein and protein-mRNA interactions to execute the appropriate alternative splicing program.

Transgenic models and the Importance of Genetic Approaches

As discussed in Chapter 1, experimental manipulation of lncRNAs with shRNA vs. transgenic mice can produce different results. This has been most recently illustrated with lincRNA-p21; an inducible knockout mouse was generated that did not display any of the same gene-expression changes described upon siRNA-mediated knockdown (Dimitrova et al., 2014). Instead, the lincRNA-p21 deficient fibroblasts showed misregulation of *p21*, the nearest gene neighbor. The downregulation of *p21* expression was found to be the major driver of the phenotype. A transgenic model was therefore able to unveil a strong *cis* regulatory effect that had not been seen with siRNA studies. This example is particularly important to consider for the *Pnky* lncRNA, which is transcribed ~3 kb upstream of known neurogenic regulator *Brn2.* It is possible that a *Pnky* transgenic mouse will demonstrate reduced *Brn2* expression, although we were not able to detect such a change with shRNA-mediated knockdown in V-SVZ cultures. We are currently working to generate an inducible *Pnky* knockout mouse to address these questions and better understand the function of *Pnky in vivo*.

Conclusion

The work presented here demonstrates that high-throughput methods can be applied to an *in vivo* stem cell system for the identification of a relatively small number of lncRNAs of interest from a list of thousands. Intensive study of a lncRNA that was uniquely recovered and previously unannotated revealed it has a role in the key transition between stem cell and transit-amplifying cell. Overall, the work presented here advances both the field of lncRNA biology and neural stem cell biology by: 1) Presenting a publically accessible catalogue of candidate lncRNAs involved in V-SVZ neurogenesis; 2) Demonstrating that this catalogue can be used to identify candidates that are required for the differentiation of specific lineages; 3) Demonstrating

that depletion of a single noncoding transcript can dramatically increase neurogenesis both *in vitro* and *in vivo,* both in adult and embryonic NSCs; and 4) Providing evidence that a lncRNA can affect alternative splicing choices which in turn affect NSC behavior. Beyond the study of NSCs and lncRNAs, the work also has implications for human disease and evolution, and opens up future directions for the further study of *Pnky* and hundreds of other lncRNAs in neural development.

References

Alvarez-Buylla, A., and Lim, D.A. (2004). For the long run: maintaining germinal niches in the adult brain. Neuron *41*, 683–686.

Alvarez-Dominguez, J.R., Hu, W., Yuan, B., Shi, J., Park, S.S., Gromatzky, A.A., van Oudenaarden, A., and Lodish, H.F. (2014). Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. Blood *123*, 570–581.

Anderson, S.A., Eisenstat, D.D., Shi, L., and Rubenstein, J.L. (1997). Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes. Science *278*, 474–476.

Aprea, J., Prenninger, S., Dori, M., Ghosh, T., Monasor, L.S., Wessendorf, E., Zocher, S., Massalini, S., Alexopoulou, D., Lesche, M., et al. (2013). Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. The EMBO Journal *32*, 3145–3160.

Ayoub, A.E., Oh, S., Xie, Y., Leng, J., Cotney, J., Dominguez, M.H., Noonan, J.P., and Rakic, P. (2011). Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. Proceedings of the National Academy of Sciences *108*, 14950–14955.

Barry, G., Briggs, J.A., Vanichkina, D.P., Poth, E.M., Beveridge, N.J., Ratnu, V.S., Nayler, S.P., Nones, K., Hu, J., Bredy, T.W., et al. (2013). The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. Mol. Psychiatry.

Batista, P.J., and Chang, H.Y. (2013). Long Noncoding RNAs: Cellular Address Codes in Development and Disease. Cell *152*, 1298–1307.

Bekhor, I., Bonner, J., and Dahmus, G.K. (1969). Hybridization of chromosomal RNA to native DNA. Proc. Natl. Acad. Sci. U.S.a. *62*, 271–277.

