Molecular characterization and functional analysis of a novel long noncoding RNA in the mouse

Dissertation

for the award of the degree

''Doctor rerum naturalium''

of the Georg-August-Universtät Göttingen

within the doctoral program Genes and Development

of the Georg-August-Universtät School of Science (GAUSS)

submitted by

Parth Devesh Joshi

from Aliabada, India

Göttingen 2018

Thesis committee

Prof. Dr. Gregor Eichele, Dept. of Genes and Behavior, Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Reinhard Lührmann, Dept. of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Ernst A. Wimmer, Dept. of Developmental Biology, Georg-August-Universtät, Göttingen

Prof. Dr. Bernhard Herrmann, Dept. of Developmental Genetics, Max Planck Institute for Molecular Genetics, Berlin

Members of the Examination Board

Prof. Dr. Gregor Eichele (1st reviewer), Dept. of Genes and Behavior, Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Ernst A. Wimmer (2nd reviewer), Dept. of Developmental Biology, Georg-August-Universtät, Göttingen

Further members of the Examination Board

Prof. Dr. Reinhard Lührmann, Dept. of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Anastassia Stoykova, Dept. of Genes and Behavior, Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Gregor Bucher, Dept. of Developmental Biology, Georg-August-Universtät, Göttingen

Prof. Dr. Ahmed Mansouri, Dept. of Molecular Cell Differentiation, Max Planck Institute for Biophysical Chemistry, Göttingen

Date of Oral Examination: February 25, 2019

I hereby declare that the Ph.D. thesis entitled "*Molecular characterization and functional analysis of a novel long noncoding RNA in the mouse*" was written independently and with no other sources and aids than quoted.

Parth Devesh Joshi

Göttingen, December 2018

Table of Contents

Abstract

Majority of mammalian transcriptomes encode long noncoding RNAs (lncRNAs) that are not translated into proteins. LncRNAs are mRNA-like transcripts that are longer than 200 nucleotides, transcribed by RNA-Pol-II and undergo post-transcriptional modifications and splicing. Although they are less evolutionary conserved than protein-coding genes, lncRNAs have been shown to regulate several biochemical and cellular processes at a transcriptional level. Yet to date, the mechanism of only a handful of them has been studied in mammalian organisms. Therefore, in order to understand the function of lncRNAs in the mouse, we have performed large-scale RNA-sequencing of whole E14.5 mouse embryos. Results obtained revealed ~7000 putative lncRNA transcripts expressed at this developmental stage. From this list, we focused our attention on a small group of divergent lncRNAs that are transcribed in close proximity to physiologically important protein-coding genes and that showed significant evolutionary conservation in mammals. Our RNA *in situ* hybridization (ISH) data in mouse embryos confirmed that a few of these divergent lncRNAs genes share similar expression profiles with their neighboring protein-coding genes.

From this list, we selected an uncharacterized and mammalian conserved divergent lncRNA *Gm14204* that has an expression similar to its adjacent neurotransmitter transporter gene *Slc32a1*. *Gm14204* is divergently transcribed with respect to *Slc32a1* on the opposite strand with a distance of ~50 bp between the two genes. Keeping in mind the prevailing idea in the literature, we hypothesized that *Gm14204* lncRNA might regulate *Slc32a1* transcription. First, we performed an in-depth molecular characterization of *Gm14204* expression using RNA-ISH and observed that it has a nervous system-specific expression pattern in the developing mouse embryos and an ubiquitous expression throughout the adult brain, which is broadly comparable to *Slc32a1* mRNA expression. Next, using a series of RNA-FISH and single-molecule (sm) FISH experiments, we showed that *Gm14204*

lncRNA and *Slc32a1* mRNA transcripts are co-expressed in a subset of GABAergic interneuron population in the adult mouse brain.

To investigate the function of *Gm14204*, we have generated the first genetic mouse model for *Gm14204* by knock-in of transcription termination sequences into the first intron of this gene. LncRNA gene deletion was not advisable because *Gm14204* and *Slc32a1* genes are situated very close to each other and deleting a region of *Gm14204* might disrupt *cis*-regulatory elements in that locus. Thus, our novel strategy prematurely terminated the transcription of all *Gm14204* lncRNA isoforms, leaving an intact *Slc32a1* gene locus. Furthermore, using quantitative RT-PCR analysis of embryonic mouse brains, we showed that the expression of *Slc32a1* gene is not altered in the lncRNA mutants. Contrary to the general notion in the field that divergent lncRNAs regulate their adjacent genes, in this study, we provide evidence that the divergent *Gm14204* lncRNA does not regulate *Slc32a1* despite the fact that they share similar expression patterns. Additionally, this lncRNA also does not regulate other protein-coding genes in the *Gm14204* genomic region. Moreover, our smFISH data in adult brain have clearly indicated that *Gm14204* localizes to multiple locations inside the nucleus in a subset of neurons, suggesting a *trans* function of the lncRNA. Identification of these neurons with molecular markers might provide us with a deeper knowledge about the possible function of lncRNA.

Only a few lncRNA knockouts in mouse have been shown to be lethal whereas many others have reported the presence of subtler phenotypes. In our study, we found that genetic deletion of *Gm14204* was not lethal. The mice were viable, healthy, showed no visible abnormalities and were born at a normal Mendelian ratio. We still search for a subtle phenotype in the nervous system where *Gm14204* shows strong expression during mouse development and in the adult.

1.0 INTRODUCTION

The Central Dogma of life as first proposed by Francis H. Crick in his paper in 1970 (Crick, 1970) stated the transfer of information from DNA to proteins *via.* messenger RNA (mRNA). These mRNAs serve as a template for protein synthesis. Ribosomal and Transfer RNAs (rRNA and tRNAs) had already been discovered in the 1950s to play roles in protein synthesis. In the last 40 years, many other regulatory RNAs such as snoRNAs, small nuclear RNAs, and Piwi-interacting RNAs were discovered to have roles in gene regulation. The non-coding RNA field expanded during the last 18 years after the emergence of micro-RNAs (miRNAs). MiRNAs majorly have functions during the posttranscriptional events of gene regulation, particularly in higher eukaryotic organisms (reviewed in Cech & Steitz, 2014). A breakthrough came when deep sequencing results showed that the majority of an organisms' genome is transcribed (Carninci et al., 2005) and does not code for proteins. These transcripts were arbitrarily named as long non-coding (lnc) RNAs because they were longer than 200 nucleotides (nts.) and in order to differentiate them from small ncRNAs that were shorter than 200 nts.

1.1 The concept of Pervasive Transcription and discovery of Long noncoding RNAs

The notion that a large percentage of living organisms' genome is transcribed is known as 'Pervasive transcription'. This idea got high attention after the ENCODE project consortium provided convincing evidence that most of the human genome is pervasively transcribed (Birney et al., 2007). With the advent of massive RNA sequencing (RNA-seq) in the last decade, a few studies have suggested that mammalian cells produce a large number of large non-coding transcripts which were not been previously reported (Carninci et al., 2005; Kapranov et al., 2002). These transcripts were discarded earlier as transcriptional noise due to

their low abundance in numbers compared to protein-coding transcripts and also due to their low protein-coding ability (Kapranov et al., 2002). Although a couple of non-coding RNAs had already been studied in the early 90s, such as *Xist* lncRNA required for X-chromosome inactivation (Brockdorff et al., 1992; Brown et al., 1991) and *H19* – an imprinted lncRNA required for mouse development (Bartolomei, Webber, Brunkow, & Tilghman, 1993), not much attention was given to search for more noncoding transcripts in the genome. The discovery of lncRNAs was on a slow pace until the lncRNA *HOTAIR* was discovered to regulate the transcription of multiple genes on the 40 kb region of *HOXD* loci in human cell lines (Rinn et al., 2007). More recently, the existence of active transcription from the intergenic regions in several human and mouse cell lines came from looking for chromatin signatures, such as DNaseI hypersensitivity sites, transcription factors (TFs) binding sites and histone modification marks like H3K9ac, H3K4me3 and H3K36me3 on gene bodies (ENCODE et al., 2012; Guttman et al., 2009; Khalil et al., 2009). All these studies gave confidence for the presence of novel lncRNA genes in mammalian genomes. Subsequently, many biologically significant lncRNAs such as *Evf2* (Berghoff et al., 2013), *Fendrr* (Grote et al., 2013), *Bvht* (Klattenhoff et al., 2013), *linc-Brn1b*, *linc-Pint*, *Peril* etc. (Sauvageau et al., 2013) were discovered to have roles in mouse development.

1.2 Characteristics of long noncoding RNAs

The discovery of a plethora of lncRNAs has been achieved by identifying some of the hallmarks of lncRNA genes. The first identification is by looking for specific chromatin signatures. Chromatin signatures are combinations of histone modifications that correspond to active gene transcription. The tri-methylation of histone H3 on lysine 4 (H3K4me3) is a hallmark for active gene promoters and the tri-methylation of histone H3 on lysine 36 corresponds to active transcription of the entire gene body. This is called the 'K4-K36' domain (Mikkelsen et al., 2007). Such signatures used to look for active transcription outside of protein-coding genes revealed nearly 1600 regions in mouse and 2500 regions in the human genome that were actively transcribed and producing lncRNAs (Guttman et al., 2009; Khalil et al., 2009). Moreover, chromatin state maps revealed enhancer regions containing short stretches of mono-methylation of histone H3 on lysine 4 (H3K4me1) which give rise to a number of lncRNAs, generally named as enhancer lncRNA or eRNAs in short (Ørom et al., 2010).

The second major characteristic of a lncRNA gene is its position in the genome with respect to protein-coding genes. LncRNAs are classified into different broad categories depending on their overlap or non-overlap with coding genes (Kung, Colognori, & Lee, 2013; Ma, Bajic, & Zhang, 2013; Rinn & Chang, 2012) (**Fig. 1**):

1) **Sense overlapping Transcripts or Sense lncRNAs:** LncRNA genes overlapping partially the protein-coding genes and transcribed in the same direction are called Sense lncRNAs. e.g.: *H19/H19 upstream 1 and 2* transcripts.

2) **Natural Antisense Transcripts (NATs):** LncRNA genes overlapping other coding genes but are transcribed in the opposite direction than latter ones are known as NATs. These lncRNAs tend to be enriched mostly near the 5' (promoter) or 3' (terminator) ends of coding genes. e.g.: *Kcnq1/Kcnq1ot1* gene pair, *Igf2r/Airn* gene pair, etc.

3) **Intergenic lncRNAs (LincRNAs):** These are the most abundant lncRNA forms present in the genome. LincRNAs do not overlap other genes but are present at random locations at a substantial far away distance from coding genes. Studies have shown that nearly 70% of lncRNAs are intergenic type, e.g.: *MALAT1, HOTAIR* lncRNAs etc.

4) **Intronic lncRNAs:** Many of the introns of protein-coding genes have been found to harbor long noncoding transcripts, but only a few of those have been studied to-date.

5) **Divergent lncRNAs:** LncRNAs transcribed from the vicinity of transcription start sites of protein-coding genes in an antisense direction are termed divergent lncRNAs. The distance between the TSS of both genes is less than 1 kb. This category of lncRNAs is known to regulate the expression of their adjacent protein-coding genes. Moreover, it is argued that these lncRNA-mRNA gene pairs share similar spatiotemporal expression profiles in various tissues (Grote et al., 2013; Liu et al., 2017; Sigova et al., 2013; Wu & Sharp, 2013).

Finally, the last characteristic of lncRNAs is their inability to code for proteins. Determining whether any gene produces coding or non-coding transcripts is very challenging because of small functional peptides encoded by such transcripts that are often missed by large-scale proteomic approaches (Dinger, Pang, Mercer, & Mattick, 2008). Experimental techniques, for instance, ribosomal profiling have provided evidence that majority of lncRNAs do not code for proteins. This method measures the ribosome occupancy on RNAs, thus differentiating them into coding or non-coding (Guttman, Russell, Ingolia, Weissman, & Lander, 2013). Computational methods such as 'PhyloCSF (codon substitution frequency)' algorithm used by UCSC genome browser or 'Coding Potential assessment tools (CPAT)' (L. Wang et al., 2013) are now been widely used to check if a lncRNA encodes short peptides. In addition, the

functional importance of such peptides in lncRNAs, if found, needs proper evaluation and experimentation *in vivo*.
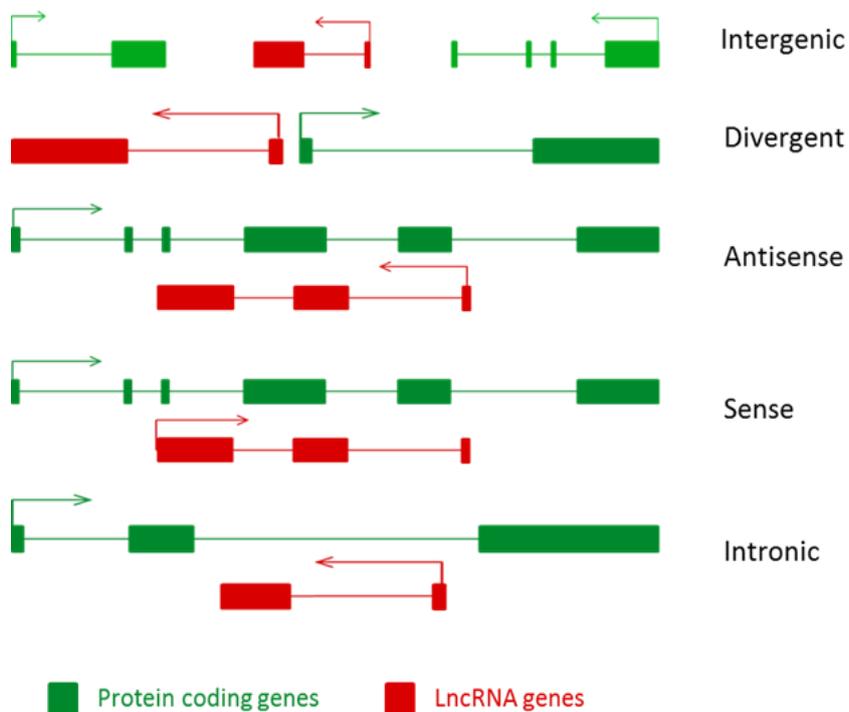

Figure 1: Classification of lncRNAs based on their location in the genome. LncRNAs classified based on their position with respect to protein-coding genes. Reference: (Rinn et al., 2007; S. U. Schmitz, Grote, & Herrmann, 2016).

1.3 Evolution and Conservation of Long noncoding RNAs

The evolutionary changes in gene sequences especially for protein-coding genes and changes in their expression patterns are known to underlie differences in lineage-specific phenotypes between different species. On the other hand, since only a small fraction of long noncoding RNAs have been well characterized, the evolutionary analysis of these sequences still remains poorly understood. The conservation of lncRNAs across different species can provide some clues on their function, but the absence of lncRNA annotations in several organisms so far provides an obstacle in carrying out such analysis. In recent years, several studies have attempted to perform large-scale evolutionary comparisons using available RNA-seq data from various organisms. The first study performed by Ulitsky, Shkumatava, Jan, Sive, & Bartel, (2011) showed that few intergenic lncRNAs (lincRNAs) are highly conserved from zebrafish to mammals. Using phastCons scores from the UCSC genome browser, they showed that the exons of lincRNAs were less conserved than mRNA exons and UTRs, but more conserved than mRNA introns. This observation was consistent with previous reports for mammalian lincRNAs (Guttman et al., 2009; Khalil et al., 2009; Ponting, Oliver, & Reik, 2009). Necsulea et al., (2014) attempted to determine the lncRNA evolutionary conservation across 11 tetrapod species. They identified lncRNAs that had originated nearly 300 Myr (million years) ago and had a tendency to evolve rapidly. Assessing the conservation of expression pattern of lncRNAs, they observed that the transcription and expression of lncRNA genes evolved rapidly across species as compared to coding genes. Moreover, the turnover rates for lncRNAs are also more rapid than protein-coding genes, as observed in previous studies mentioned above. Interestingly, their ChIP-seq transcription factor (TF) binding data showed that lncRNA promoters bound TFs more frequently than random intergenic regions and the binding site sequence conservation was more than for protein-coding gene promoters. Apart from lncRNA exons showing less conservation than mRNA

exons, the promoters of lncRNAs are on average more conserved than their exons and almost similarly conserved as promoters of coding genes.

All these studies, along with others (Guttman et al., 2009; Ponjavic, Ponting, & Lunter, 2007) indicate that the evolution of lncRNA sequences tends to be weakly constrained compared to coding gene sequences and that lncRNAs are frequently under constant constraint. Certainly, it is evident that the number of lncRNAs has increased during animal evolution, indicating that growing numbers of lncRNAs are linked to organism complexity (Hezroni et al., 2015; Necsulea et al., 2014).

1.4 Long noncoding RNAs in gene regulation

Long noncoding RNAs now are considered the most diverse and largest group of nucleic acid-based molecules in the mammalian transcriptomes. Several lncRNAs are shown to function in a wide range of cellular and biochemical processes with different modes of action. This section will mention a few mechanisms of lncRNAs reported so far.

- **Modes of transcriptional regulation by lncRNAs**

Transcriptional regulation involves an interplay of the cell and tissue-specific transcription factors (TFs) and chromatin remodeling factors, which collectively act on promoters and enhancers to facilitate the assembly of transcriptional machinery on gene promoters to regulate gene transcription. LncRNAs could regulate protein-coding gene transcription positively or negatively and in *cis* or in *trans*.

- **LncRNA regulation in *cis***

When lncRNAs regulates the transcription of protein-coding genes on the same chromosome then they are termed as *cis*-acting lncRNAs. These lncRNAs can act *via* two different modes. The first one is by lncRNA product. A classic example of this category is the X-inactivation specific transcript or *Xist* lncRNA. *Xist* is expressed from either of the two X chromosomes in females and induces the silencing of the whole chromosome required to maintain dosage compensation in mammals (reviewed in Lee, 2009). Many recent studies have shown that *Xist* lncRNA binds various proteins (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015) to carry out silencing of X-chromosome. *Xist* has various A-repeat elements, which are required for transcriptional silencing of genes on X-chromosome (Hoki et al., 2009) possibly by recruiting several polycomb proteins (PRC2) (Zhao, Sun, Erwin, Song, & Lee, 2008). *Xist* is known to alter the nuclear three-dimensional architecture of the X-chromosome by spreading over the entire chromosome (Engreitz et al., 2013). Another example of lncRNA acting *via* their products is human *HOTTIP* lncRNA. *HOTTIP* is expressed in the *HOXA* cluster whereby it activates the transcription of flanking genes by binding to WDR5 in the MLL histone modifier complex. This whole complex then recruits H3K4me3 (histone H3 lysine 4 tri-methylation) on active gene promoters (K. C. Wang et al., 2011). *Fendrr* lncRNA, which recruits polycomb protein PRC2 to regulate the expression of its target genes involved in the development of lateral plate mesoderm. *Fendrr* lncRNA transcript binds and delivers PRC2 to the promoter of *Foxf1* gene on the same chromosome to fine-tune *Foxf1* expression (Grote et al., 2013).

The second mode of regulation is *via* lncRNA transcription itself. Transcription mediated silencing is defined as the act of transcription of one gene can repress in *cis* the functional transcription of another gene (Kornienko, Guenzl, Barlow, & Pauler, 2013). LncRNA transcription could cause silencing of protein-coding genes by depositing

nucleosomes in an unfavorable manner for TFs binding on gene promoters. For example, an eukaryotic *SRG1* lncRNA silences *SER3* coding gene by transcriptional overlap. *SRG1* transcription increases the number of nucleosomes at the *SER3* promoter, repressing its transcription (Martens, Laprade, & Winston, 2004). Another example of transcriptional interference is caused by *Airn* lncRNA. This lncRNA is inherited on paternal chromosomes and overlaps the *Igf2r* protein-coding gene (Latos et al., 2012) that is an essential embryonic growth suppressor (Ludwig et al., 1996). *Airn* knockout experiments in mice have shown that its transcriptional overlap interfered with the accumulation of functional RNAPII on the *Igf2r* promoter, repressing its transcription. Thus suggesting that *Airn* transcription is continuously required for *Igf2r* silencing (Latos et al., 2012).

- **LncRNA regulation in *trans***

Certain lncRNAs products can regulate the transcription of protein-coding genes on distant chromosomes. Regulation in *trans* can act directly on one specific gene or on a set of locus-specific genes. For instance, a 331-nucleotide 7SK lncRNA sequesters the positive transcriptional elongation factor (PTEF-b) and prevents it to phosphorylate the RNAPII carboxy-terminal domain, thereby preventing elongation of several genes in embryonic stem cells (Castelo-Branco et al., 2013). *HOTAIR* lncRNA expressed from the *HOXC* cluster represses the transcription of 40 kb genes in the *HOXD* cluster on a different chromosome (Rinn et al., 2007). *HOTAIR* physically interacts with a polycomb protein PRC2 and regulates the localization of H3K27me3 on hundreds of sites on the genome (Rinn et al., 2007; Tsai et al., 2010). *Fendrr* lncRNA also acts in *trans* to regulate the transcription of the *Pitx2* gene by binding to PRC2 and TrxG/MLL protein complexes (Grote et al., 2013).

- **LncRNAs targeting chromatin regulators**

Throughout the years since the discovery of RNA molecules, many studies have demonstrated that RNA is an essential component required for proper structural organization of chromatin and recruitment of chromatin-modifying complexes to DNA (Bernstein & Allis, 2005). However, until the last decade, the specific RNA molecules associated with this interaction remained elusive. In the early 90s, several genetic studies revealed a few lncRNAs that were involved in heterochromatin formation {*Xist* lncRNA (Brockdorff et al., 1992; Brown et al., 1991)} and imprinting {*H19* lncRNA (Bartolomei et al., 1993)}. In the last decade, several studies have reportedly shown that lncRNAs can associate with many chromatin modifying complexes involved in gene activation and gene silencing (Marchese, Raimondi, & Huarte, 2017; Rinn & Chang, 2012; Vance & Ponting, 2014). LncRNAs can deposit H3K4me3 marks on gene promoters to activate and maintain the transcription of protein-coding genes (K. C. Wang et al., 2011) or lncRNAs can bind DNA methyltransferases such as DNMT1 and DNMT3b to repress transcription (Mohammad, Mondal, Guseva, Pandey, & Kanduri, 2010; K. M. Schmitz, Mayer, Postepska, & Grummt, 2010). Apart from all these, the polycomb complex PRC2 has been the most studied chromatin complex regarding its relation to the lncRNAs regulating gene expression.

Polycomb repressive complex 2 (PRC2), a histone methyltransferase is essential for epigenetic silencing during development. Thousands of lncRNAs have been shown to associate with PRC2 *in vivo* and this has gained much attention during the last decade (Guttman et al., 2009; Khalil et al., 2009; Marchese et al., 2017). The lncRNA *HOTAIR* expressed from the *HOXC* gene locus in mouse and humans interferes with the transcriptional activation of genes spanning > 40 kb in the *HOXD* gene locus. RNA immunoprecipitation (RNA-IP) experiments demonstrated that this lncRNA binds to PRC2 and deposits H3K27me3 repressive marks on the *HOXD* locus. This mechanism of action of lncRNA was

proved to act in *trans* (Rinn et al., 2007). Another imprinting specific lncRNA that showed this interaction was *Kcnq1ot1*. Transcribed from the paternal chromosome in mouse, this lncRNA binds PRC2 and is tethered to chromatin, where it deposits repressive histone marks, thereby silencing multiple genes in the 1 Mb domain of *Kcnq1* locus (Pandey et al., 2008). Many other well-studied lncRNAs clearly show how PRC2 binding to lncRNAs is essential to carry out biochemical processes. An in-depth review of mechanisms of PRC2-lncRNA action can be found in (Davidovich & Cech, 2015).