Belgard, T.G., Marques, A.C., Oliver, P.L., Abaan, H.O., Sirey, T.M., Hoerder-Suabedissen, A., García-Moreno, F., Molnár, Z., Margulies, E.H., and Ponting, C.P. (2011). A transcriptomic atlas of mouse neocortical layers. Neuron *71*, 605–616.

Berghoff, E.G., Clark, M.F., Chen, S., Cajigas, I., Leib, D.E., and Kohtz, J.D. (2013). Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. Development *140*, 4407–4416.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes Dev *19*, 1635–1655.

Bertani, S., Sauer, S., Bolotin, E., and Sauer, F. (2011). The Noncoding RNA Mistral Activates Hoxa6 and Hoxa7 Expression and Stem Cell Differentiation by Recruiting MLL1 to Chromatin.

Molecular Cell *43*, 1040–1046.

Bond, A.M., Vangompel, M.J.W., Sametsky, E.A., Clark, M.F., Savage, J.C., Disterhoft, J.F., and Kohtz, J.D. (2009). Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. Nat Neurosci *12*, 1020–1027.

Boutz, P.L., Stoilov, P., Li, Q., Lin, C.-H., Chawla, G., Ostrow, K., Shiue, L., Ares, M., and Black, D.L. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev *21*, 1636–1652.

Bult, C.J., Eppig, J.T., Kadin, J.A., Richardson, J.E., Blake, J.A., Mouse Genome Database Group (2008). The Mouse Genome Database (MGD): mouse biology and model systems. Nucleic Acids Res *36*, D724–D728.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes & Development *25*, 1915–1927.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., et al. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. Journal of Neuroscience *28*, 264–278.

Carpenter, S., Aiello, D., Atianand, M.K., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J.B., et al. (2013). A long noncoding RNA mediates both activation and repression of immune response genes. Science *341*, 789–792.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature *491*, 454–457.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell *147*, 358–369.

Cheng, L.-C., Pastrana, E., Tavazoie, M., and Doetsch, F. (2009). miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. Nat Neurosci *12*, 399–408.

Cheung, H.C., Hai, T., Zhu, W., Baggerly, K.A., Tsavachidis, S., Krahe, R., and Cote, G.J. (2009). Splicing factors PTBP1 and PTBP2 promote proliferation and migration of glioma cell lines. Brain *132*, 2277–2288.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell *44*, 667–678.

Cui, K., Zang, C., Roh, T.-Y., Schones, D.E., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell *4*, 80–93.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa,

M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P., et al. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. Nature *503*, 371–376.

Dimitrova, N., Zamudio, J.R., Jong, R.M., Soukup, D., Resnick, R., Sarma, K., Ward, A.J., Raj, A., Lee, J.T., Sharp, P.A., et al. (2014). LincRNA-p21 Activates p21 In cis to Promote Polycomb Target Gene Expression and to Enforce the G1/S Checkpoint. Mol Cell.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Soldà, G., Simons, C., et al. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Research *18*, 1433–1445.

Doetsch, F., Petreanu, L., Caille, I., Garcia-Verdugo, J.M., and Alvarez-Buylla, A. (2002). EGF converts transit-amplifying neurogenic precursors in the adult brain into multipotent stem cells. Neuron *36*, 1021–1034.

Eißmann, M., Gutschner, T., Hämmerle, M., Günther, S., Caudron-Herger, M., Groß, M., Schirmacher, P., Rippe, K., Braun, T., Zörnig, M., et al. (2012). Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. RNA Biology *9*, 1076–1087.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods *4*, 207–214.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St Laurent, G., Kenny, P.J., and Wahlestedt, C. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat Med *14*, 723–730.

Fasano, C.A., Phoenix, T.N., Kokovay, E., Lowry, N., Elkabetz, Y., Dimos, J.T., Lemischka, I.R., Studer, L., and Temple, S. (2009). Bmi-1 cooperates with Foxg1 to maintain neural stem cell self-renewal in the forebrain. Genes & Development *23*, 561–574.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet *15*, 7–21.

Feng, J., Bi, C., Clark, B.S., Mady, R., Shah, P., and Kohtz, J.D. (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes & Development *20*, 1470–1484.

Fey, E.G., Krochmalnic, G., and Penman, S. (1986). The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. J. Cell Biol. *102*, 1654–1665.

Gabory, A., Jammes, H., and Dandolo, L. (2010). The H19 locus: role of an imprinted non-coding RNA in growth and development. Bioessays *32*, 473–480.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.-K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E.O., et al. (2011). An alternative splicing switch

regulates embryonic stem cell pluripotency and reprogramming. Cell *147*, 132–146.

Ghanem, N., Jarinova, O., Amores, A., Long, Q., Hatch, G., Park, B.K., Rubenstein, J.L.R., and Ekker, M. (2003). Regulatory roles of conserved intergenic domains in vertebrate Dlx bigene clusters. Genome Research *13*, 533–543.

Goecks, J., Nekrutenko, A., Taylor, J., Galaxy Team (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.

Gong, C., and Maquat, L.E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature *470*, 284–288.

Gonzales-Roybal, G., and Lim, D.A. (2013). Chromatin-based epigenetics of adult subventricular zone neural stem cells. Front Genet *4*, 194.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology *29*, 644–652.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science *329*, 643–648.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., et al. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Developmental Cell *24*, 206–214.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071–1076.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

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

Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nature Structural &Amp; Molecular Biology *21*, 198–206.

Han, L., Zhang, K., Shi, Z., Zhang, J., Zhu, J., Zhu, S., Zhang, A., Jia, Z., Wang, G., Yu, S., et

al. (2012). LncRNA profile of glioblastoma reveals the potential role of lncRNAs in contributing to glioblastoma pathogenesis. Int. J. Oncol. *40*, 2004–2012.

He, L., and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nature Reviews Genetics *5*, 522–531.

Hirabayashi, Y., and Gotoh, Y. (2010). Epigenetic control of neural precursor cell fate during development. Nat. Rev. Neurosci. *11*, 377–388.

Holmes, D.S., Mayfield, J.E., Sander, G., and Bonner, J. (1972). Chromosomal RNA: its properties. Science *177*, 72–74.

Horvath, S., and Dong, J. (2008). Geometric interpretation of gene coexpression network analysis. PLoS Comput. Biol. *4*, e1000117.

Hsieh, J. (2012). Orchestrating transcriptional control of adult neurogenesis. Genes & Development *26*, 1010–1021.

Hu, G., Tang, Q., Sharma, S., Yu, F., Escobar, T.M., Muljo, S.A., Zhu, J., and Zhao, K. (2013). Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. Nat. Immunol. *14*, 1190–1198.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols *4*, 44–57.

Huang, R.C., and Bonner, J. (1965). Histone-bound RNA, a component of native nucleohistone. Proc. Natl. Acad. Sci. U.S.a. *54*, 960–967.

Huang, R.C., and Huang, P.C. (1969). Effect of protein-bound RNA associated with chick embryo chromatin on template specificity of the chromatin. J. Mol. Biol. *39*, 365–378.

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

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nature Genetics *43*, 621–629.

Hwang, W.W., Salinas, R.D., Siu, J.J., Kelley, K.W., Delgado, R.N., Paredes, M.F., Alvarez-Buylla, A., Oldham, M.C., and Lim, D.A. (2014). Distinct and separable roles for EZH2 in neurogenic astroglia. Elife *3*, e02439.

Ihrie, R.A., and Alvarez-Buylla, A. (2011). Lake-front property: a unique germinal niche by the lateral ventricles of the adult brain. Neuron *70*, 674–686.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell Proteomics *4*, 1265–1272.

Ishitani, T., Ninomiya-Tsuji, J., Nagai, S., Nishita, M., Meneghini, M., Barker, N., Waterman, M., Bowerman, B., Clevers, H., Shibuya, H., et al. (1999). The TAK1-NLK-MAPK-related pathway antagonizes signalling between beta-catenin and transcription factor TCF. Nature *399*, 798–802.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Jim nez, C.R., Huang, L., Qiu, Y., and Burlingame, A.L. (2001). In-Gel Digestion of Proteins for MALDI-MS Fingerprint Mapping (Hoboken, NJ, USA: John Wiley & Sons, Inc.).