1.5 Sub-cellular localization of Long noncoding RNAs

The spatiotemporal expression profile and cellular localization of lncRNAs are of utmost importance in hypothesizing their mechanism and potential role in a biological context. One can argue that unlike mRNAs, which are exported to the cytoplasm for their translation, lncRNAs should localize in the nucleus since they lack a coding ORF. This is true for the majority, but not for all lncRNAs. A few examples described here demonstrate that lncRNAs can localize in diverse cytoplasmic and nuclear spaces from where they exert different modes of action. An early example was *Xist* lncRNA, a key regulator of X-inactivation (Brockdorff et al., 1992; Brown et al., 1991). *Xist* was discovered to localize on the X-chromosome inside the nucleus (Clemson, McNeil, Willard, & Lawrence, 1996). One study identified a novel lncRNA specifically expressed in the post-mitotic neurons, called *Gomafu*. *Gomafu* was identified to localize in a novel nucleoplasm compartment, which did not co-localize with known nuclear domain markers (Sone et al., 2007). Another well-studied example is lncRNA *MALAT1*, which was shown to localize in nuclear speckles and to interact with many pre-mRNA splicing factors. *MALAT1* regulates cellular levels of phosphorylated SR (serine/arginine) proteins thereby regulating alternative splicing (Tripathi et al., 2010). Several other lncRNAs such as *Evf2* (Berghoff et al., 2013), *Fendrr* (Grote et al., 2013),

Neat1 (Souquere, Beauclair, Harper, Fox, & Pierron, 2010), etc. are shown to localize strictly inside the nucleus. LncRNAs can exhibit diverse expression patterns inside a cell nucleus. A detailed study performed by M. N. Cabili et al., (2015) showed a wide range of localization patterns inside the nuclei and cytoplasm using a novel technique called single-molecule RNA FISH (fluorescent *in situ* hybridization) (**Fig. 2**). A few lncRNAs are found to be exported in the cytoplasm where they bind to ribosomal complexes and the translational machinery of coding mRNAs. Since cytoplasmic lncRNA topic is outside the scope of this project, detailed reviews and original papers can be found elsewhere (Rashid, Shah, & Shan, 2016; Van Heesch et al., 2014).
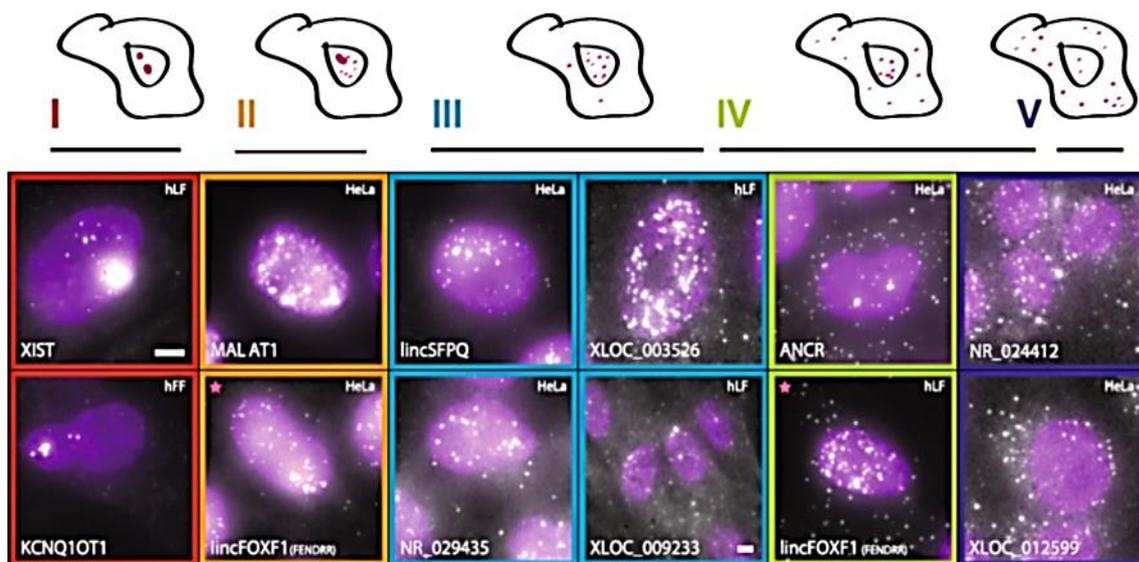

Figure 2: Sub-cellular localization of lncRNAs. Top panel: Scheme of lncRNA localized inside a cell. The color-coded numbers correspond to the colored boxes in the images. **I**: Only two lncRNA spots observed inside the nucleus; **II**: Multiple lncRNA foci inside the nuclei; **III**: Majority of lncRNA spots are localized to the nuclei, but a few lncRNA transcripts are transported to the cytoplasm; **IV**: Equal distribution of lncRNAs in the nucleus and cytoplasm; **V**: Cytoplasmic lncRNAs. Bottom panel: examples of different lncRNAs and their distribution in HeLa and hLF (human lung fibroblasts) cell lines. *DAPI* was used to stain the nucleus. Scale bar: 5 µm. Picture modified from (M. N. Cabili et al., 2015).

1.6 Long noncoding RNAs involved in mouse development and the scope of this project

In the last decade, much evidence has been provided by different studies that lncRNAs exert important functions through a wide range of mechanisms. However, the majority of this evidence has been deduced in cell culture models and only a handful of lncRNA functions are studied in the mouse by generating genetic knockout mouse models. A major reason was effective targeting strategies were not developed to knockout lncRNAs *in vivo*. Many lncRNAs either overlap important protein-coding genes or there are essential transcriptional regulatory elements overlapping the lncRNA gene (Li & Chang, 2014), making lncRNA deletions difficult. One group attempted to generate 18 knockout mouse lines by replacing the entire lncRNA gene with *lac*Z reporter cassette (Sauvageau et al., 2013). It was observed that for 13 of these lncRNA lines no strong phenotypes were reported and the mice were viable and fertile. For other five-lncRNA knockouts, the group reported mild to strong phenotypes. For instance, lncRNAs *Fendrr* and *Peril* deletions showed no/reduced viability. They observed embryonic growth defects in *linc-Pint*$^{-/-}$ and *Mdgt*$^{-/-}$ mice, and abnormal cortical development in *linc-Brn1b*$^{-/-}$ mice. A conclusion drawn by the authors of this study was these defects were due to the lncRNA deletions only. Although one cannot rule out a strong possibility that such strong phenotypes might have occurred by deletion of *cis*-regulatory elements present in the lncRNA gene loci that are essential for normal function of proximal protein-coding genes (Bassett et al., 2014; Li & Chang, 2014).

Other approaches include either deleting the promoter of lncRNA gene or by promoter inversion. Such strategies so far are less proven to knockout complete lncRNA transcripts because of bidirectionality of the promoter. Additionally, if lncRNA and its adjacent mRNA gene share same promoter sequence then promoter deletions can disrupt the expression levels of that protein-coding gene. Moreover, lncRNA genes can have alternative promoters and in

that case, one promoter deletion might not be successful. All the genetic strategies used to study lncRNA function so far have been described by Bassett et al., (2014) and Li & Chang, (2014).

Another experimental approach that is well used to knockout lncRNA transcripts is the integration of polyadenylation cassettes (polyA) in the lncRNA gene loci. Successful lncRNA terminations were possible when the polyA sites were introduced close to transcription start sites (TSS) of lncRNA genes. This mechanism prematurely terminates lncRNA transcription, thus producing no lncRNA product (Grote et al., 2013; Latos et al., 2012; Sleutels, Zwart, & Barlow, 2002) and yet leaving intact gene loci. Until now, only a handful of lncRNA functions have been investigated *in vivo*. With the growing number of lncRNAs present in the mammalian genome, many questions arise regarding the mechanistic roles of such transcripts. It is already clear from the above-cited literature that lncRNAs, although a few are as important as protein-coding genes to fine-tune several cellular processes. Therefore, this study aims to discover novel, putative lncRNAs expressed during mouse development and generate a knockout mouse model for a better understanding of lncRNA mechanisms *in vivo*.

1.7 Objectives of this project

In the present study, we first began to explore how many lncRNAs were expressed during the E14.5 stage of mouse embryonic development. E14.5 stage was chosen because at this time point organogenesis and neurogenesis are mostly complete and the majority of key brain development events have already started to occur (Molyneaux, Arlotta, Menezes, & Macklis, 2007). To-date there is no such existing compilation of lncRNAs expressed in E14.5 mouse embryos. Keeping this in mind, we wanted to develop a comprehensive bio-resource of E14.5 lncRNAs. To do this, we performed RNA-sequencing (RNA-seq) of whole E14.5 embryos and subjected the raw data under a stringent lncRNA selection pipeline (see details in Result section). We next combined a systematic approach filtering them based on location on physiologically important and conserved gene loci. Additionally, we focused primarily on studying the mechanism of a divergent class of lncRNAs since in the past years several studies have pointed out their importance in regulating the expression of their neighboring protein-coding genes (Grote et al., 2013; Sigova et al., 2013; Vance et al., 2014).

Moreover, in recent years, many studies have focused on lncRNAs having roles during mouse Central Nervous System (CNS) development. Genomic loci of many lncRNAs were present in close proximity of protein-coding genes involved in CNS development (reviewed in Briggs, Wolvetang, Mattick, Rinn, & Barry, 2015; Roberts, Morris, & Wood, 2014). A study performed by Mercer, Dinger, Sunkin, Mehler, & Mattick, (2008) using RNA *in situ* hybridization data from Allen Brain Atlas found that many lncRNAs are associated with coding genes linked to neuroanatomical development. Realizing the wealth of lncRNA information present in mouse nervous system, we focused our attention on divergent lncRNAs expressed in mouse E14.5 CNS. We selected a handful of lncRNAs using two main criteria: (i) mammalian syntenic conservation of gene loci and (ii) lncRNAs divergently transcribed to physiologically important protein-coding genes. In the next step, RNA *in situ* hybridization in

E14.5 mouse embryos was performed to elucidate the expression profile of these selected lncRNAs. Our first aim was to validate whether or not divergent lncRNAs shared similar and overlapping expression patterns with their adjacent protein-coding mRNAs. If divergent lncRNA-mRNA pairs share similar expression in different tissues, then we can hypothesize that such gene pairs might involve sharing of common transcriptional network and a lncRNA could regulate the expression of its adjacent mRNA gene.

Our second aim was to generate a genetic knockout mouse model for a novel and previously uncharacterized lncRNA gene called *Gm14204* that is divergently transcribed to an important mammalian neurotransmitter transporter gene *Slc32a1* in the mouse. *Slc32a1* (VGAT, Viaat) codes for a neurotransmitter transporter, which is responsible for the GABA uptake into the synaptic vesicles (McIntire, Reimer, Schuske, Edwards, & Jorgensen, 1997). Loss of Viaat in mice causes a drastic reduction of neurotransmitter release in the GABAergic neurons. Viaat knockout mice die between E18.5 and birth and the embryos at an earlier stage display several developmental defects (Wojcik et al., 2006). There is less knowledge about the transcriptional regulation of *Slc32a1* gene. Our hypothesis was that *Gm14204* lncRNA might regulate the expression of *Slc32a1* at a transcriptional level. If this were true, the lncRNA knockout mice would help us better understand the mechanism.

Furthermore, *Gm14204* knockout mouse model will help us to investigate other roles of lncRNA in CNS development and function.

2.0   Results

## 2.1 Part 1: Screening for long non-coding RNAs

2.1.1 Identification of long non-coding RNAs expressed in developing mouse stage E14.5

In order to obtain a comprehensive list of lncRNAs expressed at E14.5 mouse developmental stage, we extracted total RNA from three whole embryos at E14.5 and conducted single-stranded strand-specific RNA-sequencing. The gene models of lncRNAs were *de novo* assembled from RNA-seq data based by the methods as described in the study of M. Cabili et al., (2011). To reduce transcriptional noise and ensure high quality of transcripts, single-exonic lncRNAs were eliminated. In order to achieve better annotation of lncRNAs, we integrated another set of mouse lncRNAs assembled from public RNA-seq data (Ensembl, RefSeq, and UCSC), in order to remove repetition of lncRNAs. After eliminating redundancy, approximately ~7000 multi-exonic lncRNAs (RPKM $\geq$ 0.1) have been assembled. Notably, lncRNAs from other public databases were also found in our RNA-seq dataset, thus confirming a high quality of our sequencing analysis (**Fig. 3 scheme**).

Based on their genomic locations relative to protein-coding genes, these lncRNAs were divided into 5 categories (Rinn & Chang, 2012): Divergent (774), Intergenic (3292), Sense overlapping (1072), Antisense overlapping (1445) and Intronic (640) (numbers indicate total lncRNAs within each category) (**Fig. 3**). The expression values obtained from RNA-Seq were in terms of RPKM (Reads Per Kilobase of exons per Million mapped reads). The intergenic and divergent class of lncRNAs have an advantage that they do not overlap any protein-coding genes, which facilitates their manipulation for the establishment of knockout models for functional analysis. In our study, we have focused primarily on the divergent class of lncRNAs. Divergent lncRNAs have been implicated to regulate their activity of

transcription of their adjacent coding partners through binding to different transcription factors or regulatory proteins (Grote et al., 2013; Latos et al., 2012; Vance et al., 2014). Because of the high number of divergent lncRNA transcripts (>700) obtained from our sequencing data, further criteria were applied to narrow down lncRNA candidates that will be used for further analysis. First, we looked for divergent lncRNAs that were highly conserved amongst mammalian species. Next, from this conserved lncRNA list, we selected those lncRNAs which were in close proximity to important protein-coding genes that encode regulators of mouse development and physiology. Finally, using an unbiased approach we randomly selected 10 divergent lncRNAs which were next to regulatory protein-coding genes and performed RNA *in situ* hybridization (RNA-ISH) on mouse embryonic E14.5 sagittal sections. To visualize the expression of their adjacent protein-coding genes, we made use of the GenePaint database (www.gp3.mpg.de) and compared the expression of lncRNA-mRNA gene pairs. Finally, we have narrowed down to one lncRNA which resides on a mammalian conserved genomic locus for generating mouse knockout model and studying its physiological role during mouse development.
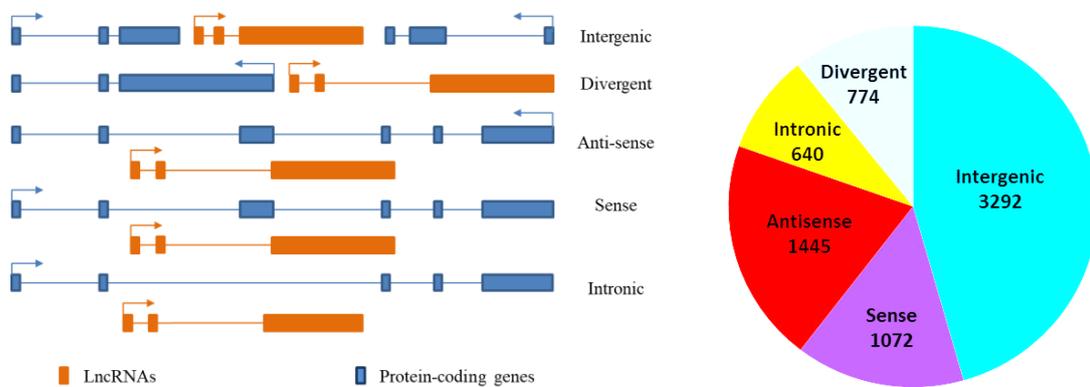
a
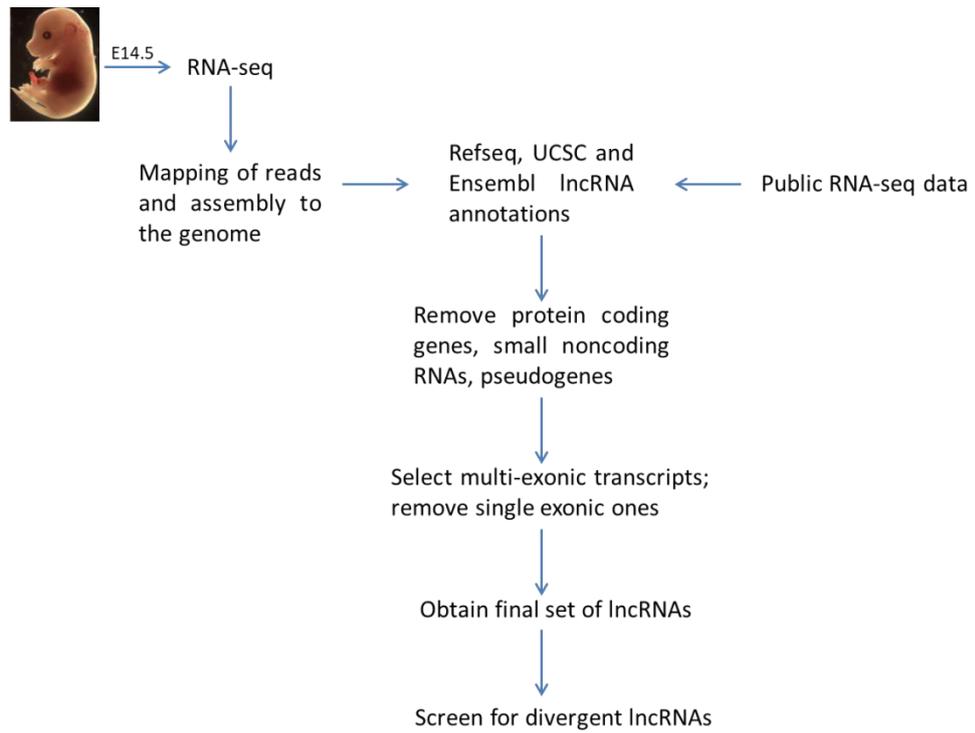
b

Figure 3: Identification of lncRNAs expressed in the E14.5 mouse embryo.

(a) Schematic workflow of *de novo* assembly of RNA-seq data of mouse embryos (n=3 embryos) and integration with public annotation of lncRNAs (see text for details), (b) Classification of lncRNAs based on their position in the genome with respect to protein-coding genes as described in (Rinn & Chang, 2012). The pie-chart shows the number of lncRNAs within each category found in our lncRNA sequencing data.

2.1.2 Genomic and transcriptional co-localization of divergent lncRNA-mRNA pairs in the developing mouse embryos

There is evidence that adjacent lncRNA-mRNA gene pairs give rise to separate transcripts which share similar spatiotemporal expression patterns in mouse (Ponjavic, Oliver, Lunter, & Ponting, 2009). In this study, we have performed non-radioactive RNA *in situ* hybridization for 10 divergent lncRNAs and compared the expression patterns with the expression of protein-coding genes, published in the GenePaint database. The selections of lncRNAs were based on the criteria that they are in the vicinity of important transcription factors (TFs) or genes with important developmental and physiological roles and are conserved in vertebrates.

Interestingly, our analysis indicated that most of the divergent lncRNAs share similar expression with their coding gene pairs. For example in Figure 4, we found that previously uncharacterized lncRNAs *Gm11266*, *9130024F11Rik,* and *Cuff.552755* shared high identical expression with transcription factors *Nfib* (nuclear factor I B), *Satb2* (Special AT-rich sequence binding protein 2) and *Mll5* (lysine methyltransferases), respectively. *Gm11266* and *Nfib* have similar expression pattern in the developing mouse central nervous system including E14.5 neocortex, septum and the spinal cord (**Fig. 4**). The lncRNA *9130024F11Rik* and *Satb2* TF are expressed in the E14.5 cortex, axial skeleton and Meckel's cartilage (**Fig. 4**) while lncRNA *Cuff.557255* and *Mll5* TF have a similar expression in the superior colliculus, dorsal root ganglion and the cerebellum (**Fig. 4**). The *Fendrr* lncRNA has been shown to have an important role during embryonic development and its knockout leads to embryonic lethality (Grote et al., 2013; Sauvageau et al., 2013). We found that this lncRNA was present in our RNA-seq data and shares similar expression patterns with its adjacent *Foxf1* TF in the lung, intestine, and gut (**Fig. 4**).
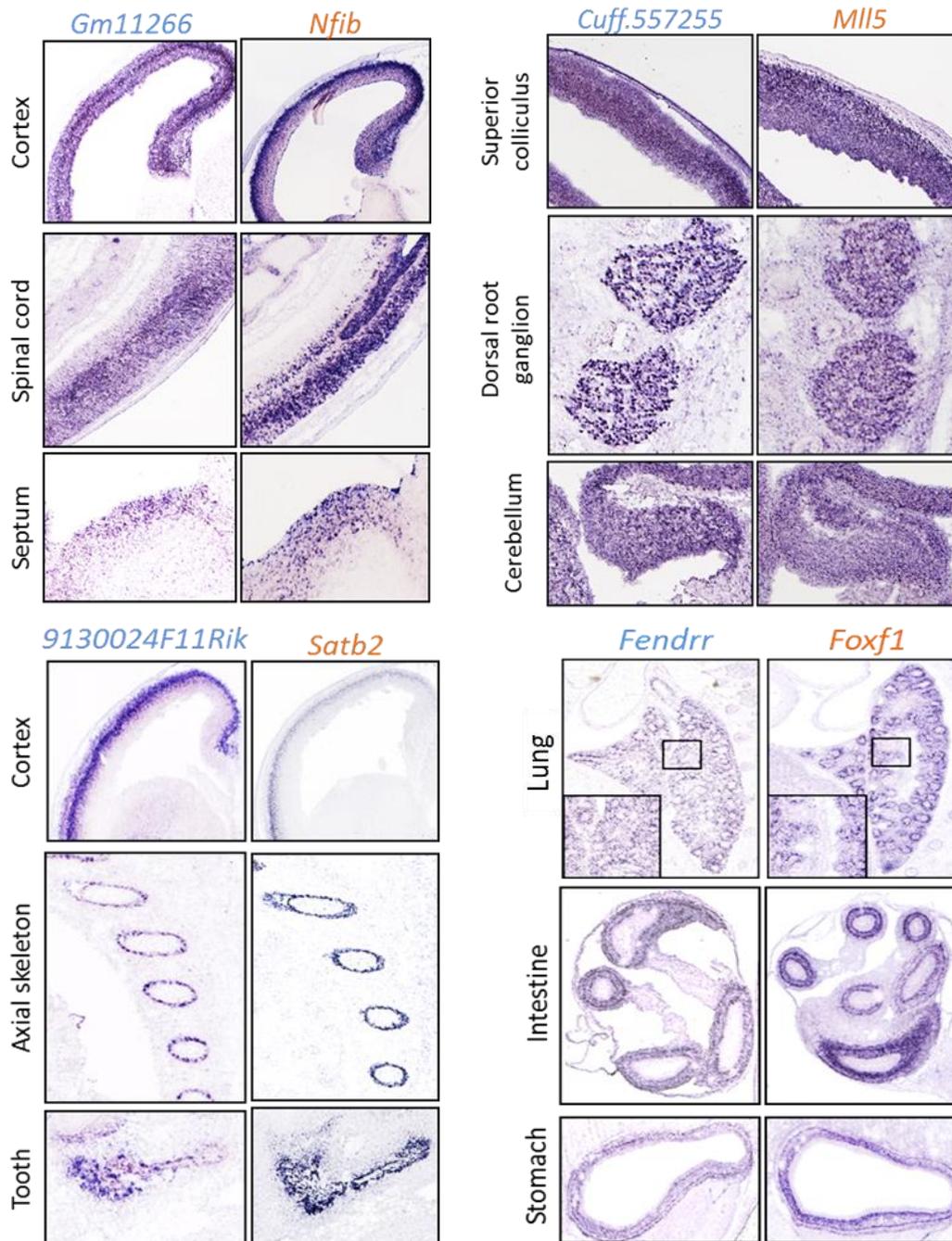
Figure 4: RNA *in situ* hybridization of divergent lncRNA-mRNA pairs in E14.5 mouse embryos.