Kallen, A.N., Zhou, X.-B., Xu, J., Qiao, C., Ma, J., Yan, L., Lu, L., Liu, C., Yi, J.-S., Zhang, H., et al. (2013). The imprinted H19 lncRNA antagonizes let-7 microRNAs. Mol Cell *52*, 101–112.

Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A., and Reinberg, D. (2014). Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin. Mol Cell *53*, 290–300.

Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., and Reik, W. (2012). The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol *14*, 659–665.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome Research *12*, 656–664.

Keppetipola, N., Sharma, S., Li, Q., and Black, D.L. (2012). Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. Crit. Rev. Biochem. Mol. Biol. *47*, 360–378.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci USA *106*, 11667–11672.

Kim, E.J., Leung, C.T., Reed, R.R., and Johnson, J.E. (2007). In vivo analysis of Ascl1 defined progenitors reveals distinct developmental dynamics during adult neurogenesis and gliogenesis. Journal of Neuroscience *27*, 12764–12774.

Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal *3*, ra8.

Klattenhoff, C.A., Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P.A., Steinhauser, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S., et al. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell *152*, 570–583.

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res *35*, W345–W349.

Kraus, P., Sivakamasundari, V., Lim, S.L., Xing, X., Lipovich, L., and Lufkin, T. (2013). Making sense of Dlx1 antisense RNA. Dev. Biol. *376*, 224–235.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature *493*, 231–235.

Kretz, M., Webster, D.E., Flockhart, R.J., Lee, C.S., Zehnder, A., Lopez-Pajares, V., Qu, K., Zheng, G.X.Y., Chow, J., Kim, G.E., et al. (2012). Suppression of progenitor differentiation requires the long noncoding RNA ANCR. Genes & Development *26*, 338–343.

Kriegstein, A., and Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. Annu. Rev. Neurosci. *32*, 149–184.

Kuwajima, T., Nishimura, I., and Yoshikawa, K. (2006). Necdin promotes GABAergic neuron differentiation in cooperation with Dlx homeodomain proteins. Journal of Neuroscience *26*, 5383–5392.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 1–7.

Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: a model for nuclear organelles. Nat Rev Mol Cell Biol *4*, 605–612.

LaMonica, B.E., Lui, J.H., Wang, X., and Kriegstein, A.R. (2012). OSVZ progenitors in the human cortex: an updated perspective on neurodevelopmental disease. Curr. Opin. Neurobiol. *22*, 747–753.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lee, J.T. (2012). Epigenetic regulation by long noncoding RNAs. Science *338*, 1435–1439.

Legnini, I., Morlando, M., Mangiavacchi, A., Fatica, A., and Bozzoni, I. (2014). A Feedforward Regulatory Loop between HuR and the Long Noncoding RNA linc-MD1 Controls Early Phases of Myogenesis. Mol Cell *53*, 506–514.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Lien, W.-H., Guo, X., Polak, L., Lawton, L.N., Young, R.A., Zheng, D., and Fuchs, E. (2011). Genome-wide maps of histone modifications unwind in vivo chromatin states of the hair follicle lineage. Cell Stem Cell *9*, 219–232.

Lim, D.A., Huang, Y.-C., Swigut, T., Mirick, A.L., Garcia-Verdugo, J.M., Wysocka, J., Ernst, P., and Alvarez-Buylla, A. (2009). Chromatin remodelling factor Mll1 is essential for neurogenesis

from postnatal neural stem cells. Nature *458*, 529–533.

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

Lin, N., Chang, K.-Y., Li, Z., Gates, K., Rana, Z.A., Dang, J., Zhang, D., Han, T., Yang, C.-S., Cunningham, T.J., et al. (2014). An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment. Mol Cell.

Liu, H., Sadygov, R.G., and Yates, J.R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. *76*, 4193–4201.