Expression of *Gm11266* lncRNA and its coding partner *Nfib* transcription factor as observed in neopallial cortex, spinal cord and developing septum. Expression of *9130024F11Rik* lncRNA and its coding partner *Satb2* transcription factor in the cortical plate, axial skeleton and Meckel's cartilage. Expression of *Cuff557255* lncRNA and its coding partner *Mll5* transcription factor in the superior colliculus, dorsal root ganglion, and the cerebellum. Expression of *Fendrr* lncRNA and its coding partner *Foxf1* transcription factor in the lung, intestine, and stomach.

2.1.3 The lncRNA *Gm14204* expression pattern is similar to the mammalian inhibitory neurotransmitter transporter gene *Slc32a1*

Divergent lncRNA-mRNA gene pairs share substantially high degree of similar expression with their coding gene partner because they might be transcriptionally co-active on the genomic loci or even these lncRNAs could potentially regulate the expression and functions of adjacent coding genes. The latter phenomenon has been experimentally verified in studies of lncRNAs with important roles for mouse development such as *Fendrr* (Grote et al., 2013), *Evf2* (Berghoff et al., 2013) and *lnc-Kdm2b* (Liu et al., 2017).

Amongst the identified divergent lncRNAs from our screen was one novel lncRNA *Gm14204* with unknown characteristics so far. It is located on the reverse strand of chromosome 2 and in close vicinity to *Slc32a1* protein-coding gene also known as Solute Carrier Family 32 Member 1 (or Vesicular GABA Transporter (VGAT) / Vesicular Inhibitory Amino Acid Transporter (Viaat)) (UniProtKB: Q9H598). This transporter is responsible for GABA and Glycine uptake into the synaptic vesicles. *Gm14204* lncRNA is located just 50 bp upstream of *Slc32a1* and is transcribed divergently (**Fig. 5a**). Our RNA-seq data shows that the lncRNA has 4 different isoforms of different lengths (**Fig. 5a**) on the contrary to the NCBI database that shows only 2 isoforms (not shown). The lncRNA gene extends over 15 kb region on the genomic locus and the longest pre-dominant isoform is 5401 bp in length. *Gm14204* lncRNA is 5'-capped and 3'-polyadenylated which is one of an important characteristic of lncRNAs (Source: NCBI). The promoter region of *Gm14204* shows high H3K4me3 marks and the gene body shows high H3K36me3 marks which indicate that the gene is actively transcribed (**Fig. 5b**). Since lncRNAs have undergone high evolutionary changes, they often tend to be less conserved across species. To check for the conservation of *Gm14204*, we made use of Multiz Alignment of 60 vertebrates from the UCSC genome browser and saw that the lncRNA is conserved across mammals, with 5' region of lncRNA

showing a high degree of conservation (**Fig. 5c**). Apparently, we found conservation of whole lncRNA gene only in Rat. In other mammals, only the first exon and intron of the lncRNA gene are well-conserved with the remaining genomic regions showing a lesser degree of conservation. We did not find *Gm14204* conserved across non-mammalian species.

Recent evidence has shown that lncRNAs might code for small proteins or micro peptides which are less than 100 amino acids and might have previously escaped protein analysis primarily due to overlooking small ORFs in RNA transcripts (D. M. Anderson et al., 2015; Nelson et al., 2016). To investigate whether *Gm14204* does code for any proteins, we used the public available Coding-Potential Assessment Tool to assess the coding potentiality of the lncRNA. **Fig. 5d** shows *Gm14204* does not code for any known/unknown proteins, *Xist* and *Malat1* lncRNAs used as controls which also do not encode any peptides. *Gapdh* and *Slc32a1* were used as positive controls.
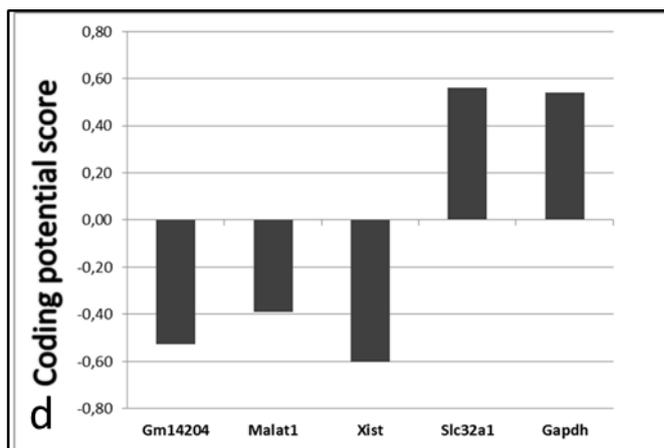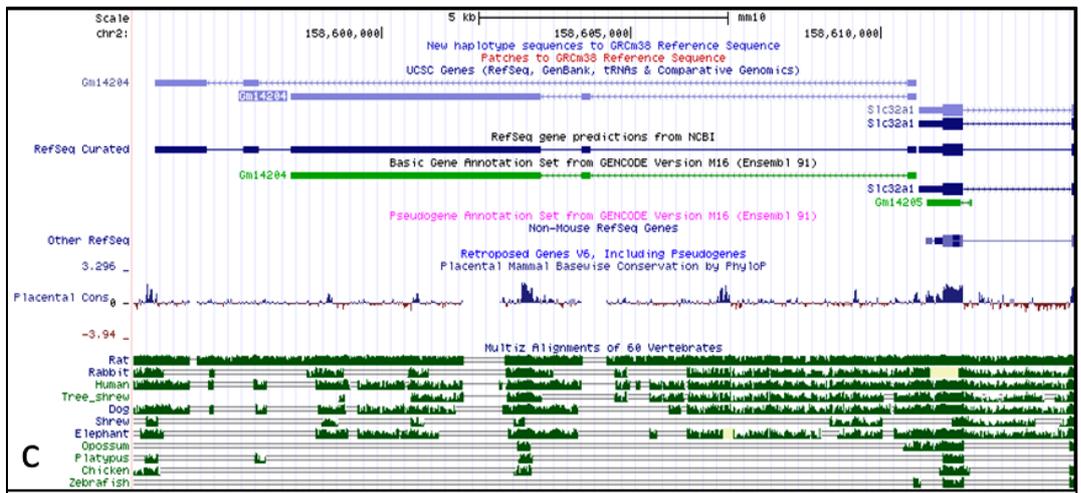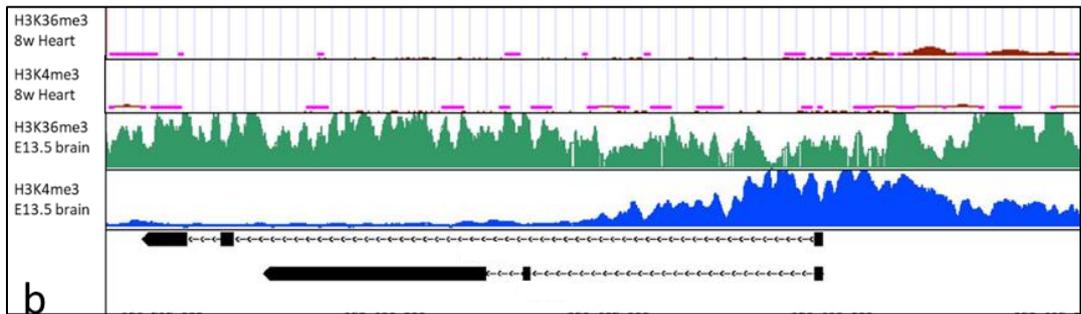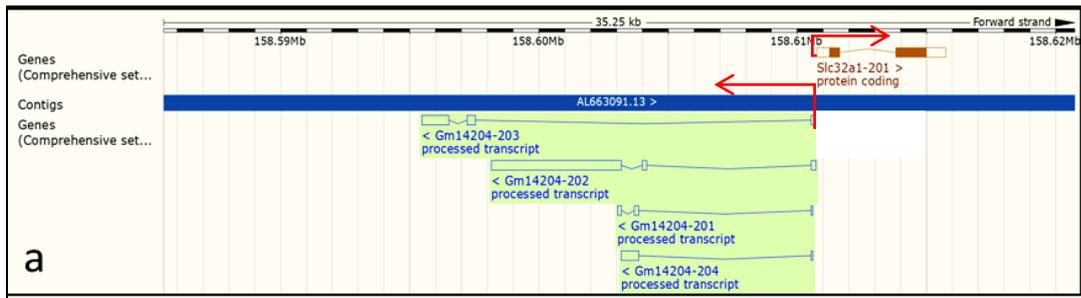
Figure 5: Characteristics of *Gm14204* lncRNA.

(a) Genomic organization of *Gm14204* long noncoding RNA. *Gm14204* is located on chromosome 2 and is transcribed on the reverse strand and divergently (red arrows) of *Slc32a1* protein-coding gene. The distance between *Gm14204* and *Slc32a1* is 50 bp. The lncRNA has different isoforms of varying lengths, which are alternatively spliced. (b) Transcriptional hallmarks of *Gm14204* lncRNA. The promoter region of lncRNA has high H3K4me3 marks (blue) and the gene body has high H3K36me3 marks (green) which shows that this gene is actively transcribed and expressed in the mouse brain, but not in the heart. (c) The evolutionary conservation of *Gm14204* lncRNA over its entire length as depicted using the University of California, Santa Cruz (UCSC) genome browser. The 5' region of the lncRNA is highly conserved than the rest of the gene body (green blocks). (d) *Gm14204* lncRNA does not code for any known proteins or micro peptides as determined by Coding-Potential Assessment Tool (L. Wang et al., 2013). A negative score indicates the absence of any coding ORF. *Gapdh* and *Slc32a1* protein-coding genes were used as positive controls. *Xist* and *Malat1* lncRNA do not code for any proteins or micro peptides.

2.1.4 Spatiotemporal expression profiling of *Gm14204* lncRNA during mouse development

In order to determine when the expression of *Gm14204* begins during development, embryonic brains (E11.5, E12.5, E13.5, E15.5, and E17.5) and adult mouse brain (P56, postnatal day stage 56) were collected for total RNA isolation and subsequent quantitative reverse transcriptase PCR (qRT-PCR) analysis. The results indicated the expression of lncRNA is evident from E11.5 brain and the expression gradually increases as embryogenesis progresses. The maximum expression was observed in E15.5 brain but thereafter starts to decrease. At this stage, low expression was also detectable in the eye. Notably, in adult mouse (P56), significant expression was detectable only in brain and spleen. No lncRNA expression was observed in the embryonic (E14.5) or adult (P56) heart, liver and kidney (**Fig. 6a**). Remarkably, a very similar expression was also detectable for the adjacent coding gene *Slc32a1* mRNA showing a steep increase from E11.5 brain to E17.5 brains. However, the expression continues to remain high in the adult mouse brain. Weak expression was also observed in the embryonic eye, but not in other organs (**Fig. 6b**).
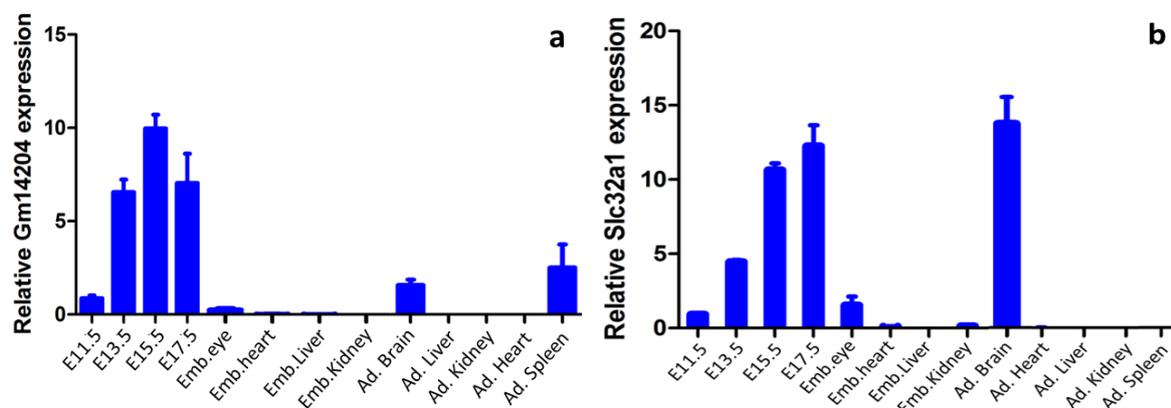

Figure 6: Quantitative RT-PCR analysis of (a) *Gm14204* lncRNA and (b) *Slc32a1* mRNA in different embryonic (E14.5) and postnatal mouse (P56) organs. Values are represented as Mean ± S.E.M; n = 3 / organ. All values are normalized to *Gapdh* housekeeper mRNA.

In order to study in detail such similar expression pattern between the *Gm14204* lncRNA and *Slc32a1* mRNA, we generated Digoxigenin-labeled RNA probe (700-900 nts.) targeting the longer isoform for *Gm14204* and preformed non-radioactive RNA *in situ* hybridization on sagittal sections of E13.5, E14.5, E15.5 embryos, P1 head, P7, and P56 mouse brains. Overall, *Gm14204* lncRNA showed a strong expression throughout the central nervous system in mouse embryos at E14.5 (**Fig. 7**). Weaker expression was observed in the mouse peripheral nervous system, sensory organs such as the eye neural retina and lens fibers, ear labyrinths (inner ear), the tip of the lower lip and the tongue. Additionally, very weak expression was also detected in other tissues such as the limb mesenchyme and the kidneys (**Fig. 8a-h**). All *Gm14204* isoforms showed identical expression in the mouse embryos and adult mouse brain (data not shown).

We next checked if *Gm14204* shares expression patterns with *Slc32a1* mRNA in mouse E14.5 embryos. Both the genes have strong overlapping expression profiles in the mouse central nervous system, weak in the peripheral nervous system, sensory organs and kidneys (**Fig. 9**). *Gm14204-Slc32a1* RNA transcripts were present in the developing cortex, medial and lateral ganglionic eminences, pre-thalamus, hypothalamus, pre-tectum, tegmentum, superior colliculus, cerebellum, pons, medulla and spinal cord (**Fig. 9a-h**). Weak expression was detected in the neural retina of the eye, trigeminal and dorsal root ganglia (**Fig. 9i, j**).

These results are consistent with previous studies which show divergent lncRNAs have a similar expression with their adjacent coding genes in mouse tissues and gave us an indication that *Gm14204* lncRNA might regulate the transcription or expression of *Slc32a1* gene at an RNA level.

Mouse embryo E14.5
Distance from Midline: 0.600 mm, right side
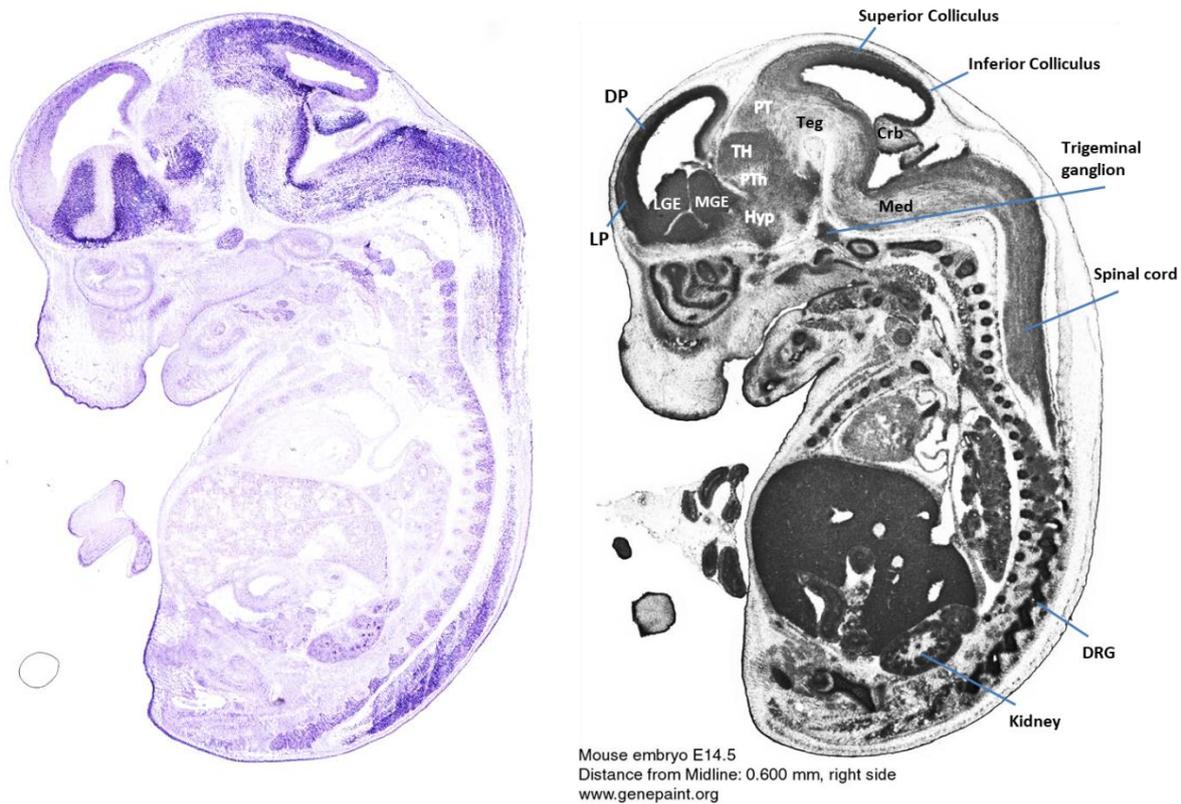www.genepaint.org

Figure 7: Overview of *Gm14204* long noncoding RNA expression at embryonic day (E) 14.5 of mouse development (left) as determined by RNA in situ hybridization.

Expression was observed in the dorsal pallium (DP) cortex, lateral and medial ganglionic eminence, thalamus, hypothalamus, superior colliculus, tegmentum, cerebellum, medulla, spinal cord, dorsal root ganglia, trigeminal ganglion, and kidneys. The annotation of expression pattern for this lncRNA is shown in the right figure (source: GenePaint database). Crb: cerebellum; DP: dorsal pallium; DRG: dorsal root ganglia; Hyp: hypothalamus; LP: lateral pallium; LGE: lateral ganglion eminence; Med: medulla; MGE: medial ganglionic eminence; PTh: pre-thalamus; PT: pre-tectum; Teg: tegmentum; Th: thalamus.
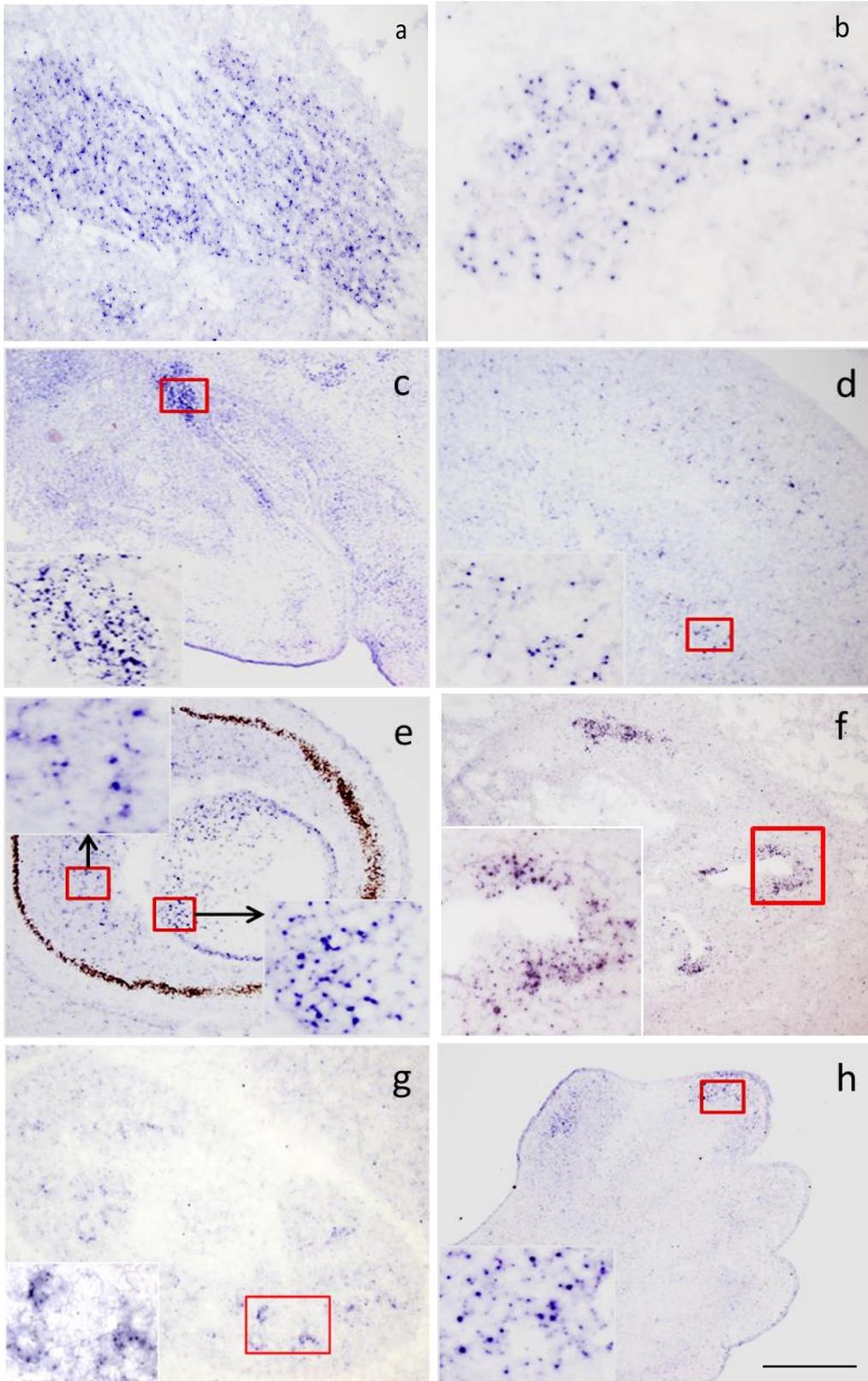
Figure 8: Expression patterns of *Gm14204* lncRNA outside the mouse central nervous system.