Liu, J.K., Ghattas, I., Liu, S., Chen, S., and Rubenstein, J.L. (1997). Dlx genes encode DNA-binding proteins that are expressed in an overlapping and sequential pattern during basal ganglia differentiation. Dev. Dyn. *210*, 498–512.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.-H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nature Genetics *42*, 1113–1117.

Long, J.E., Garel, S., Alvarez-Dolado, M., Yoshikawa, K., Osumi, N., Alvarez-Buylla, A., and Rubenstein, J.L.R. (2007). Dlx-dependent and -independent regulation of olfactory bulb interneuron differentiation. Journal of Neuroscience *27*, 3230–3243.

Lui, J.H., Hansen, D.V., and Kriegstein, A.R. (2011). Development and evolution of the human neocortex. Cell *146*, 18–36.

Lundgren, D.H., Hwang, S.-I., Wu, L., and Han, D.K. (2010). Role of spectral counting in quantitative proteomics. Expert Rev Proteomics *7*, 39–53.

Makeyev, E.V., Zhang, J., Carrasco, M.A., and Maniatis, T. (2007). The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Molecular Cell *27*, 435–448.

Mao, Y., Ge, X., Frank, C.L., Madison, J.M., Koehler, A.N., Doud, M.K., Tassa, C., Berry, E.M., Soda, T., Singh, K.K., et al. (2009). Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling. Cell *136*, 1017–1031.

Matsuda, T., and Cepko, C.L. (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. Proc. Natl. Acad. Sci. U.S.a. *101*, 16–22.

Mayfield, J.E., and Bonner, J. (1971). Tissue differences in rat chromosomal RNA. Proc. Natl. Acad. Sci. U.S.a. *68*, 2652–2655.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature Biotechnology *28*, 495–501.

Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. Nature Structural &Amp; Molecular Biology *20*, 300–307.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proceedings of the National Academy of Sciences *105*, 716–721.

Mercer, T.R., Gerhardt, D.J., Dinger, M.E., Crawford, J., Trapnell, C., Jeddeloh, J.A., Mattick, J.S., and Rinn, J.L. (2012). Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nature Biotechnology *30*, 99–104.

Mercer, T.R., Qureshi, I.A., Gokhan, S., Dinger, M.E., Li, G., Mattick, J.S., and Mehler, M.F. (2010). Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. BMC Neurosci *11*, 14.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Molofsky, A.V., Pardal, R., Iwashita, T., Park, I.-K., Clarke, M.F., and Morrison, S.J. (2003). Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature *425*, 962–967.

Molyneaux, B.J., Arlotta, P., Menezes, J.R.L., and Macklis, J.D. (2007). Neuronal subtype specification in the cerebral cortex. Nat. Rev. Neurosci. *8*, 427–437.

Morán, I., Akerman, İ., van de Bunt, M., Xie, R., Benazra, M., Nammo, T., Arnes, L., Nakić, N., García-Hurtado, J., Rodríguez-Seguí, S., et al. (2012). Human β Cell Transcriptome Analysis Uncovers lncRNAs That Are Tissue-Specific, Dynamically Regulated, and Abnormally Expressed in Type 2 Diabetes. Cmet *16*, 435–448.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin. Science *322*, 1717–1720.

Nakagawa, S., Ip, J.Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T., and Prasanth, K.V. (2012). Malat1 is not an essential component of nuclear speckles in mice. Rna *18*, 1487–1499.

Ng, S.-Y., Bogu, G.K., Soh, B.S., and Stanton, L.W. (2013). The Long Noncoding RNA RMST Interacts with SOX2 to Regulate Neurogenesis. Molecular Cell *51*, 349–359.

Ng, S.-Y., Johnson, R., and Stanton, L.W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. The EMBO Journal *31*, 522–533.

Nickerson, J.A., Krochmalnic, G., Wan, K.M., and Penman, S. (1989). Chromatin architecture and nuclear RNA. Proc. Natl. Acad. Sci. U.S.a. *86*, 177–181.

Oldham, M.C., Horvath, S., and Geschwind, D.H. (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc. Natl. Acad. Sci. U.S.a. *103*, 17973–17978.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D.H. (2008). Functional organization of the transcriptome in human brain. Nat Neurosci *11*,

1271–1282.

Onoguchi, M., Hirabayashi, Y., Koseki, H., and Gotoh, Y. (2012). A noncoding RNA regulates the neurogenin1 gene locus during mouse neocortical development. Proceedings of the National Academy of Sciences *109*, 16939–16944.

Pandey, R., Mondal, T., Mohammad, F., and Enroth, S. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell *32*, 232–246.

Pastrana, E., Cheng, L.-C., and Doetsch, F. (2009). Simultaneous prospective purification of adult subventricular zone neural stem cells and their progeny. Proceedings of the National Academy of Sciences *106*, 6387–6392.

Paul, J., and Duerksen, J.D. (1975). Chromatin-associated RNA content of heterochromatin and euchromatin. Mol. Cell. Biochem. *9*, 9–16.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annu. Rev. Genet. *36*, 233–278.

Poitras, L., Ghanem, N., Hatch, G., and Ekker, M. (2007). The proneural determinant MASH1 regulates forebrain Dlx1/2 expression through the I12b intergenic enhancer. Development *134*, 1755–1765.

Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. Nature *443*, 167–172.

Ponti, G., Obernier, K., and Alvarez-Buylla, A. (2013). Lineage progression from stem cells to new neurons in the adult brain ventricular-subventricular zone. Cell Cycle *12*, 1649–1650.

Prensner, J.R., Iyer, M.K., Sahu, A., Asangani, I.A., Cao, Q., Patel, L., Vergara, I.A., Davicioni, E., Erho, N., Ghadessi, M., et al. (2013). The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nature Genetics *45*, 1392–1398.

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., and Turner, D.J. (2008). A large genome center's improvements to the Illumina sequencing system. Nature Methods *5*, 1005–1010.

Ramos, A.D., Diaz, A., Nellore, A., Delgado, R.N., Park, K.-Y., Gonzales-Roybal, G., Oldham, M.C., Song, J.S., and Lim, D.A. (2013). Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. Cell Stem Cell *12*, 616–628.

Rapicavoli, N.A., and Blackshaw, S. (2009). New meaning in the message: noncoding RNAs and their role in retinal development. Dev. Dyn. *238*, 2103–2114.

Rapicavoli, N.A., Poth, E.M., and Blackshaw, S. (2010). The long noncoding RNA RNCR2 directs mouse retinal cell specification. BMC Dev. Biol. *10*, 49.

Rapicavoli, N.A., Poth, E.M., Zhu, H., and Blackshaw, S. (2011). The long noncoding RNA

Six3OS acts in trans to regulate retinal development by modulating Six3 activity. Neural Development *6*, 32.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Roan, N.R., Chu, S., Liu, H., Neidleman, J., Witkowska, H.E., and Greene, W.C. (2014). Interaction of Fibronectin With Semen Amyloids Synergistically Enhances HIV Infection. J. Infect. Dis.

Saito, T. (2006). In vivo electroporation in the embryonic mouse central nervous system. Nat Protoc *1*, 1552–1558.

Salditt-Georgieff, M., and Darnell, J.E. (1982). Further evidence that the majority of primary nuclear RNA transcripts in mammalian cells do not contribute to mRNA. Molecular and Cellular Biology *2*, 701–707.

Salomonis, N., Schlieve, C.R., Pereira, L., Wahlquist, C., Colas, A., Zambon, A.C., Vranizan, K., Spindler, M.J., Pico, A.R., Cline, M.S., et al. (2010). Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. Proceedings of the National Academy of Sciences *107*, 10514–10519.

Sanchez-Elsner, T. (2006). Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. Science *311*, 1118–1123.

Sanuki, R., Onishi, A., Koike, C., Muramatsu, R., Watanabe, S., Muranishi, Y., Irie, S., Uneo, S., Koyasu, T., Matsui, R., et al. (2011). miR-124a is required for hippocampal axogenesis and retinal cone survival through Lhx2 suppression. Nat Neurosci *14*, 1125–1134.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M., et al. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife *2*, e01749.