Apart from the E14.5 embryonic central nervous system, scattered and medium level *Gm14204* lncRNA expression was detected in the trigeminal ganglion (a), dorsal root ganglia (b), tip of lower lip (c), in the tongue (d), neural retina and lens fibers (e), ear labyrinths (f), very weak in the kidneys (g) and the limb mesenchyme (h). Scale bar: 500 µm.
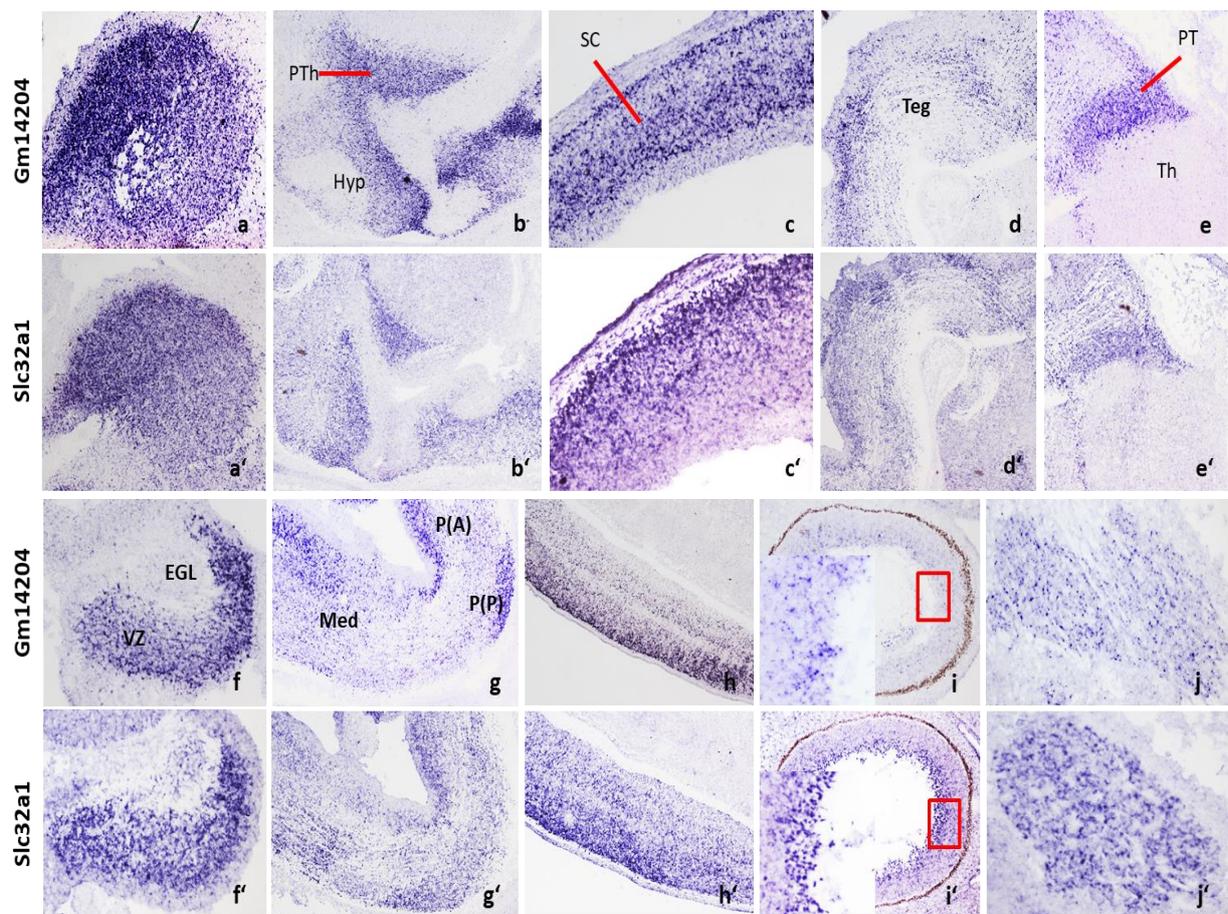

Figure 9: Expression patterns of *Gm14204* lncRNA and *Slc32a1* mRNA in E14.5 embryos established by RNA *in situ* hybridization.

Note the striking similarity of expression of both genes in the central and peripheral nervous systems: (a-a') Striatum, (b-b') Pre-thalamus (PTh) and Hypothalamus (Hyp), (c-c') Superior colliculus (SC), (d-d') Tegmentum, (e-e') Pre-tectum (PT), (f-f') Cerebellum (VZ: ventricular zone; EGL: extra germinal layer), (g-g') Anterior P(A) and Posterior P(P) pons and Medulla, (h-h') Spinal cord, (i-i') Neural retina of eye (red box), (j-j') Trigeminal ganglion, (i-i') Neural retina of eye, (j-j') Trigeminal ganglion.

2.1.5 Molecular characterization of *Gm14204* lncRNA

2.1.5a Regionalized expression of *Gm14204* in developing telencephalon

The mouse embryonic forebrain is separated into two major structures namely the telencephalon and the diencephalon. During the course of embryonic development, the dorsal part of the telencephalic vesicle gives rise to the neocortex, while the ventral part gives rise to medial and lateral ganglionic eminences (MGE, LGE), which further generate the basal ganglia (striatum and pallidum). The new-born neurons produced by the ventricular zone (VZ) are responsible for the formation of mouse cortex in an inside first (layer 6, 5), outside last (layer 2, 1) pattern.

In the developing mouse telencephalon at embryonic day (E) E13.5 - E15.5, the ventricular zone showed weak expression of *Gm14204* lncRNA in the VZ of pallium and sub-pallium. On the contrary, a very strong expression for the lncRNA was observed in the intermediate zone (IZ) and the cortical plate (CP) of the neocortex (**Fig. 10a, d, g**). These are the migrating neurons in the IZ, which have already started migrating from the VZ to form the upper layers (4-2) of the cortex. In the sub-pallium, *Gm14204* was strongly expressed in the mantle zone of both the lateral and medial ganglionic eminences (LGE and MGE) (**Fig. 10c, f, i**), which are populated mostly by interneurons. The striatum (**Fig. 10b, e, h**) also showed a strong expression for *Gm14204* lncRNA. It is already known the cortical interneurons travel tangentially from the lateral ganglionic eminence to the intermediate zone (IZ) of the cortex from where they migrate to form the cortical plate (S. A. Anderson, Eisenstat, Shi, & Rubenstein, 1997).

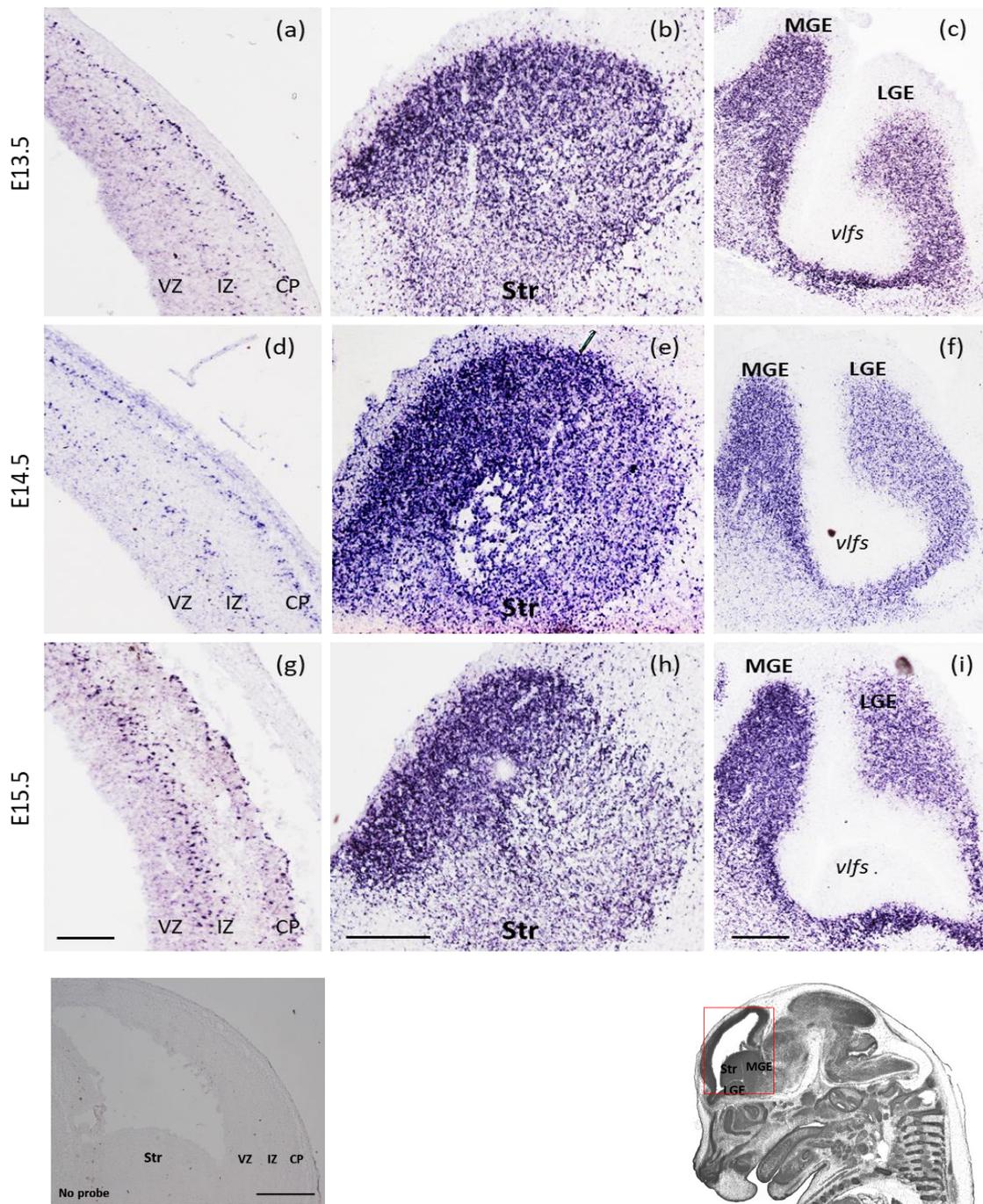**Figure 10: Expression of *Gm14204* lncRNA in sagittal section (25 µm) of a developing mouse telencephalon.**

In the embryonic cortex, the expression was observed in the cortical plate and intermediate zone at three developmental time points (a, d, g). In the basal ganglia, the lncRNA was confined to the Striatum (b, e, h), the medial ganglionic eminence and the lateral ganglionic eminence (c, f, i). CP: cortical plate; IZ: intermediate zone; LGE: lateral ganglionic eminence; MGE: medial ganglionic eminence; Str: striatum; *vlfs*: lateral ventricle, septal fork; VZ: ventricular zone. Scale bar: 500 µm.

2.1.5b Regionalized expression of *Gm14204* lncRNA in the developing diencephalon

The developing mouse diencephalon gives rise to two important brain structures, namely the thalamus and the hypothalamus. As both these structures comprise of diverse cell types and neurons, there are no well-defined molecular markers that delineate different neuronal subtypes, which makes the analysis of these structures difficult. According to the Prosomeric model, specific expression of many transcription factors and regulatory molecules define morphological boundaries of the developing diencephalon and segment like domains called prosomeres (Puelles & Rubenstein, 2003). The three major segments in the posterior diencephalon include the prosomere 1 (p1, epi-thalamus), prosomere 2 (p2, thalamus) and prosomere 3 (p3, pre-thalamus). A hallmark of this model is the boundary between p2 and p3, the zonal limitans intrathalamic (Zli) that release factors required for the patterning of developing thalamus (Bulfone et al., 1993).

In our study, we found that *Gm14204* was strongly expressed in the ventral or pre-thalamus and in the hypothalamus, but not in dorsal thalamus. In order to characterize in more detail, the expression of lncRNA, we took advantage of the extensively studied expression pattern of *Dlx5* TF gene in the diencephalon (Puelles & Rubenstein, 2003). Comparison of expression of *Gm14204* with *Dlx5,* at matched medial to lateral sections of embryonic brains revealed several striking features. The lncRNA was expressed in a prosomere restricted manner confined to p1 (epi-thalamus, medium expression) and in p3 (pre-thalamus, strong expression). We did not observe any expression in the p2 (thalamus) (**Fig. 11a-c**). The derivatives of p3, the reticular nucleus (Rt) and zona incerta (ZI) were strongly positive for *Gm14204* lncRNA (**Fig. 11c**). No expression was detected in the Eminentia thalami (Em) (**Fig. 11b**).

More detailed analysis helped to identify very strong and restricted lncRNA expression in the distinct hypothalamic areas such as Preoptic Area (POA), Anterior hypothalamus (AHA, AHP), Suprachiasmatic nucleus (SCV), Periventricular nucleus (Pv), Dorsal medial hypothalamic nucleus (DMH), Tuberal hypothalamus (TM) and Arcuate nucleus (Ar). Unlike DMH, the ventral medial hypothalamic nucleus (VMH) was negative for lncRNA expression (**Fig. 11a-b'**). The posterior hypothalamus and the mammillary hypothalamus did not express *Gm14204* (**Fig. 11a**). Similarly, the supraopto-periventricular area (SPV) was negative for lncRNA.
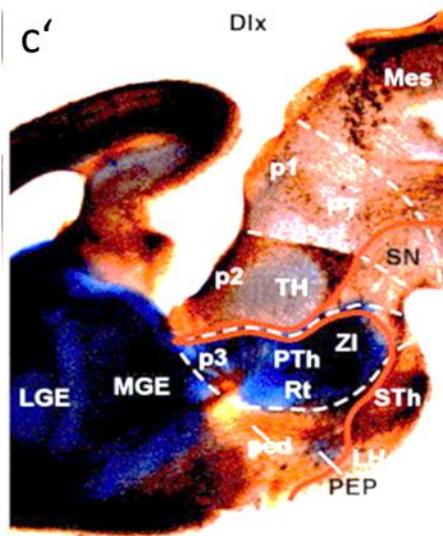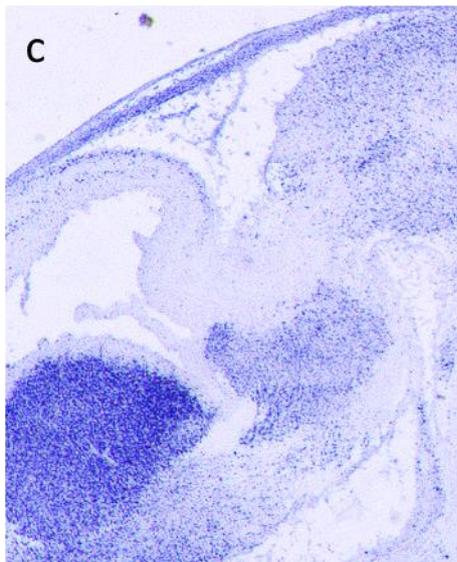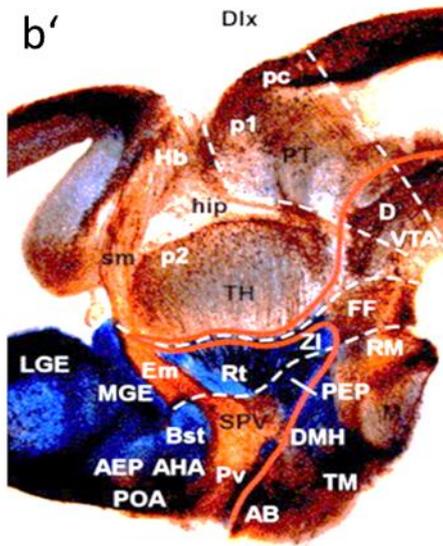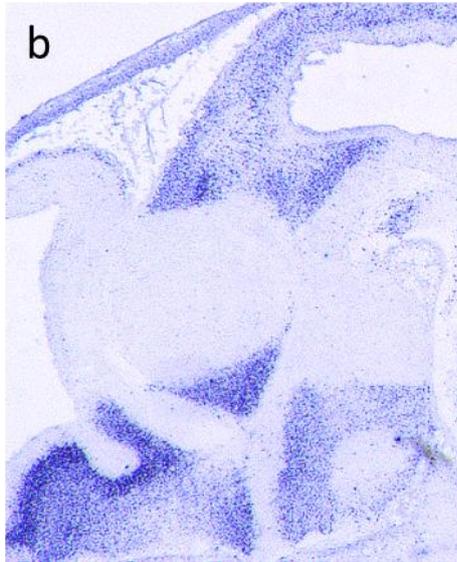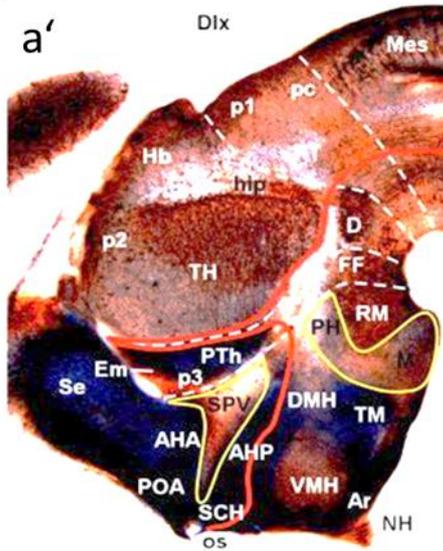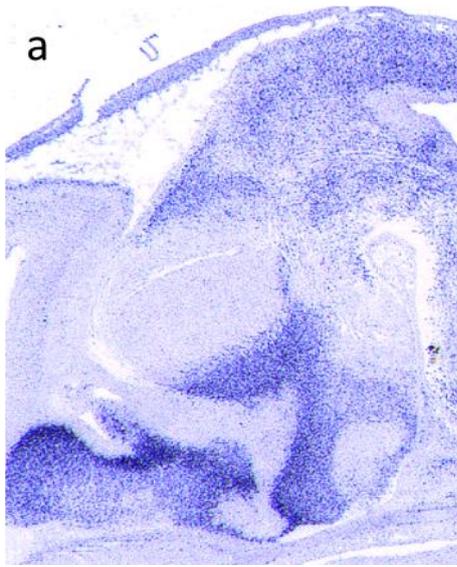
Figure 11: Expression pattern of *Gm14204* lncRNA (a-c) in the telencephalic forebrain sub-divisions as shown in three sagittal sections (25 μm thick) of E14.5 developing mouse embryonic brain.

Sections (a-c) and (a´-c´) represent in blue the expression of *Gm14204* lncRNA and *Dlx5*, respectively. *Dlx5* images taken from (Puelles & Rubenstein, 2003), on approximately matched sagittal E14.5 (a-c) and E15.5 (a´-c´) brain sections from medial (a, a´) - to- lateral (c-c´) levels. Prosomeric boundaries are shown as white broken lines. The alar-basal boundary is shown in red line. AB: anterobasal nucleus; AEP: anterior entopeduncular area; AHA: anterior hypothalamus, anterior area; AHP: anterior hypothalamus, posterior area; Ar: arcuate nucleus; Bst: bed nucleus of stria terminalis; D: nucleus of Darkschewitsch; DMH: dorsomedial hypothalamic nucleus; Em: eminentia thalami; FF: Forel fields; Hb: habenula (epithalamus); hip: habenulo interpedunclar tract; LGE: lateral ganglionic eminence; LH: lateral hypothalamus; M: mammillary complex; Mes: mesencephalon; MGE: medial ganglionic eminence; p1-p3: prosomeres; POA: preoptic area, pc: posterior commisure; ped: telencephalic peduncle; PEP: posterior entopeduncular area; PH: posterior hypothalamus; PT: pretectum; PTh: pre-thalamus; Pv: anterior periventricular nucleus; RM: retromammilary area; Rt: reticular nucleus; SCH: suprachiasmatic area; Se: septum; SPV: supraopto-paraventricular area; TH: thalamus; TM: tuberomammilary area, VMH: ventromedial hypothalamic nucleus; ZI: zona incerta.

2.1.5c Expression of *Gm14204* lncRNA in the mid- and hind-brain

We next characterized *Gm14204* expression in embryonic mouse midbrain and hindbrain. In the developing mesencephalon, strong expression was observed in the superior colliculus (SC), the tegmentum (Teg.) and the pre-tectum (PT) at E13.5 – E15.5 developmental stages (**Fig. 12**), while the inferior colliculus remained negative (data not shown). In the superior colliculus, the expression was only present in the mantle layer of differentiated cells. In the developing hindbrain (E13.5-E15.5), we observed *Gm14204* expression in the anterior and posterior pons {P(A) and P(P)} and the medulla (Med.) (**Fig. 13b, d, f**).

The major morphological units of the cerebellum are the cerebellar cortex, white matter, and the cerebellar nuclei. The cortex of a mammalian cerebellum is a layered structure, although produced in a different manner as compared to cerebellar cortex (**Fig. 13 scheme**). The inner granule cell layer (GCL) of the cerebellum is composed of granular cells and sparsely distributed interneurons. Above GCL is the Purkinje cell layer (PCL), followed by the outermost molecular cell layer (MCL). The cerebellar neurons are generated from two progenitor zones: the ventricular zone (VZ) and the upper rhombic lip (RL) (Martinez et. al, 2013). The VZ produces cells of PCL, Bergmann glia, interneurons and astrocytes, whereas RL produces projection neurons of the cerebellar nuclei and the granule cells of the GCL. In the beginning, the proliferating precursors of the GCL migrate towards the cerebellar primordium to form the extra granule layer (EGL) where they still divide continuously to expand the progenitor pool. After becoming post-mitotic neurons, the EGL cells descend radially to form the inner GCL (Martinez et. al, 2013). From our *in situ* hybridization results in E13.5-E15.5, *Gm14204* lncRNA expression remained confined to the cells of presumptive GCL (**Fig. 13a, c, e**). No lncRNA expression was detected in cerebellar germinative zones, VZ, and RL/EGL.
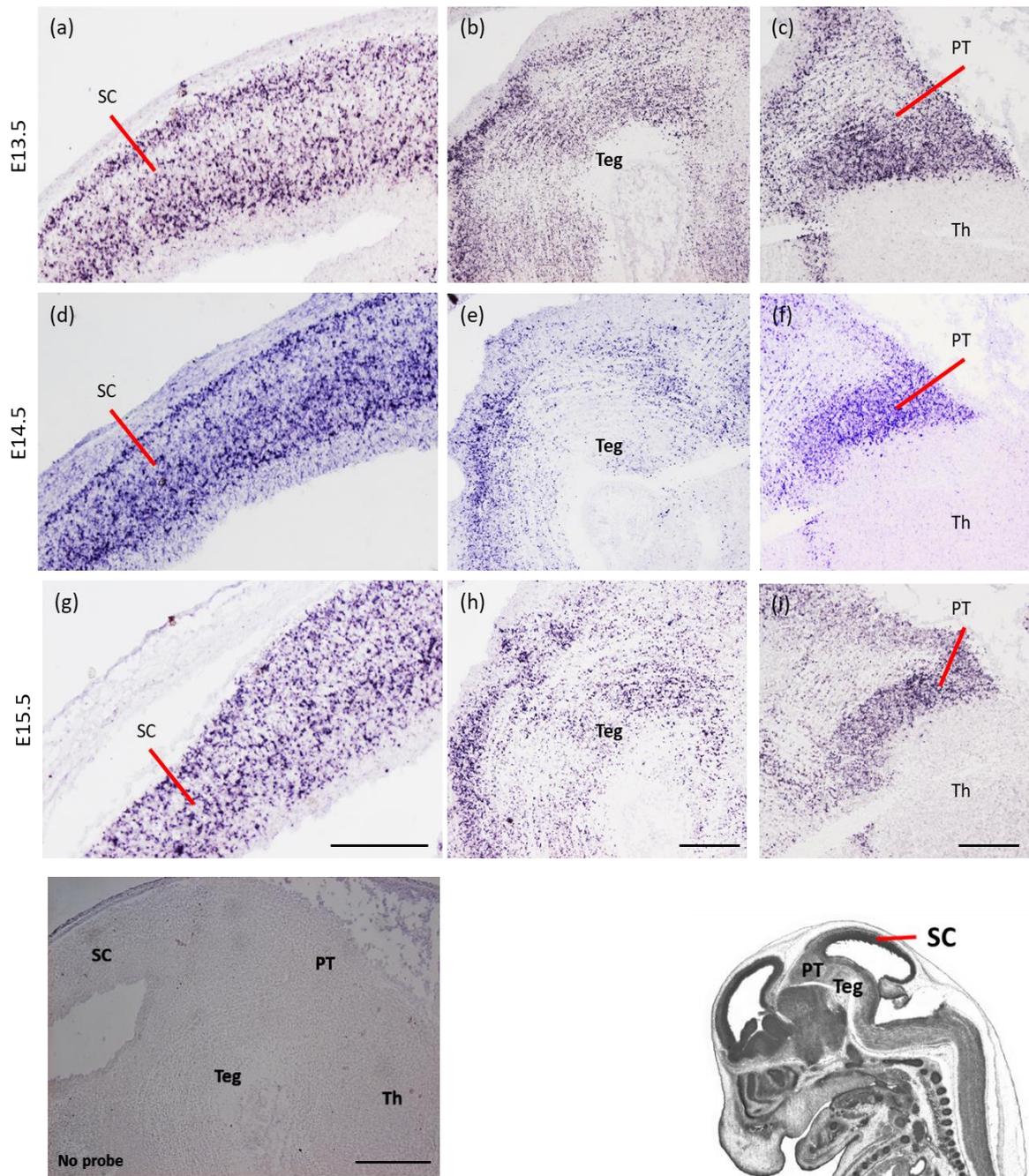
Figure 12: Expression of *Gm14204* lncRNA in sagittal sections (25 µm) of a developing mouse mesencephalon.

Strong expression was observed in the superior colliculus (a, d, g), tegmentum (b, e, h) and pre-tectum (c, f, i). PT: Pre-Tectum; SC: superior colliculus; Teg: tegmentum; Th: dorsal Thalamus. Scale bar: 500 µm (a, d, g); 200 µm (b-c, e-f, h-i).
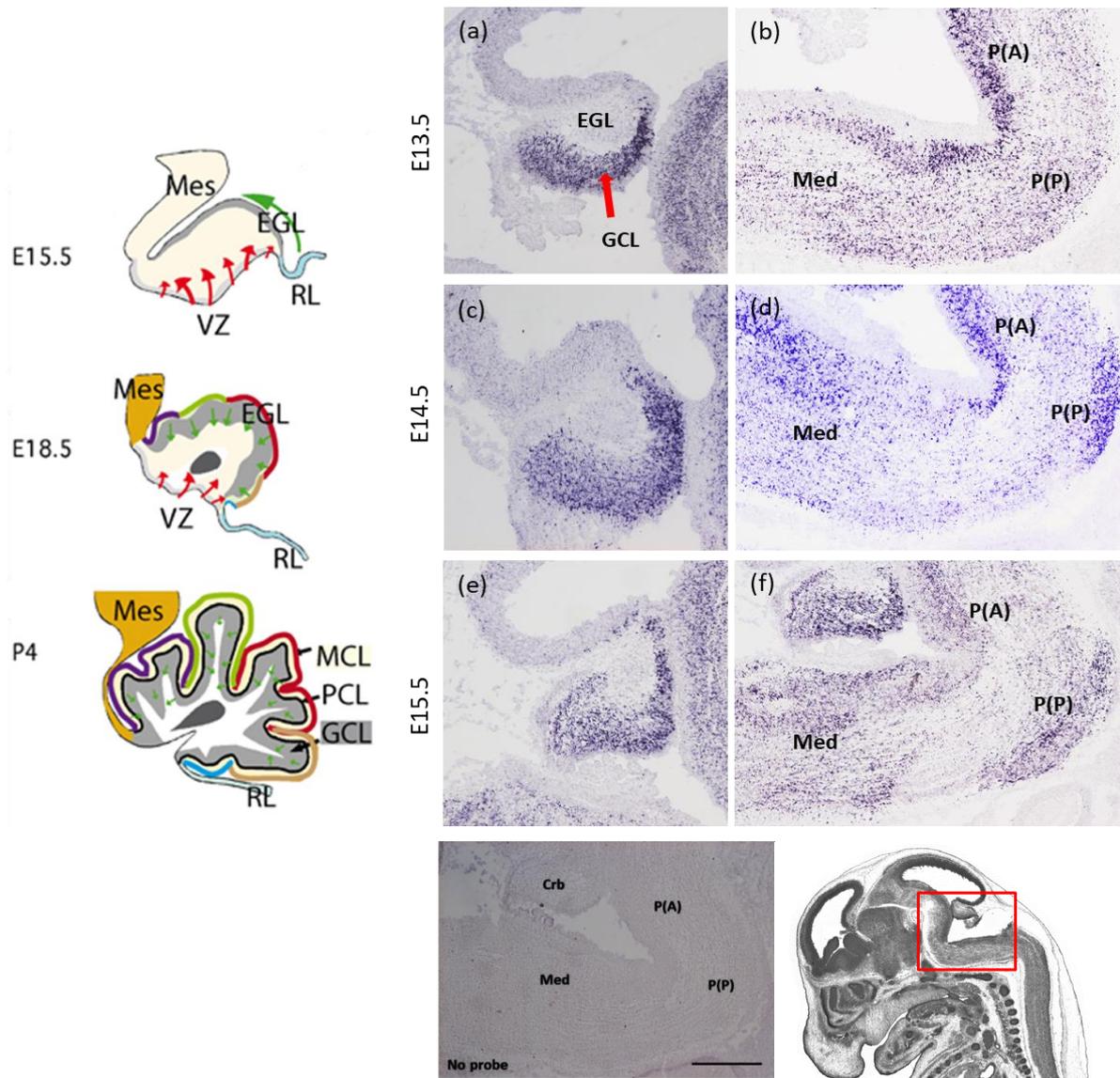
Figure 13: Expression of *Gm14204* lncRNA in sagittal sections (25 µm) of the developing rhombencephalon.

Strong expression was observed only in the presumptive GCL layer (red arrow in (a)), but not in the VZ or RL/EGL of the developing cerebellum (a, c, e). Additionally, *Gm14204* is also strongly expressed in the anterior and posterior pons and the medulla (b, d, f). Crb: cerebellum, EGL: external germinal layer, Med: medulla, P(A): anterior pons, P(P): posterior pons. Scale bar: 200µm (a, c, e); 500 µm (b, d, f). Scheme on the left is adapted from (Martinez et al., 2013); for details see text.

2.1.6 Expression of *Gm14204* lncRNA in the developing mouse eye

In mouse, eye development is first visible with the formation of bilateral indentations in the prospective forebrain called optic sulci or optic pits at E8.0. At this time point until E9.0, optic vesicles are formed and they reach in close contact with the surface ectoderm, which will later form the lens placode. During development, optic vesicles and the lens placode invaginate co-ordinally to form the bilayered optic cup. Lens vesicle is formed when the lens placode invaginates to form the lens cup, which will separate from the surface ectoderm. This detached surface ectoderm will later proliferate to give rise to the corneal epithelium. Finally, the innermost layer of the optic cup will give rise to the neural retina where retinal neurogenesis will occur during development. Such a complex process of eye development requires extensive coordination of transcription factors, many of which are layer specific{reviewed in (Heavner & Pevny, 2012)}.

From our RNA-ISH data, interestingly we have found that the expression of *Gm14204* lncRNA increases in the retina during development. Very weak and scattered expression was first observed at E13.5 in the retina and the lens fibers (**Fig. 14a**). The expression was strong at E14.5 in both these tissues (**Fig. 14b**) but almost disappeared in the lens fiber at E15.5. At E15.5, retina had still the same level of expression (**Fig. 14c**). In the new-born pups at Postnatal day P1, very strong and regionalized expression was observed in the retina. Comparatively, the expression was stronger than visible in the embryonic stages. In P1 eye, the optic cup margin and the central optic cup layer also showed scattered but weak *Gm14204* expression (**Fig. 14d**).
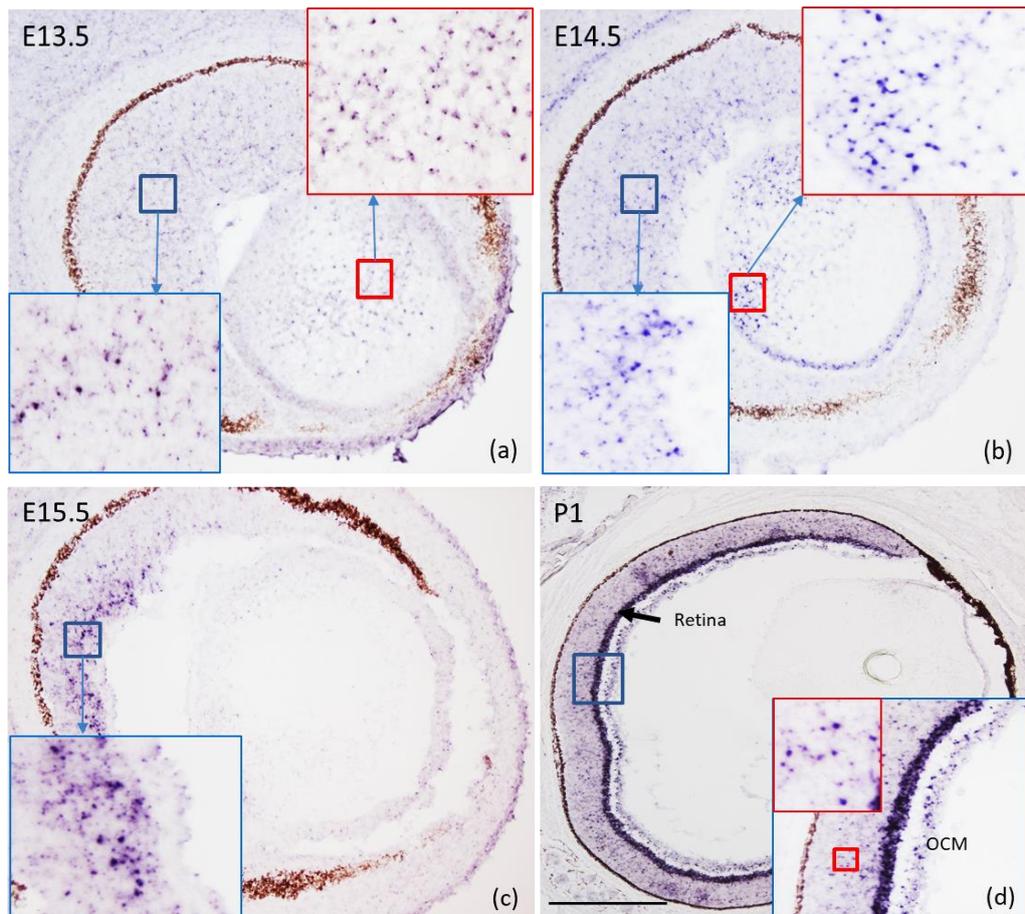
Figure 14: Expression of *Gm14204* lncRNA in sagittal sections (25 μm) of the developing mouse eye.

Weak expression was observed in (a) E13.5 embryonic retina (blue inlet) and lens fibers (red inlet). The expression seems to gradually increase only in the eye retina, throughout development as seen in E14.5 (b) and E15.5 (c) mouse embryos. In E15.5, no expression was detectable in the lens fibers. (d) Robust expression was observed in the mouse postnatal day 1 (P1) eye retina, while the Optic Cup Margin (OCM) and the central optic cup layer exhibited low expression (red inlet in (d)). No expression was observed in the lens fibers at P1. Magnification: 5X; inlets: 20X. Scale bar: 500μm.

2.1.7 Expression of *Gm14204* lncRNA in the developing hippocampus

The development of murine hippocampus in mouse starts as early as E12.5, where the hippocampal neuroepithelium (HNE) and the dentate neuroepithelium (DNE) begins to be visible (Urban & Guillemot, 2014). These structures become clearly distinguishable at E14.5-E15.5 giving rise to presumptive CA (Cornu Ammonis) fields (CA1, CA2, and CA3). At E17.5 the hippocampal fissure starts to form where the dentate precursor cells migrate to and accumulate there. HNE gives rise to hippocampal neurons which then migrate to the location CA1-CA3 (Urban & Guillemot, 2014). There has been evidence that mature CA1 and CA3 fields are distinctively visible only after birth (P0), and the pyramidal neurons from these two fields mix together to give rise to small CA2 region (Tole, Christian, & Grove, 1997).

The expression of *Gm14204* lncRNA was first detected at E15.5 stage in the CA1 layer, but not before (**Fig. 15a, b, c**). Here the expression was restricted to the *stratum radiatum* layer and not in the *stratum oriens* (**Fig. 15c**). In postnatal day P1 mice, *Gm14204* expression was seen scattered throughout the hippocampal field with few positive neurons in the CA and DG regions (**Fig. 15d**). By P7, the expression had become restricted to pyramidal cell layer of CA1 and CA3 with no expression in the granule cell layer of the DG. A few neurons in the polymorph layer of DG showed positive *Gm14204* expression (**Fig. 15e**). In the adult mouse brain (P56), very strong expression was observed in the CA1, CA3 and the polymorph regions of the mature hippocampus (**Fig. 15f**). It is surprising that the expression is not specialized to the CA layers in P1 mouse brain, but as brain development progresses, the lncRNA expression becomes more confined to CA1 and CA3 layers at P7 developmental stage. This confinement of expression continues in the adult mouse brain (P56). The dentate gyrus, however, never expresses *Gm14204* lncRNA in developing or adult hippocampus.
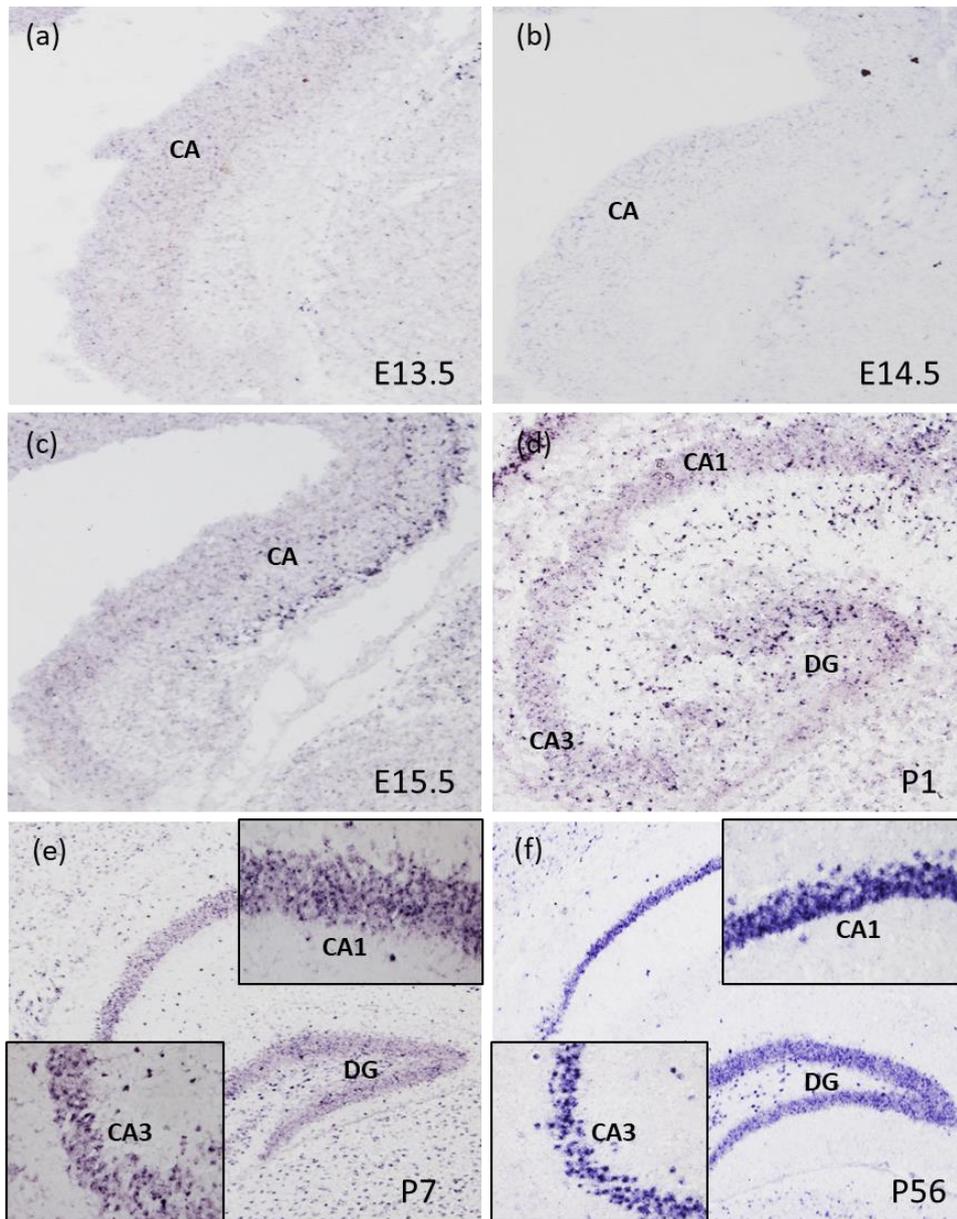
Figure 15: Expression of *Gm14204* lncRNA in sagittal sections (25 µm) of the developing mouse hippocampus.

No expression was observed in the CA region of E13.5 (a) or E14.5 (b) mouse embryos. Expression starts to be visible in the presumptive CA1 region of E15.5 embryos (c). Scattered pyramidal neurons in the P1 hippocampus showed strong expression for the lncRNA (d). In P7 and P56 mouse brain, CA1 and CA3 showed a very strong expression for *Gm14204* lncRNA (e, f). CA: Cornu Ammonis, DG: dentate gyrus. Magnification: (a, b, c: 10X), (d, e, f: 5X), inlets: 20X.

2.1.8 Expression profile of *Gm14204* lncRNA in young P1, P7, and adult P56 mouse brain

We next performed RNA-ISH on young mouse P1 head, P7 brain and adult P56 brains to determine whether *Gm14204* was expressed also postnatally. We also wanted to know if the lncRNA overlaps with the *Slc32a1* expression in the adult mouse brain. Just immediately after birth, in P1 mouse brain, a scattered but strong expression was observed in most of the brain tissues. Similarly, P1 and P7 mouse cortex showed a diffused expression for the lncRNA (**Fig. 16a, 17a**). *Gm14204* was present in all six layers of the cortex, not showing layer-specific expression as seen in the embryo pallium. No expression was detected in the sub-ventricular zone (SV) or the white matter (WM). In the adult brain (P56), the expression had decreased in all layers (**Fig. 18a**). In the olfactory bulb, the dynamic expression was detected very strongly in the glomerular layer, mitral layer and the granule layer at both P1 and P7 stages (**Fig. 16b, 17b**), while the expression had reduced in P56 brain in all layers (**Fig. 18b**). No expression was detected in the outer plexiform layer (opl) at any of these stages. *Gm14204* had a high expression in the embryonic striatum as seen in Fig. 10 and the same level of expression was consistent in the Caudoputamen (striatum) of P1 brain (**Fig. 16c**). On the contrary, the expression decreased in the striatum of P7 brain (**Fig. 17c**) and almost very weak expression was detected in P56 brains (data not shown). In the cerebellum, *Gm14204* was expressed in the Purkinje cell layer (PCL) and the molecular cell layer (MCL) at high levels in P1 and P7 (**Fig. 16e, 17e**). In P56 mouse brain, *Gm14204* was detected very strongly in the Golgi cells (**Fig. 18b**, red arrows), Purkinje cells (PC) and weakly in the molecular layer (**Fig. 18b**). As observed at the embryonic time-points, *Gm14204* lncRNA has a pre-thalamus specific expression that was also present in P1, P7 brains in the alar plate of pre-thalamus (**Fig. 16f, 17f**). No expression was present in the dorsal thalamus, consistent with the absence of expression at the embryonic developmental stage. The adult neural stem cells which are born

in the sub-ventricular zone of a mouse brain are directed towards the olfactory bulb *via* the rostral migratory stream (RMS). Interestingly, we had detected strong expression for the lncRNA in the RMS of P1 and P7 brains (**Fig. 16d, 17d**). This expression was markedly decreased in the adult mouse brain (P56) (**Fig. 18e**) which is in accordance with a decrease in the number of NSCs in the adult brain. *Gm14204* lncRNA was observed in a similar expression pattern with the *Slc32a1* mRNA in most of the adult brain tissue, although the lncRNA was expressed much weakly than *Slc32a1* mRNA.
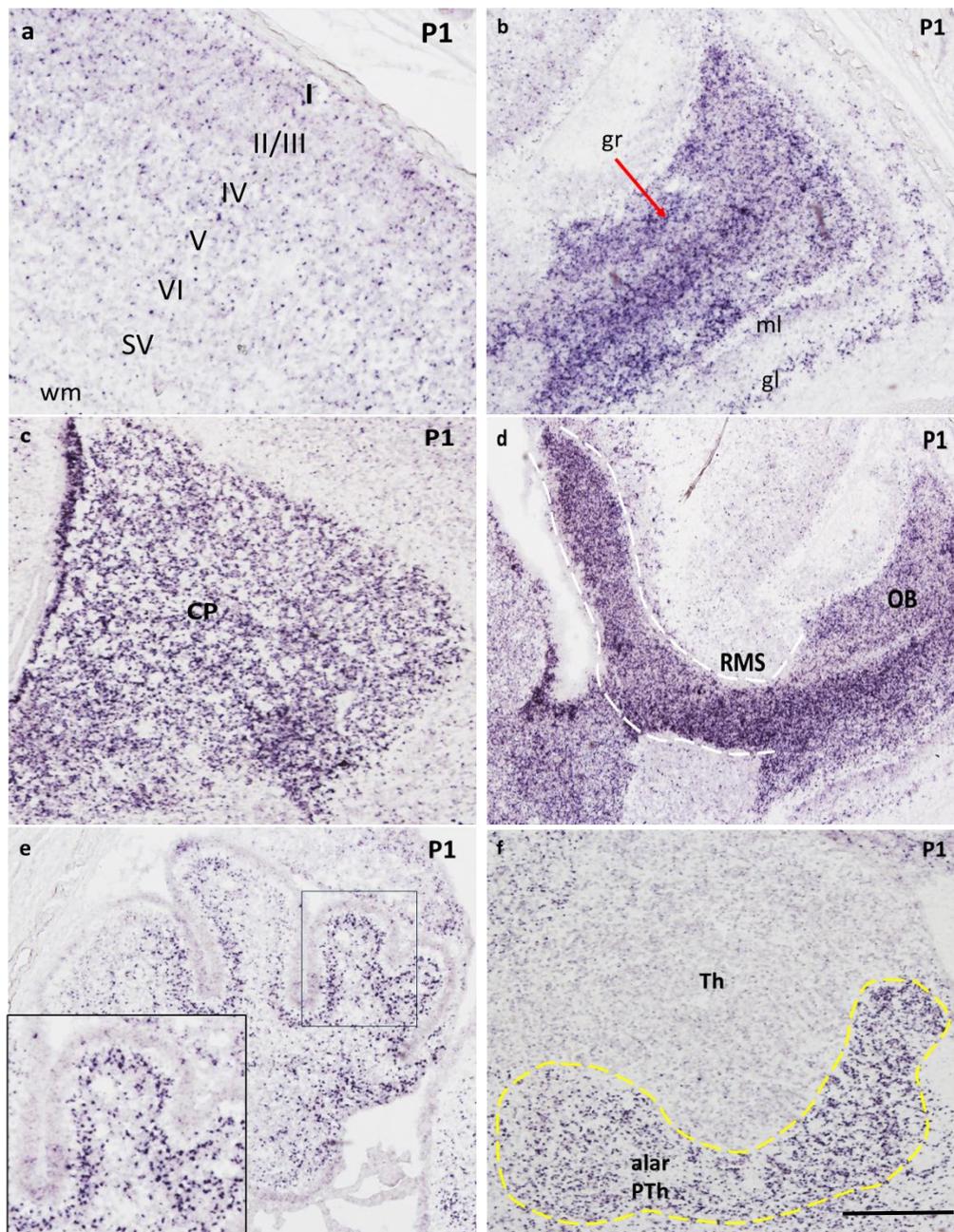
Figure 16: Expression of *Gm14204* lncRNA in P1 mouse brain.