Scheffler, B., Walton, N.M., Lin, D.D., Goetz, A.K., Enikolopov, G., Roper, S.N., and Steindler, D.A. (2005). Phenotypic and functional characterization of adult brain neuropoiesis. Proc. Natl. Acad. Sci. U.S.a. *102*, 9353–9358.

Schmitt, A.M., and Chang, H.Y. (2013). Gene regulation: Long RNAs wire up cancer growth. Nature *500*, 536–537.

Sharma, S., Findlay, G.M., Bandukwala, H.S., Oberdoerffer, S., Baust, B., Li, Z., Schmidt, V., Hogan, P.G., Sacks, D.B., and Rao, A. (2011). Dephosphorylation of the nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA-protein scaffold complex. Proceedings of the National Academy of Sciences *108*, 11381–11386.

Shibasaki, T., Tokunaga, A., Sakamoto, R., Sagara, H., Noguchi, S., Sasaoka, T., and Yoshida, N. (2013). PTB deficiency causes the loss of adherens junctions in the dorsal telencephalon and

leads to lethal hydrocephalus. Cerebral Cortex *23*, 1824–1835.

Simon, M.D., Wang, C.I., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. Proceedings of the National Academy of Sciences *108*, 20497–20502.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.a. *102*, 15545–15550.

Sugitani, Y., Nakai, S., Minowa, O., Nishi, M., Jishage, K.-I., Kawano, H., Mori, K., Ogawa, M., and Noda, T. (2002). Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons. Genes Dev *16*, 1760–1765.

Sun, L., Goff, L.A., Trapnell, C., Alexander, R., Lo, K.A., Hacisuleyman, E., Sauvageau, M., Tazon-Vega, B., Kelley, D.R., Hendrickson, D.G., et al. (2013a). Long noncoding RNAs regulate adipogenesis. Proceedings of the National Academy of Sciences *110*, 3387–3392.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013b). Jpx RNA activates Xist by evicting CTCF. Cell *153*, 1537–1551.

Talkowski, M.E., Maussion, G., Crapper, L., Rosenfeld, J.A., Blumenthal, I., Hanscom, C., Chiang, C., Lindgren, A., Pereira, S., Ruderfer, D., et al. (2012). Disruption of a large intergenic noncoding RNA in subjects with neurodevelopmental disabilities. Am. J. Hum. Genet. *91*, 1128–1134.

Thomas, E.A., Coppola, G., Tang, B., Kuhn, A., Kim, S., Geschwind, D.H., Brown, T.B., Luthi-Carter, R., and Ehrlich, M.E. (2011). In vivo cell-autonomous transcriptional abnormalities revealed in mice expressing mutant huntingtin in striatal but not cortical neurons. Hum Mol Genet *20*, 1049–1060.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology *28*, 511–515.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. Molecular Cell *39*, 925–938.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689–693.

Tsuiji, H., Yoshimoto, R., Hasegawa, Y., Furuno, M., Yoshida, M., and Nakagawa, S. (2011). Competition between a noncoding exon and introns: Gomafu contains tandem UACUAAC

repeats and associates with splicing factor-1. Genes Cells *16*, 479–490.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550.

Vance, K.W., Sansom, S.N., Lee, S., Chalei, V., Kong, L., Cooper, S.E., Oliver, P.L., and Ponting, C.P. (2014). The long non-coding RNA Paupar regulates the expression of both local and distal genes. The EMBO Journal *33*, 296–311.

Ventura, A., Meissner, A., Dillon, C.P., McManus, M., Sharp, P.A., Van Parijs, L., Jaenisch, R., and Jacks, T. (2004). Cre-lox-regulated conditional RNA interference from transgenes. Proc. Natl. Acad. Sci. U.S.a. *101*, 10380–10385.

Wallace, V.A., and Raff, M.C. (1999). A role for Sonic hedgehog in axon-to-astrocyte signalling in the rodent optic nerve. Development *126*, 2901–2909.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 1–7.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013a). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res *41*, e74–e74.

Wang, P., Xue, Y., Han, Y., Lin, L., Wu, C., Xu, S., Jiang, Z., Xu, J., Liu, Q., and Cao, X. (2014). The STAT3-Binding Long Noncoding RNA lnc-DC Controls Human Dendritic Cell Differentiation. Science *344*, 310–313.

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013b). Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Developmental Cell *25*, 69–80.

Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B., and Schultz, P.G. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science *309*, 1570–1573.

Wu, J., Akerman, M., Sun, S., McCombie, W.R., Krainer, A.R., and Zhang, M.Q. (2011). SpliceTrap: a method to quantify alternative splicing under single cellular conditions. Bioinformatics *27*, 3010–3016.

Xue, Y., Ouyang, K., Huang, J., Zhou, Y., Ouyang, H., Li, H., Wang, G., Wu, Q., Wei, C., Bi, Y., et al. (2013). Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. Cell *152*, 82–96.

Yang, L., Lin, C., Jin, C., Yang, J.C., Tanasa, B., Li, W., Merkurjev, D., Ohgi, K.A., Da Meng, Zhang, J., et al. (2013). lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature 1–8.

Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K.A., Grinstein, J.D., Dorrestein, P.C., and Rosenfeld, M.G. (2011). ncRNA- and Pc2 Methylation-Dependent Gene Relocation between Nuclear

Structures Mediates Gene Activation Programs. Cell *147*, 773–788.

Yang, Y.W., Flynn, R.A., Chen, Y., Qu, K., Wan, B., Wang, K.C., Lei, M., and Chang, H.Y. (2014). Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. Elife *3*, e02046.

Yap, K., Lim, Z.Q., Khandelia, P., Friedman, B., and Makeyev, E.V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. Genes & Development *26*, 1209–1223.

Yoon, J.-H., Abdelmohsen, K., Kim, J., Yang, X., Martindale, J.L., Tominaga-Yamanaka, K., White, E.J., Orjalo, A.V., Rinn, J.L., Kreft, S.G., et al. (2013). Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. Nat Commun *4*, 2939.

Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 Suppresses Target mRNA Translation. Mol Cell *47*, 648–655.

Young, T.L., Matsuda, T., and Cepko, C.L. (2005). The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. Curr. Biol. *15*, 501–512.

Zhang, B., Arun, G., Mao, Y.S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C.J., Wu, J., Zhang, C., et al. (2012a). The lncRNA Malat1 Is Dispensable for Mouse Development but Its Transcription Plays a cis-Regulatory Role in the Adult. Cell Rep *2*, 111–123.

Zhang, J.-X., Han, L., Bao, Z.-S., Wang, Y.-Y., Chen, L.-Y., Yan, W., Yu, S.-Z., Pu, P.-Y., Liu, N., You, Y.-P., et al. (2013). HOTAIR, a cell cycle-associated long noncoding RNA and a strong predictor of survival, is preferentially expressed in classical and mesenchymal glioma. Neuro-Oncology *15*, 1595–1603.

Zhang, X., Sun, S., Pu, J.K.S., Tsang, A.C.O., Lee, D., Man, V.O.Y., Lui, W.M., Wong, S.T.S., and Leung, G.K.K. (2012b). Long non-coding RNA expression profiles predict clinical phenotypes in glioma. Neurobiology of Disease *48*, 1–8.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Y., De, S., Garner, J.R., Smith, K., Wang, S.A., and Becker, K.G. (2010). Systematic analysis, comparison, and integration of disease based human genetic association data and mouse genetic phenotypic information. BMC Med Genomics *3*, 1.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell *40*, 939–953.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750–756.

Zhou, Q., and Anderson, D.J. (2002). The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. Cell *109*, 61–73.

Ziats, M.N., and Rennert, O.M. (2013). Aberrant expression of long noncoding RNAs in autistic brain. J. Mol. Neurosci. *49*, 589–593.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Long Noncoding RNAs with Enhancer-like Function in Human Cells. Cell *143*, 46–58.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

06/12/2014

Date