(a) In cortex, lncRNA showed strong positive scattered cells in all I-VI layers, but not in the sub-ventricular zone (SV) or the white matter; (b) Expression in olfactory bulb: the lncRNA is strongly detected in the glomerular layer (gl), mitral layer (ml) and the granule layer (gr); (c) Strong expression of lncRNA in the caudoputamen (CP); (d) The rostral migratory stream (RMS, white dashed lines) show a robust expression of lncRNA, OB: olfactory bulb; (e) In cerebellum, many GCL cells are expressing lncRNA; (f) The thalamus lacks *Gm14204*, but in the territory of pre-thalamus (PTh), a strong expression is observed in the anlage of pre-thalamic derivatives, the reticular nucleus (Rt) and Zona Incerta (ZI). Scale bar: 500 µm.
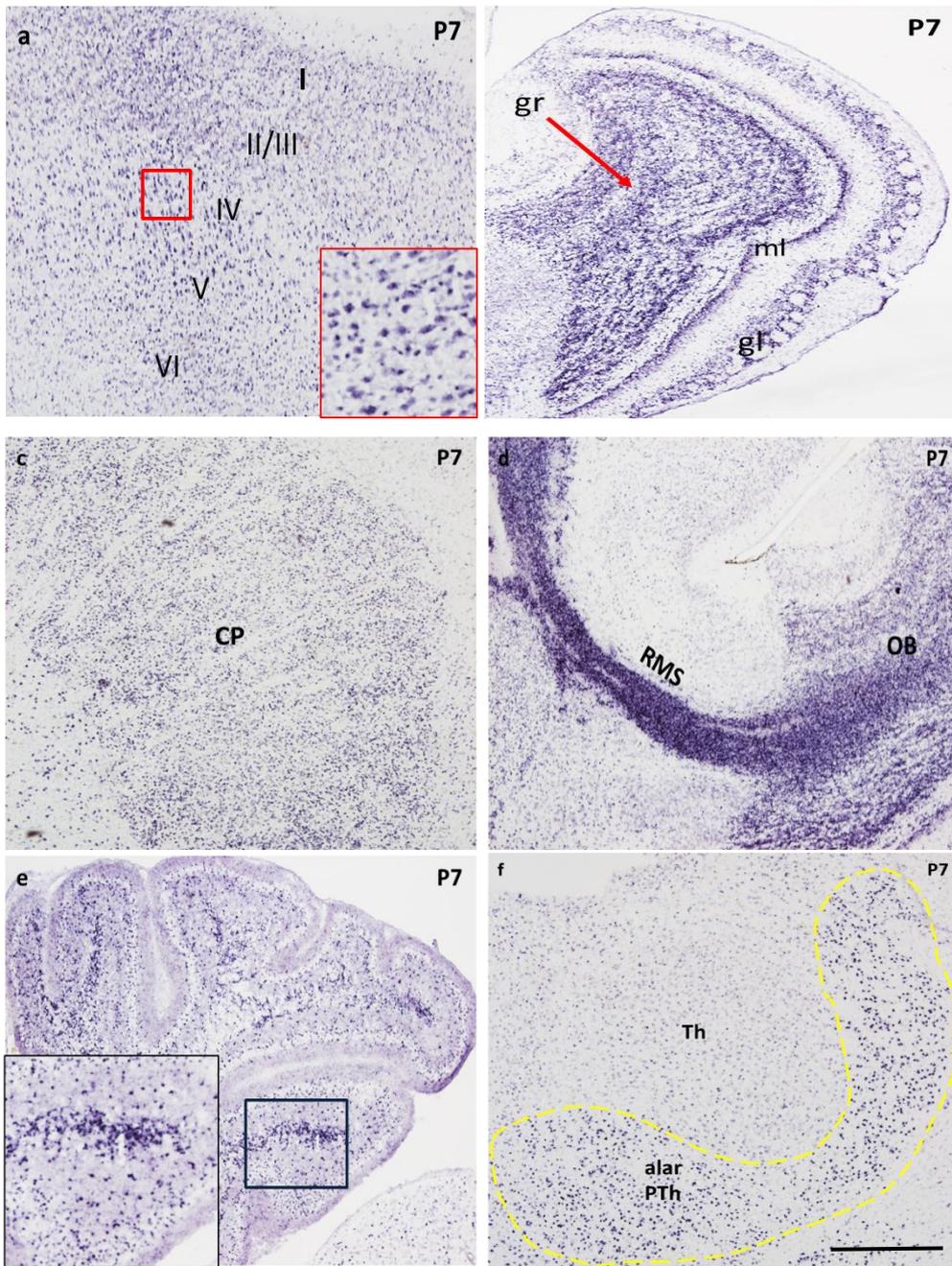
Figure 17: Expression of *Gm14204* lncRNA in P7 mouse brain.

The expression pattern of lncRNA was similar to that observed in P1 brains (see Fig. 14 for abbreviations).

Cortex   Cerebellum   Olf. Bulb
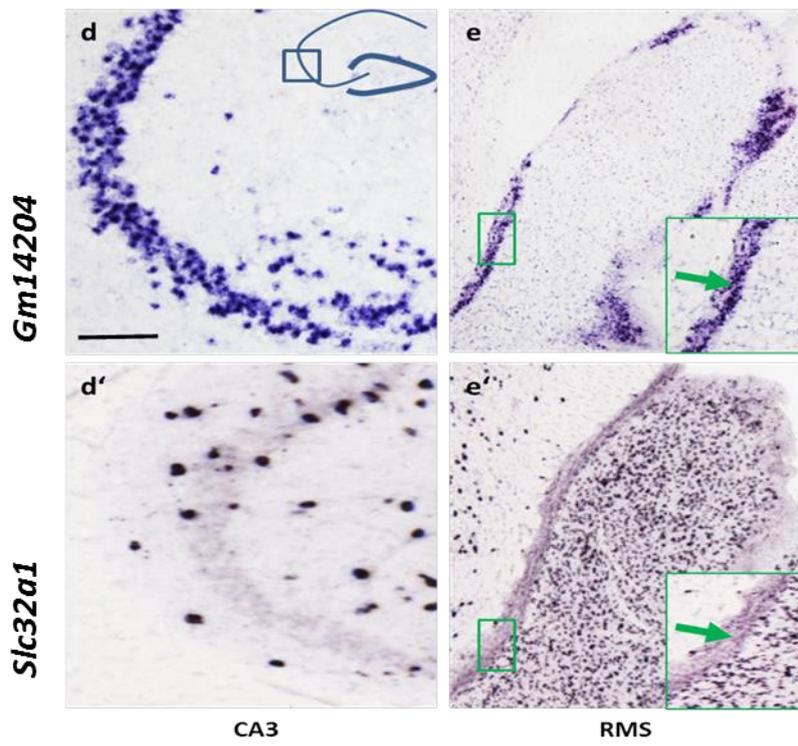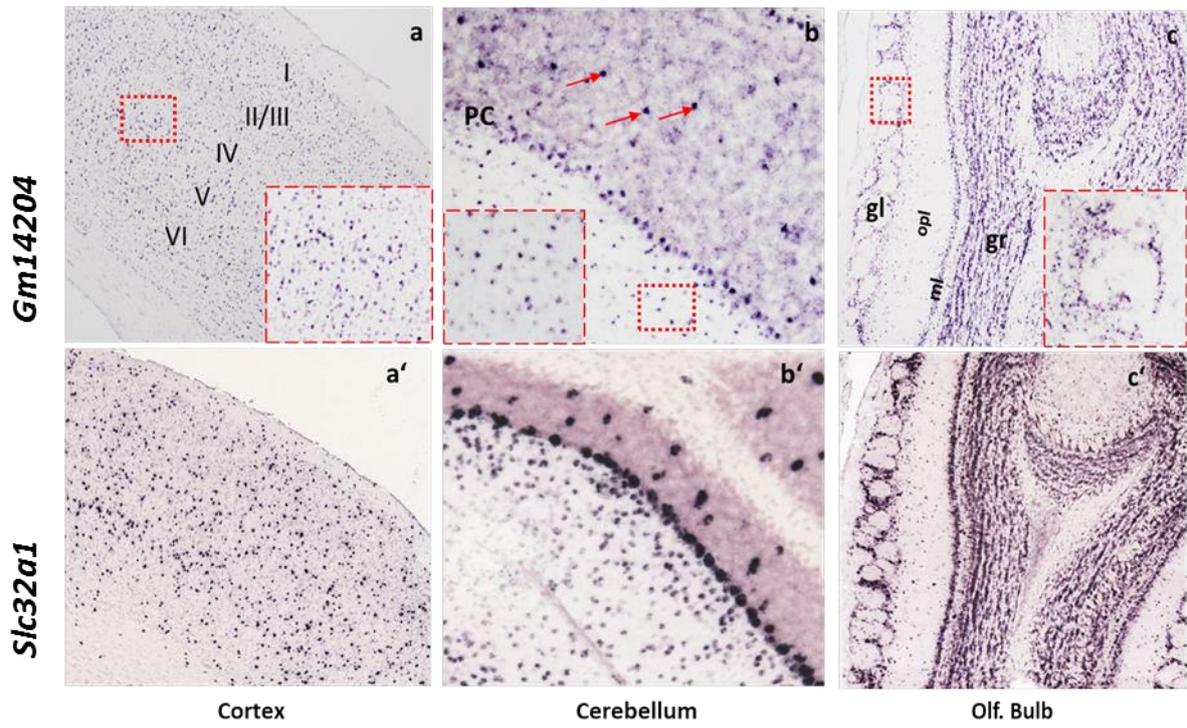
CA3   RMS

Figure 18: Expression of *Gm14204* lncRNA (a-e) and *Slc32a1* mRNA (a'-e') in adult P56 mouse brain.

(a) In the cortex, *Gm14204* lncRNA expression was scattered in all I-VI layers, but had reduced than in P1/P7 mouse brains; (b) In the cerebellum, a scattered expression was detected in the molecular cell layer (red box), while regional expression was observed in the Purkinje cell layer (PC). Few Golgi cells also marked *Gm14204* expression (red arrows); (c) In the olfactory bulb, the lncRNA was detected strongly in the glomerular layer (gl; red box), mitral layer (ml) and the granule layer (gr), while no expression was present in the outer plexiform layer (opl). (a'-c') Striking similar expression patterns of *Slc32a1* in respective tissues with that of *Gm14204*. Note that *Slc32a1* mRNA has high expression in the adult brain than lncRNA. (d-d') *Gm14204*, but not *Slc32a1* was expressed in the CA3 layer of hippocampus; (e-e') strong expression is seen in the rostral migratory stream (RMS) for *Gm14204*, but not *Slc32a1*. Scale bar: (a-e): 200 µm. (a'-e'): *Slc32a1* images are taken from the Allen Brain Atlas database.

2.1.9 Co-expression of *Gm14204* lncRNA with *Slc32a1* mRNA

As per our hypothesis, if *Gm14204* lncRNA regulates the expression of *Slc32a1* mRNA, then they should be co-expressed together in the same subset of cells in mouse tissues. To address this question, we aimed to perform double RNA *in situ* hybridization on embryonic E14.5 mouse tissue sections. Digoxigenin-labeled *Gm14204* RNA probe and FITC labeled *Slc32a1* RNA probe were prepared for RNA fluorescent ISH. *Gm14204* lncRNA was detected with anti-DIG Cy3 (red) antibody and *Slc32a1* mRNA with anti-FITC (green) antibody. Interestingly, RNA FISH on these sections showed co-expression of both RNA transcripts in the entire embryonic nervous system. The co-expression, for example, was shown for the ganglionic eminences and the cerebellum (**Fig. 19**). Since the tissue density in embryonic tissues is much compact, it is extremely difficult to visualize individual cells or neurons and determine the exact localization of RNA transcripts. Adult mouse brain sections, on the other hand, are less compact and one can clearly distinguish individual neurons. Due to this advantage, we performed RNA FISH on P56 mouse brain sections and observed co-expression of both RNA transcripts in a subset of neurons in the cortex, olfactory bulb, ventral thalamus, substantia nigra and the cerebellum (**Fig. 20**). However, we did not detect any co-labeling in a few tissues such as the hippocampus (**Fig. 21**) or the dorsal thalamus (data not shown). We observed that *Gm14204* is highly localized to the nucleus, whereas *Slc32a1* mRNA in the cytoplasm (**Fig. 20**). A few *Slc32a1* mRNA transcripts were also observed inside the nuclei. These transcripts might be nascent, pre-spliced RNAs which are not yet in their mature form.

Next, we asked a question whether *Gm14204* lncRNA and *Slc32a1* mRNA are co-transcriptionally active together and if the lncRNA plays a role in the transcription of *Slc32a1* gene. Conventional RNA FISH technique uses dual amplification strategy where an RNA probe is hybridized to detect mRNA/lncRNA transcript *in vivo* with an anti-hapten antibody

conjugated to HRP (horseradish peroxidase). Further, many biotinylated tyramine molecules are covalently attached to the vicinity of the RNA probe, which is detected with an enzymatic activity using various substrates. Due to this enzymatic reaction of biotin molecules, the signal generated at the end of the reaction looks diffused and spread around the RNA transcript (**Fig. 20a**; individual RNA transcripts not detectable, instead are seen as a cluster). It, therefore, becomes impossible to localize the exact sub-cellular localization of the RNA molecule.

To cope with the above situation, we employed a recently commercially developed technique called the single molecule FISH (smFISH; Advanced Cell Diagnostics Inc., Bio-Techne) (F. Wang et al., 2012). RNAscope Technology allows the sub-cellular detection of target RNA molecules within cells with very less background (see Methods section for details on the procedure).
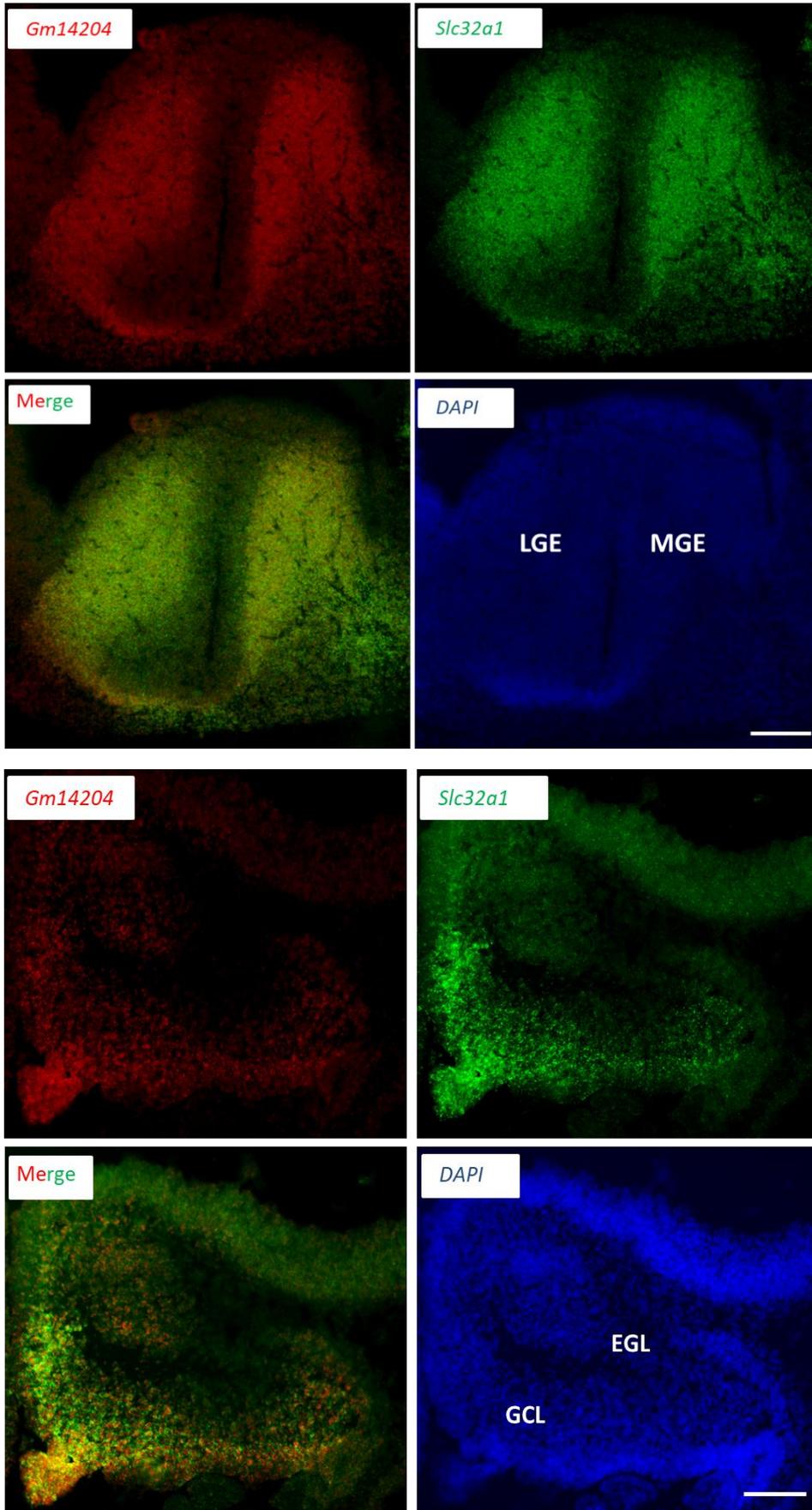
Figure 19: Double RNA FISH of *Gm14204* lncRNA and *Slc32a1* mRNA on 25 µm embryonic E14.5 sections.

Gm14204 (red), *Slc32a1* (green) and DAPI (blue) co-expression in (top panel): lateral and medial ganglionic eminences; (bottom panel): cerebellum. In the cerebellum, both transcripts are expressed only granule cell layer (GCL) but not in the extra granular layer (EGL). Scale bar: 200 µm.
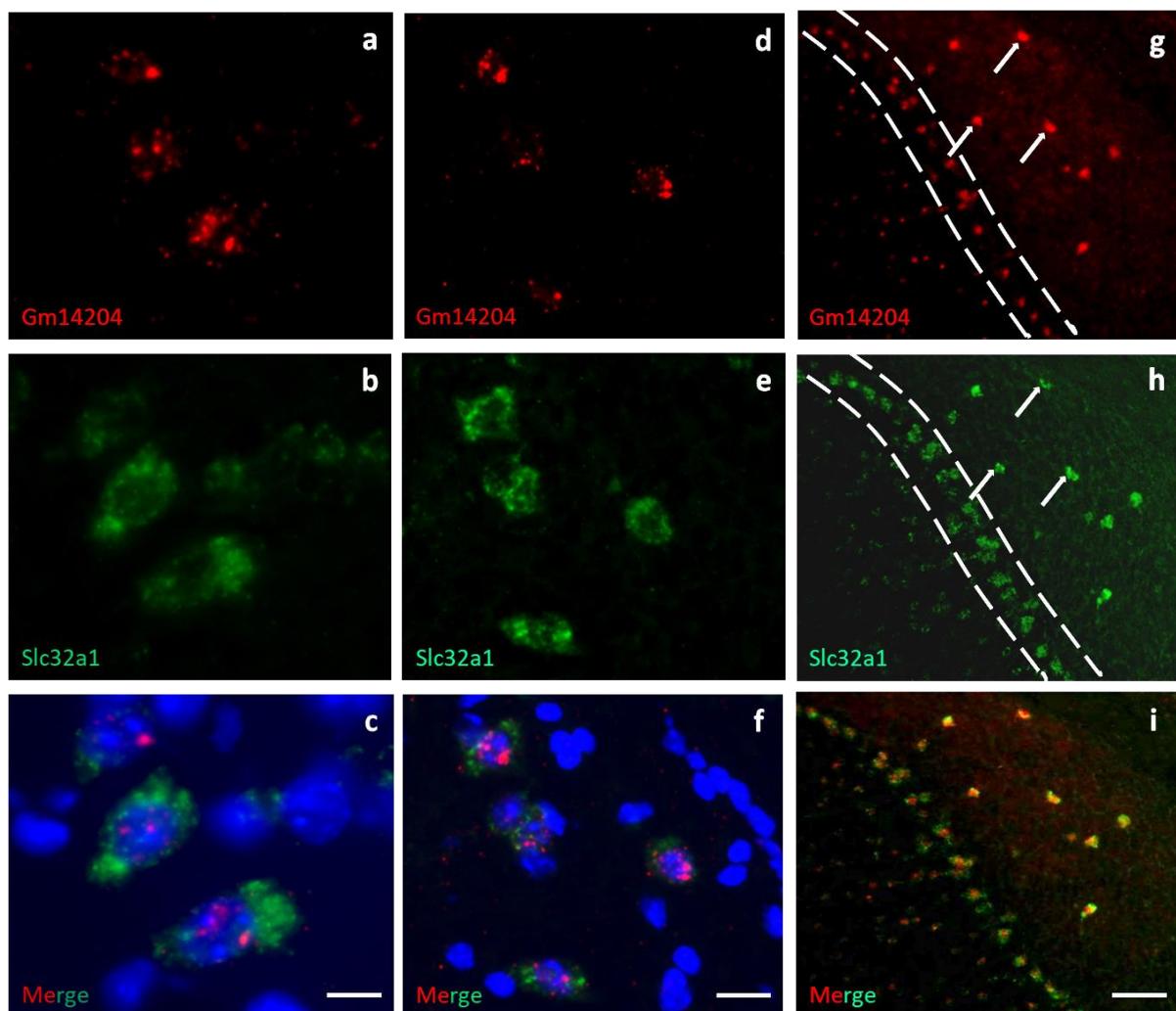
Figure 20: Double RNA FISH of *Gm14204* lncRNA and *Slc32a1* mRNA on 20 µm adult mouse brain sections.

Gm14204 (red) and *Slc32a1* (green) co-expression in (a, b, c): cortical neurons in layer III/IV; (d, e, f): substantia nigra neurons; (g, h, i): cerebellum. In the cerebellum, white arrows show co-labeling in the Golgi cells and dashed lines represent Purkinje layer. A few neurons in the molecular layer also co-label both transcripts. Blue staining in a-f is *DAPI*. *Gm14204* was highly enriched in the nucleus, whereas *Slc32a1* was detected only in the cytoplasm (a-f). Scale bar: (a-f: 5 µm); (g-i: 50 µm).
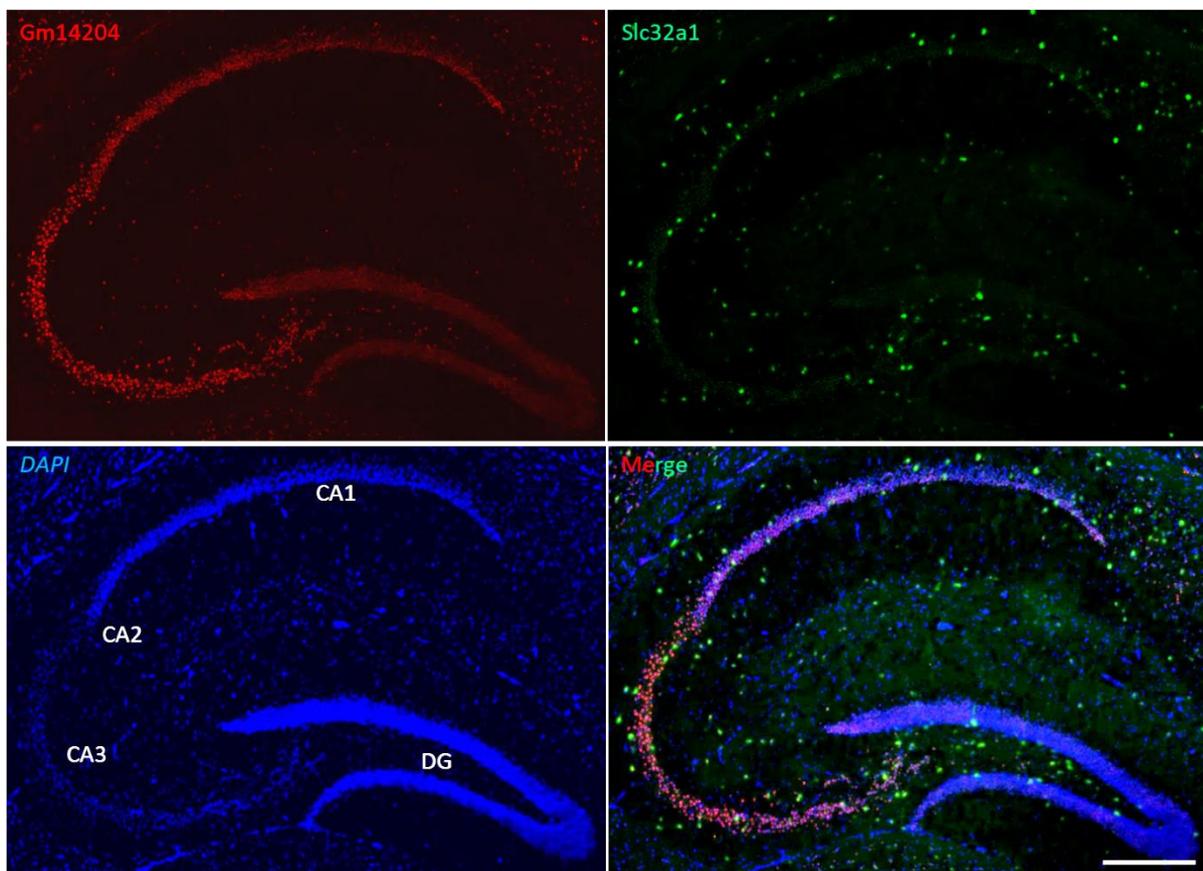

Figure 21: Double co-labeling of *Gm14204* lncRNA and *Slc32a1* mRNA in the adult mouse hippocampus.

The expression of lncRNA is highly localized to the CA regions of the hippocampus. CA3 has a strong expression than CA1 and CA2. The granular region of Dentate Gyrus was negative for the lncRNA, but the polymorph layer had a few positive co-expressed cells. Scale bar: 200 µm.

Our hypothesis was that the mature form of lncRNA transcript would regulate the transcription of *Slc32a1* gene. In order to detect the nascent pre-mRNA of *Slc32a1*, we generated intronic probes, which would bind the introns of the pre-spliced *Slc32a1* RNA transcripts. For *Gm14204*, we designed probes that would detect the mature form transcript. The probes were commercially ordered from ACDBio (Bio-Techne). We performed smFISH using these probes on adult mouse brain sections and performed confocal imaging to visualize expression profile. We chose substantia nigra and cerebellum for imaging as these tissues have distinct and separate neurons for better imaging. As expected, all the *Slc32a1* signals were present inside the nuclei as two bright spots (green, **Fig. 22**) and possibly these regions mark the transcription sites of this gene. Interestingly, *Slc32a1* signals overlapped with the *Gm14204* signals (red, **Fig. 22**) at exactly the same sites within the nuclei. Since the distance between the two genes is only 50 bp on the genome, we assumed that *Gm14204* gene is transcribed at this region, spliced, processed by some yet unknown mechanism and loaded onto the transcriptional machinery of *Slc32a1* gene. However, we cannot neglect the possibility that the co-localization of signals can arise due to the co-transcription of both genes by the same transcriptional machinery. Furthermore, we were curious to know whether *Gm14204* lncRNA resided only at its transcriptional site (from **Fig. 22**) or transported to different sites in the nucleus. Confocal imaging results clearly showed that *Gm14204* did not reside only at particular regions, but had different location sites within the nuclei (**Fig. 23**). It suggests that this lncRNA could have mechanisms not only in *cis* but also in *trans,* whereby it could regulate different genes. A single lncRNA showing varied expression phenomena within the nuclei have not been reported previously.
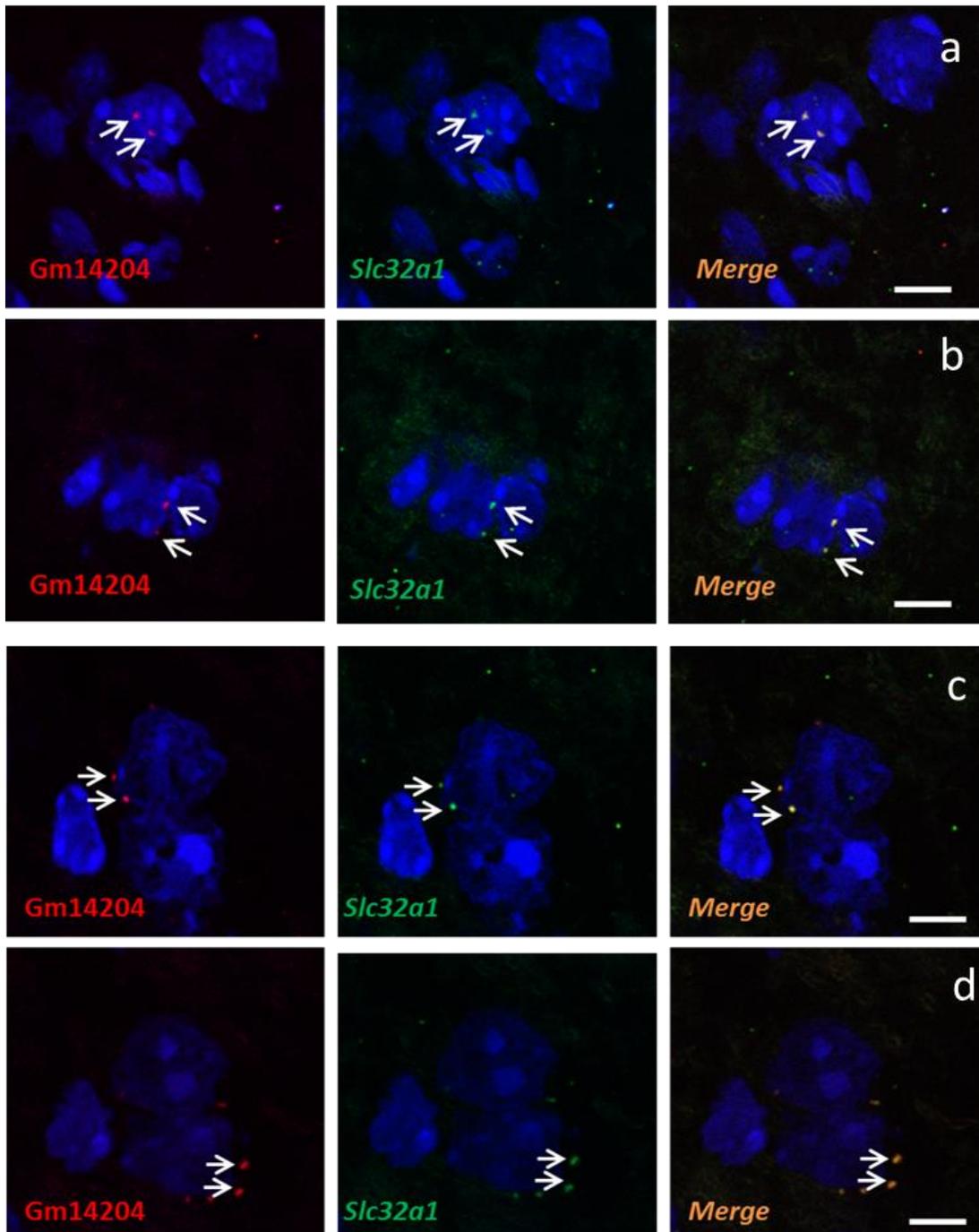
Figure 22: Single-molecule FISH data of *Gm14204* exonic lncRNA and *Slc32a1* intronic RNA.

Confocal imaging shows the mature *Gm14204* lncRNA transcript signal overlaps with the pre-mRNA of *Slc32a1* transcript in (a, b) Substantia Nigra and (c, d) Cerebellum of adult mouse brain. The overlapping of signals can also result since both the genes are situated very close to one another and transcribed co-transcriptionally. Blue: *DAPI*. Scale bar: 5 µm.
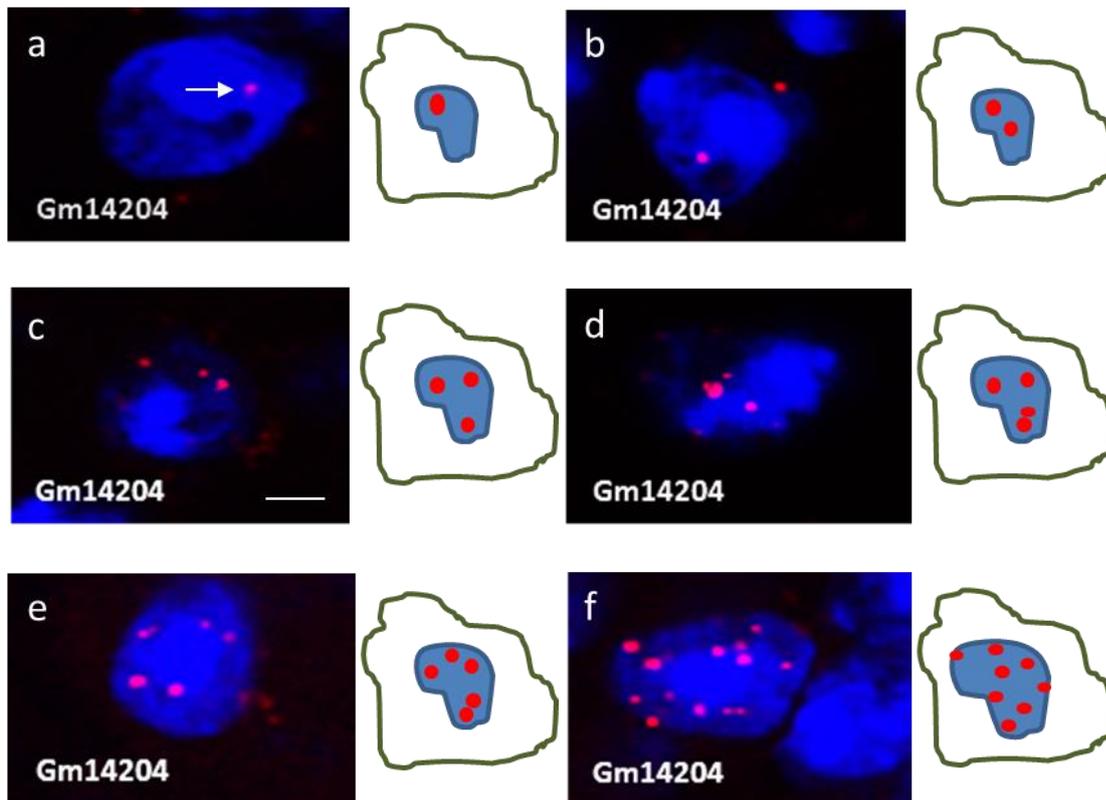
Figure 23: *Gm14204* **exhibits a variety of sub-nuclear localization patterns in the hypothalamus of adult mouse brain as determined by single molecule FISH.**

Confocal imaging shows that the lncRNA is present as a single spot (a), twin foci (b), triple foci (c), four spots (d), five spots (e) and as more than 5 spots (f) inside the cell nucleus (Blue: *DAPI*). Observe that the intensity of different lncRNA spots is not the same. Some spots were more enriched with lncRNA signals than others, meaning lncRNA is not present at equal concentrations at different places in the nuclei. Representative demonstrations of localization are shown as cartoons. Scale bar: 2 µm.

2.2 Part 2: Mouse perturbation of *Gm14204* long non-coding RNA

2.2.1 Strategy for generation of a new genetic mouse model for *Gm14204* lncRNA

To begin the understanding of *Gm14204* lncRNA function *in vivo*, we generated a knockout mouse model for this lncRNA gene. Since *Gm14204* is highly expressed during embryonic developmental stages and given that it is located next to an important mammalian neuroinhibitory transporter gene *Slc32a1,* it is possible that a deletion of this lncRNA would have lethal effects during mouse development. *Slc32a1* gene deletion in mice has been shown to be lethal (Wojcik et al., 2006) and a conventional knockout strategy to delete *Gm14204* locus would be risky. Moreover, since both these genes are only 50 bp apart, although, on opposite strands, we decided not to delete the promoter of *Gm14204* as this strategy might delete important regulatory binding elements required for the transcription of *Slc32a1* gene. The resulting phenotype might be linked to *Slc32a1* disruption and not because of *Gm14204* knockout. To overcome this situation, we aimed to insert a strong polyA termination signal followed by four artificial transcription termination sequences (TTS) (Schwalb et al., 2016) immediately at the beginning of the first intron of the lncRNA. We did not insert these sequences in the first exon or immediately after transcription start site because both these regions accommodate many LINE/SINE repeat elements. On the other hand, the first intron does not have such elements, which we thought might be safer for the insertion of termination cassette. Yet, there is the possibility of a short *Gm14204* transcript (~50 bp) being produced from the first exon of the lncRNA gene. Since the transcript will lack a 5'-G cap, it might be unstable and therefore susceptible to degradation by RNA exonucleases.

The design of the targeting construct was partially performed by me, with the help of Polygene Transgenetics, Switzerland. The generation of developing of mouse knockout was commercially done at Polygene Transgenetics. To generate this lncRNA-KO allele, the

endogenous *Gm14204* lncRNA gene locus was targeted *via* homologous recombination in mouse B6N ES cells. A vector comprising of *FRTNeo/KanFRT* selection cassette followed by a polyA signal sequence (pA) and four times TTS sequences were transfected in ES cells. The pA sequence and TTS were flanked by *loxP* sites (**Fig. 24**). This design will leave the neighboring gene and other loci elements intact with minimal changes. A targeting efficiency of 4.2 % (16 out of 384 clones) was achieved, as confirmed by Southern blotting (see Methods figure). The positive ES cells were next injected into mouse embryos to generate germ line transmitting chimeras. The heterozygous mice *Gm14204$^{pATTS(neo)/+}$* were born viable and fertile. These were then bred to *Cre-* deleter strain for the removal of *Neo/Kan* cassette. The resulting mice born were *Gm14204$^{pATTS/+}$*. These were bred to homozygosity to generate *Gm14204$^{pATTS/pATTS}$* animals. These homozygous animals were fertile, viable and no external abnormalities were visible. The mice were bred in our animal facility further, to establish a *Gm14204* KO colony.
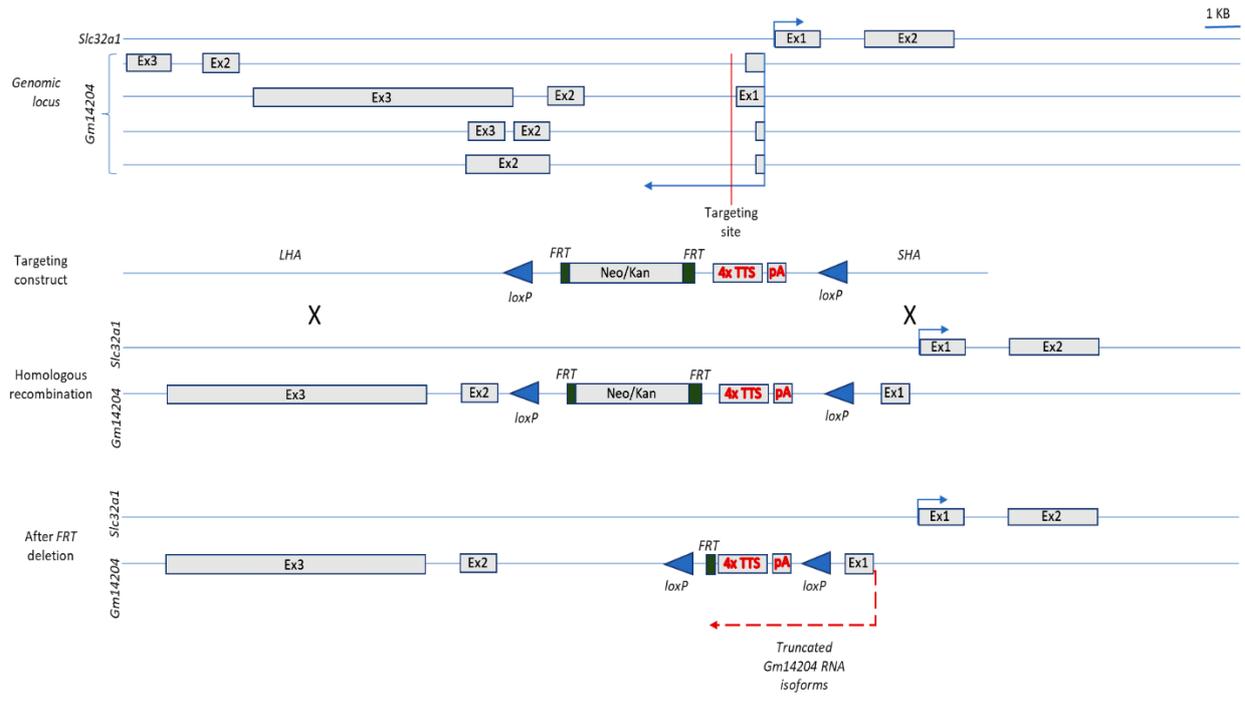
Figure 24: Schematic representation of the targeting construct.

The targeting strategy was designed in such a way that all *Gm14204* lncRNA isoforms are targeted for the knockout. The selection cassette (*Neo/Kan,* flanked with FRT sites) along with the polyA (pA) and four times TTS sequence (Transcription Termination Sequence), flanked with *loxP* sites was introduced just at the beginning of the first intron of *Gm14204* gene (red horizontal line). After homologous recombination and *Cre-* based FRT deletion, the selection cassette was removed leaving behind pA and 4xTTS sequences with one FRT site. This strategy resulted in truncation of all *Gm14204* isoforms without disrupting the *Slc32a1* gene locus. LHA: long homology arm, SHA: short homology arm.

2.2.2 Successful targeting of all *Gm14204* lncRNA isoforms

To confirm depletion of *Gm14204* lncRNA, total RNA from E12.5-E13.0 *Gm14204*$^{+/+}$ and *Gm14204*$^{pATTS/pATTS}$ mouse brains (n = 4 each) was isolated to perform quantitative RT-PCR analysis. qRT-PCR results confirmed the absence of lncRNA in *Gm14204*$^{pATTS/pATTS}$ brains as compared to *Gm14204*$^{+/+}$, confirming a complete deletion of all *Gm14204* transcripts from mouse (**Fig. 25**). Thus, these results showed that all the lncRNA isoforms were targeted using our strategy.

Next, RNA *in situ* hybridization using Digoxigenin-labeled RNA probes was performed on E14.5 mouse *Gm14204*$^{+/+}$ and *Gm14204*$^{pATTS/pATTS}$ embryonic sections. We designed the RNA probe targeting the longer isoform of *Gm14204* lncRNA. The results confirmed full knockout of lncRNA in *Gm14204*$^{pATTS/pATTS}$ tissues (**Fig. 26**), consistent with qRT-PCR results. Additionally, single-molecule FISH on adult mouse brain sections (**Fig. 27**) also verified depletion of lncRNA in *Gm14204*$^{pATTS/pATTS}$ animals.
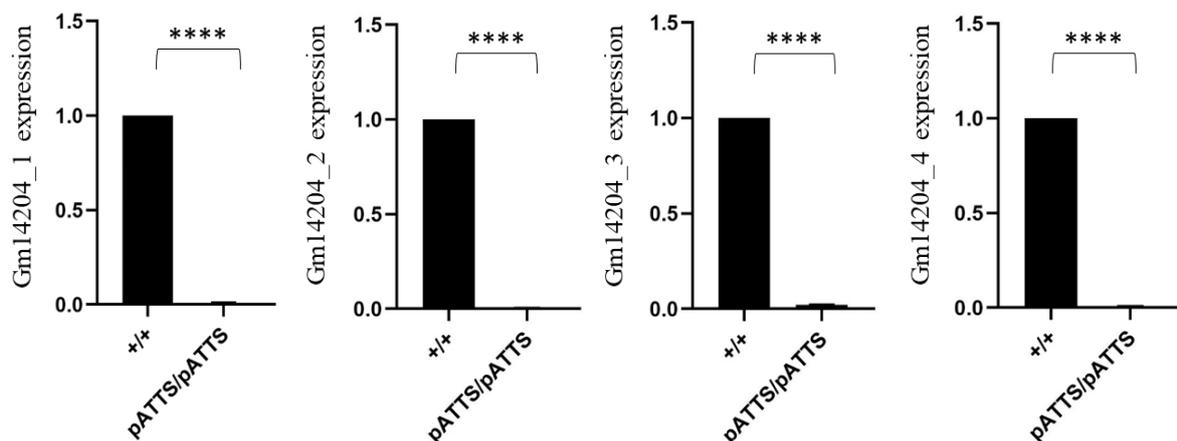

Figure 25: Quantitative RT-PCR on E12.5-E13.0 mouse brains confirms complete absence of all *Gm14204* lncRNA transcripts in mutants. All values are normalized with the housekeeper gene *Rpl13a*. Error bars indicate S.E.M. ****p < 0.0001 (n=4), Unpaired Student's t-test with Welch's correction.

Figure 26: RNA-ISH on E14.5 mouse embryonic sagittal sections (20 μm) confirm complete loss of *Gm14204* lncRNA in *Gm14204^pATTS/pATTS* mice. RNA-ISH was performed to target the predominant isoform of lncRNA. (a-a') Neocortex, expression is visible in the cortical plate and the intermediate zone only in WT; (b-b') Medial and Lateral ganglionic eminence; (c-c') superior colliculus; (d-d') PTh: pre-thalamus; Hyp: hypothalamus; (e-e'): cerebellum (GCL: granule cell layer); (f-f'): Trigeminal ganglion; (g-g'): kidney; (h-h'): developing eye (NR: neural retina, LF: lens fiber); (i-i'): Spinal cord. Scale bar: 200 μm.
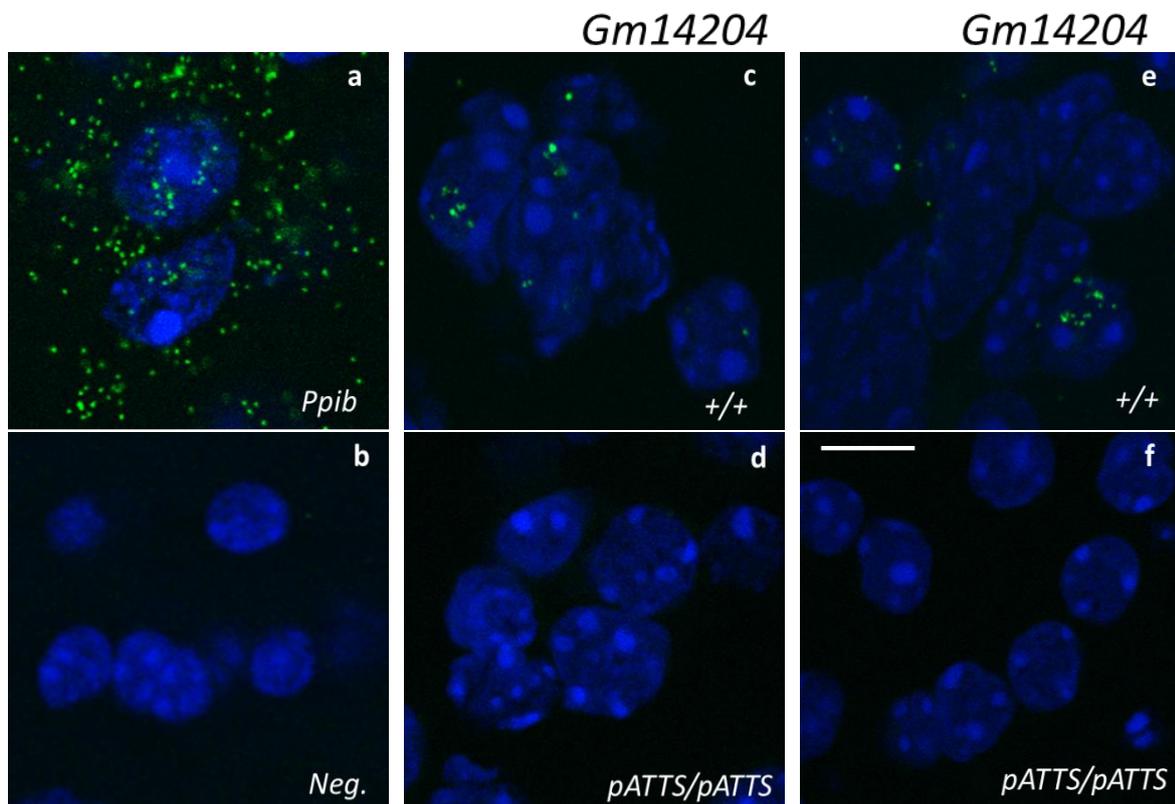

Figure 27: Single molecule FISH on 2 weeks old mouse brain sections confirms a complete loss of Gm14204 lncRNA in *Gm14204^pATTS/pATTS* mice. (a) *Ppib* mRNA: positive housekeeper control; (b) Negative control (*dapB* probe provided by ACDBio); (c-d) Cortical neurons layer IV; (e-f) Purkinje cell layer of the cerebellum. Scale bar: 5 µm.

As per our hypothesis, we had assumed that divergent *Gm14204* lncRNA might regulate the transcription of *Slc32a1* gene. The expression level of *Slc32a1* mRNA is identical to that of lncRNA in mouse embryos until E15.5 developmental stage (**Fig. 4**). Therefore, qRT-PCR was performed using RNA from E12.5-E13.0 mouse brains to observe changes in *Slc32a1* mRNA expression. We observed no significant changes in the expression levels of *Slc32a1* mRNA in *Gm14204$^{pATTS/pATTS}$* animals as compared to *Gm14204$^{+/+}$* (n=4 each) (**Fig. 28**). These results indicate that *Gm14204* lncRNA might not exert an effect on its divergent *Slc32a1* gene, but it can have possible mechanisms in *cis* on other genes on the same chromosome. Many lncRNAs are known to have mechanisms in *cis* whereby they regulate the other protein-coding genes on the same chromosome (Grote et al., 2013; Hung & Chang, 2010; Kornienko et al., 2013; Postepska-Igielska et al., 2015; Rinn & Chang, 2012; Sleutels et al., 2002), by recruiting several polycomb proteins, transcription factors or by making RNA-DNA triple helix loops. *Gm14204* lncRNA gene locus has additional four neighboring protein-coding genes in its vicinity-*Actr5*, *Dhx35*, *Ralgapb*, and *Fam83d*. We asked whether or not this lncRNA regulates the expression of these protein-coding genes.

Actr5 gene (Actin-related protein 5) is proposed to be a core component of a chromatin remodeling complex INO80, which is involved in transcriptional regulation, DNA replication and DNA repair (Jiang et al., 2010; Kitayama et al., 2009). *Dhx35* gene (DEAH (Asp-Glu-Ala-His) box polypeptide 35) is an RNA helicase protein. The function of this gene is unknown in mouse so far. *Fam83d* (Family with Sequence Similarity Member 83 D) is a microtubule-associated protein that has been shown to regulate cell proliferation, cell growth and migration in human cells (Huang et al., 2017). *Ralgapb* (Ral GTPase Activating protein ß subunit), is involved in mitotic control of cell division and spatial and temporal control of Ral GTPases (Personnic et al., 2014). All these four genes are highly expressed during embryonic mouse brain development (Source: RefSeq, ENCODE consortium), although their function in

mice has not been reported so far. Therefore, we checked the expression of these genes in *Gm14204^{pATTS/pATTS}* E12.5-E13.0 brains by qRT-PCR. No significant changes were observed in the expression of three genes (*Dhx35, Ralgapb,* and *Fam83d*) in KO animals, but we observed a significant decrease in *Actr5* mRNA in *Gm14204^{pATTS/pATTS}* brains (**Fig. 28**). This decrease in gene expression could be direct or indirect. Our results also suggest that *Gm14204* lncRNA does not have a major role in *cis*, but in *trans*.
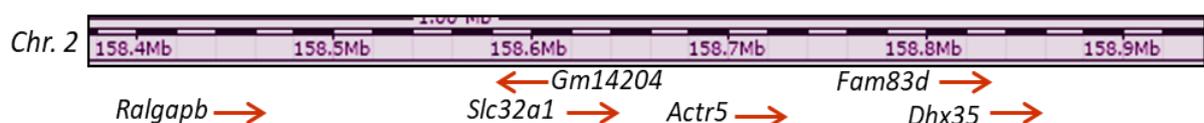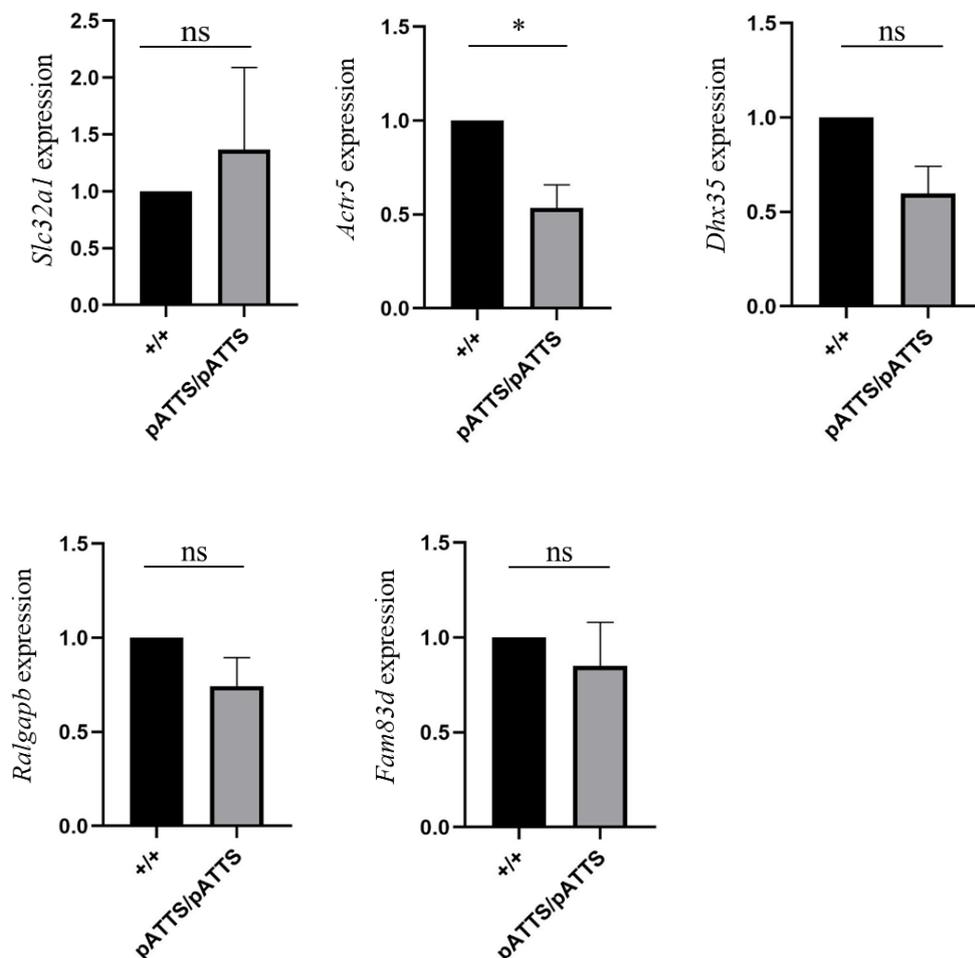
Figure 28: Quantitative RT-PCR on E12.5-E13.0 mouse brains to investigate in *cis* effects of *Gm14204* lncRNA. No change was observed in the expression of *Slc32a1*, *Dhx35*, *Ralgapb* and *Fam83d* mRNAs (ns: not significant). *Actr5* mRNA expression shows a downregulation in KO brains. All values are normalized with the housekeeper gene *Rpl13a*. Error bars indicate S.E.M. *p < 0.05 (n=4 each), Unpaired Student's t-test with Welch's correction. The scheme shows the location of protein-coding genes and *Gm14204* lncRNA.

2.2.3 Regulation of genes by *Gm14204* lncRNA involved in the development of GABAergic interneurons

The inhibitory neurons of the nervous system called the GABAergic interneurons play an important role in neuronal activity and circuits formations. Named after the release of neurotransmitter GABA by these interneurons, they are present in several regions of the brain. The origin of interneurons occurs predominantly in the ganglionic eminences (GE)-the lateral, medial and the caudal GEs and their primary roles are to synaptically release the neurotransmitter GABA to facilitate the regulation of target neurons (Kelsom & Lu, 2013). Since *Gm14204* and *Slc32a1* genes are co-expressed in the GABAergic interneurons in the developing mouse embryos (**Fig. 19-20**), we asked whether the protein-coding genes that regulate the GABAergic interneuron development in the mouse are affected by *Gm14204* knockout. In the due course of neurogenesis, GABAergic interneurons are regulated spatiotemporally by a myriad of transcription factors and their targets. *Distal-less* homeobox (*Dlx*) transcription factors *Dlx1* and *Dlx2,* play an important role in GABAergic maturation, *Nkx2.1* (NK2 homeobox) transcription factors decide interneuron fates, whereas *Lhx6* activates genes required for interneuron migration (Fukumoto et al., 2018). Glutamic Acid Decarboxylase isoforms *Gad1* and *Gad2* have been shown to regulate GABA synthesis in the interneurons and are direct targets of *Dlx1/2* genes (Le et al., 2017). *Dlx1/2* genes activate

transcription of *Gad1/2* isoforms, which in turn regulates GABA synthesis and its package into the synaptic vesicles.

In a recent publication (Fukumoto et al., 2018), the group identified several genes which are expressed in GABAergic interneurons and required for their migration and maturation to reach their targets. In their comprehensive list, they also found *Gm14204* lncRNA to be highly upregulated in these interneurons. We, therefore, assumed whether this lncRNA could regulate the transcription factor binding genes that are required for interneuron maturation. To test this hypothesis, qRT-PCR was performed on nine target genes from RNA isolated from E12.5 mouse brains. These genes included: *Gad67, Gad65, Htr3a, Sst, Lhx6, Dlx1, Dlx2, Dlx5,* and *Dlx6*. These nine genes were selected based on criteria that they are proven molecular markers for identification of GABAergic interneuron population. qRT-PCR results showed no significant difference in $Gm14204^{pATTS/pATTS}$ E12.5-E13.0 mouse brains (n=4 each), for eight of these genes, except *Dlx6. Dlx6* was significantly reduced in *Gm14204* mutants (**Fig. 29**).

Recently, a novel long noncoding RNA *Evf-2* transcribed from the conserved intergenic region between *Dlx5/Dlx6* gene loci was shown to regulate the transcription of homeodomain transcription factors DLX5 and DLX6 in developing mouse forebrain. *Evf-2* transcription regulates *Dlx5/Dlx6* genes negatively and is required for the development of GABAergic interneurons from the medial ganglionic eminences (Bond et al., 2009). Since the expression of *Evf-2* (Ponting et al., 2009) is strikingly similar with the expression profile of *Gm14204* lncRNA in developing forebrain, we asked if *Evf-2* was affected in *Gm14204* mutant brain. Q-RT-PCR analysis revealed no significant changes in the expression of *Evf-2* lncRNA in $Gm14204^{pATTS/pATTS}$ E12.5-E13.0 mouse brains as compared to WT (**Fig. 29**).
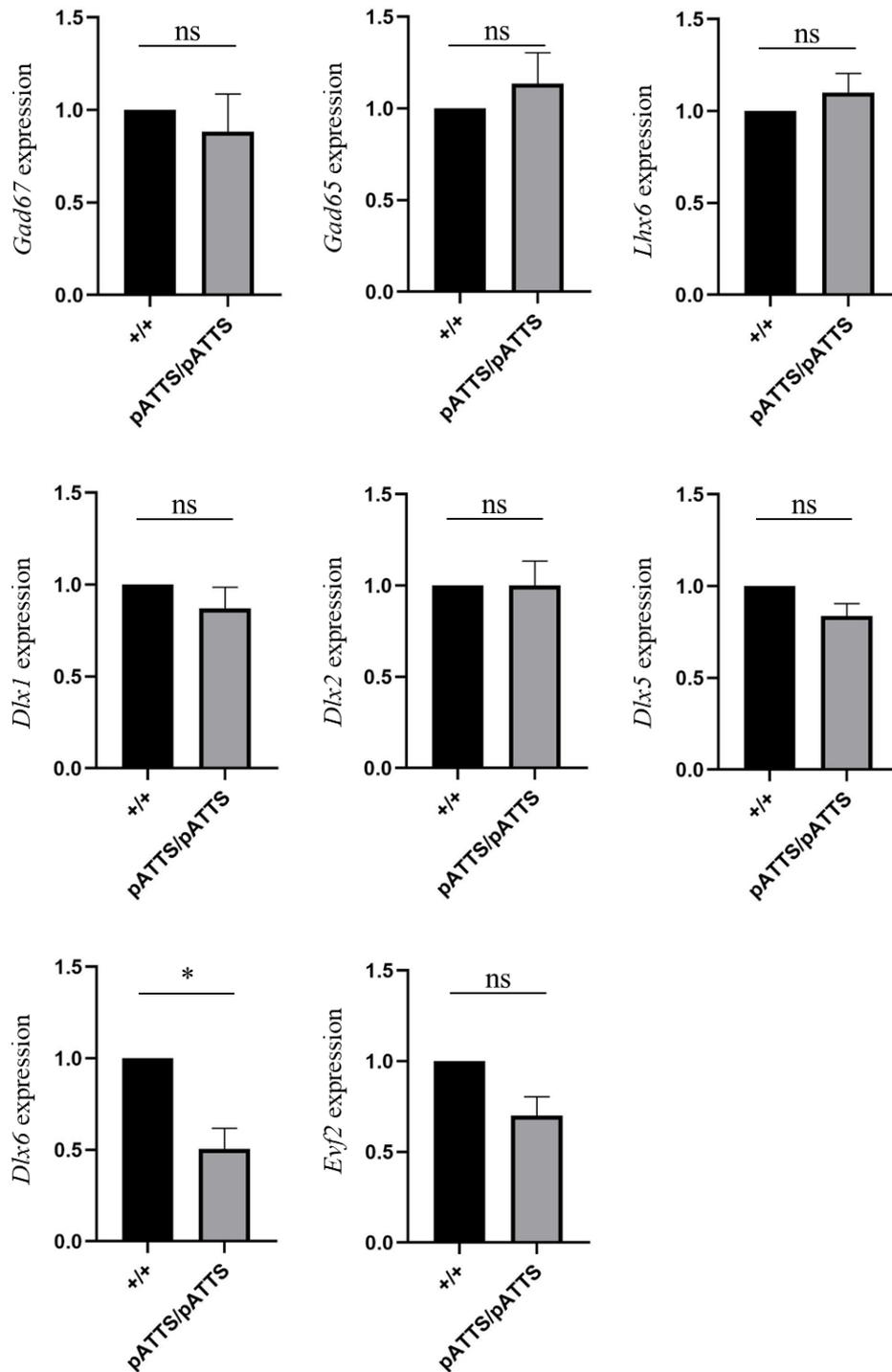
Figure 29: Quantitative RT-PCR on E12.5-E13.0 mouse brains to investigate whether *Gm14204* regulates genes involved in GABAergic interneuron development. No change was observed in the expression of *Gad67, Gad65, Lhx6, Dlx1, Dlx2, Dlx5 mRNAs and Evf2* lncRNA (ns: not significant). *Dlx6* mRNA showed significant downregulation in KO brains. All values are normalized with the housekeeper gene *Rpl13a*. Error bars indicate S.E.M. *$p < 0.05$ (n=4 each), Unpaired Student's t-test with Welch's correction.

2.2.4 *Gm14204$^{pATTS/pATTS}$* mice do not show any visible tissue abnormalities

We next wanted to know whether *Gm14204$^{pATTS/pATTS}$* mice showed any structural or morphological abnormalities. To check this, Thionine Acetate (*Nissl*) staining was performed on E14.5 embryos and P2 head sections. This method involves the use of a basic dye (here Thionine) to stain RNA in blue and that marks important neuronal structures. The Nissl substance (rough endoplasmic reticulum) gives a dark blue color due to the staining of ribosomal RNA. Inside the nuclei, the DNA is also stained blue. In E14.5 sagittal sections, we did not find any major deformities in the central or peripheral nervous system. Various tissues in the forebrain, telencephalon, mesencephalon, rhombencephalon and the spinal cord in the *Gm14204$^{pATTS/pATTS}$* showed no morphological defects as compared to *Gm14204$^{+/+}$* (**Fig. 30**). Similarly, in the young born P2 mouse heads of *Gm14204$^{pATTS/pATTS}$*, we did not observe any tissue abnormalities in the brain or developing eye (**Fig. 31**).

Next, we performed *Nissl* staining on four-week-old (P28) mouse brain coronal sections to look for the brain anatomical differences in the knockout mice. Major structures in the brain of *Gm14204$^{pATTS/pATTS}$* animals looked comparable to the WT structures with no visible differences (**Fig. 32**). Similar observations were reported from *Nissl* staining performed on ten-week-old mouse brain sections (**Fig. 33**). Thus, from the above experiments, we conclude that although *Gm14204* lncRNA has a strong expression in the developing nervous system, its constitutive ablation does not affect the overall morphology and tissue structure of the mouse nervous system. Moreover, the lncRNA knockout mice also did not show any overall body abnormalities and the mutant mice appeared healthy and viable (**Fig. 34**).
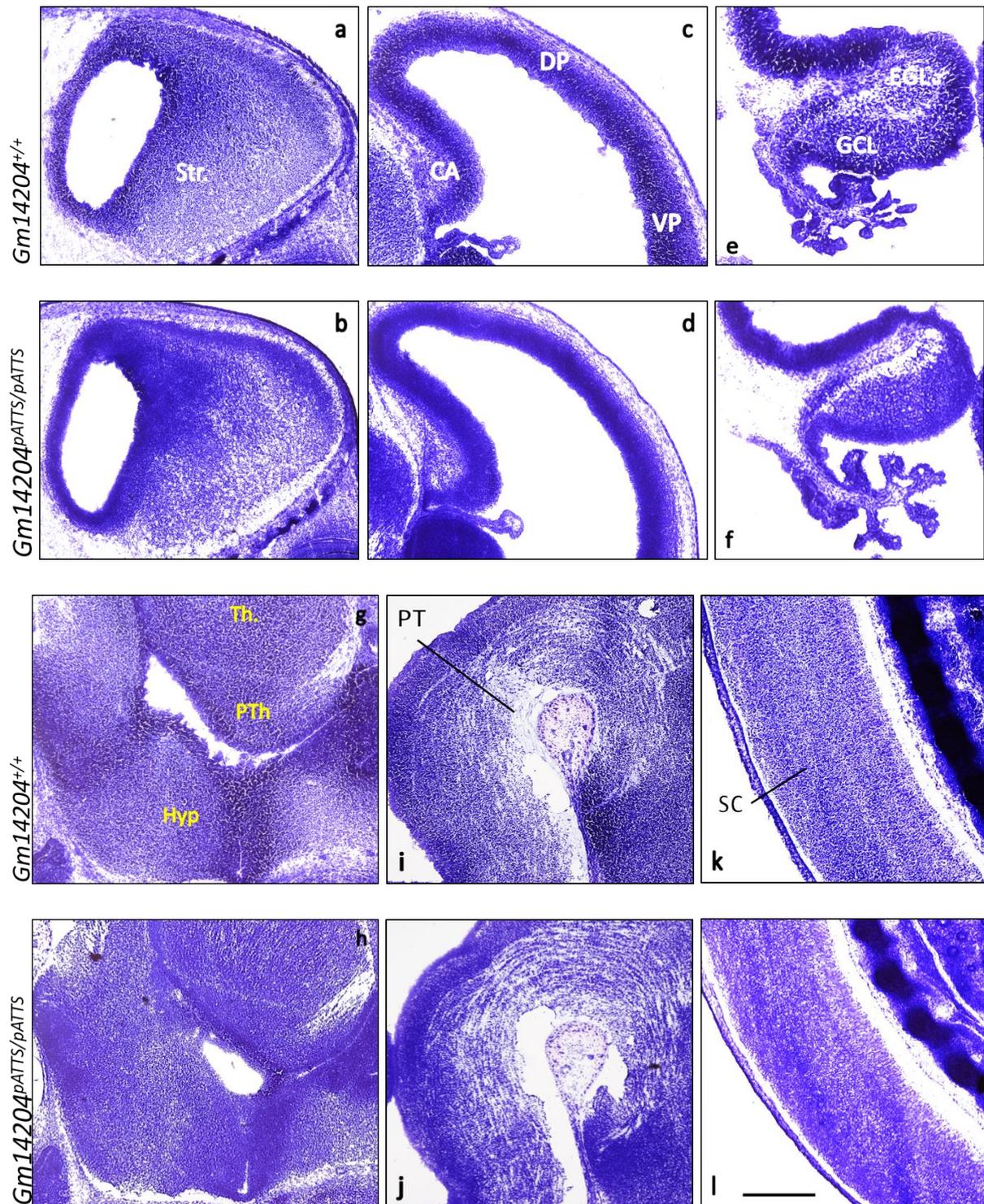
Figure 30: Thionine acetate staining on E14.5 *Gm14204^{pATTS/pATTS}* sagittal (20μm) sections did not show tissue-specific abnormalitites as compared to *Gm14204^{+/+}*. a,b: Striatum; c,d: Neocortex (DP: dorsal pallium, VP: ventral pallium, CA: cornu ammonis); e,f: Cerebellum (EGL: extra granule layer, GCL: granule cell layer); g,h: Th: Thalamus, PTh: pre-thalamus, Hyp: hypothalamus; i,j: Pre-tectum; k,l: Spinal cord (SC). Scale bar: 500 μm.
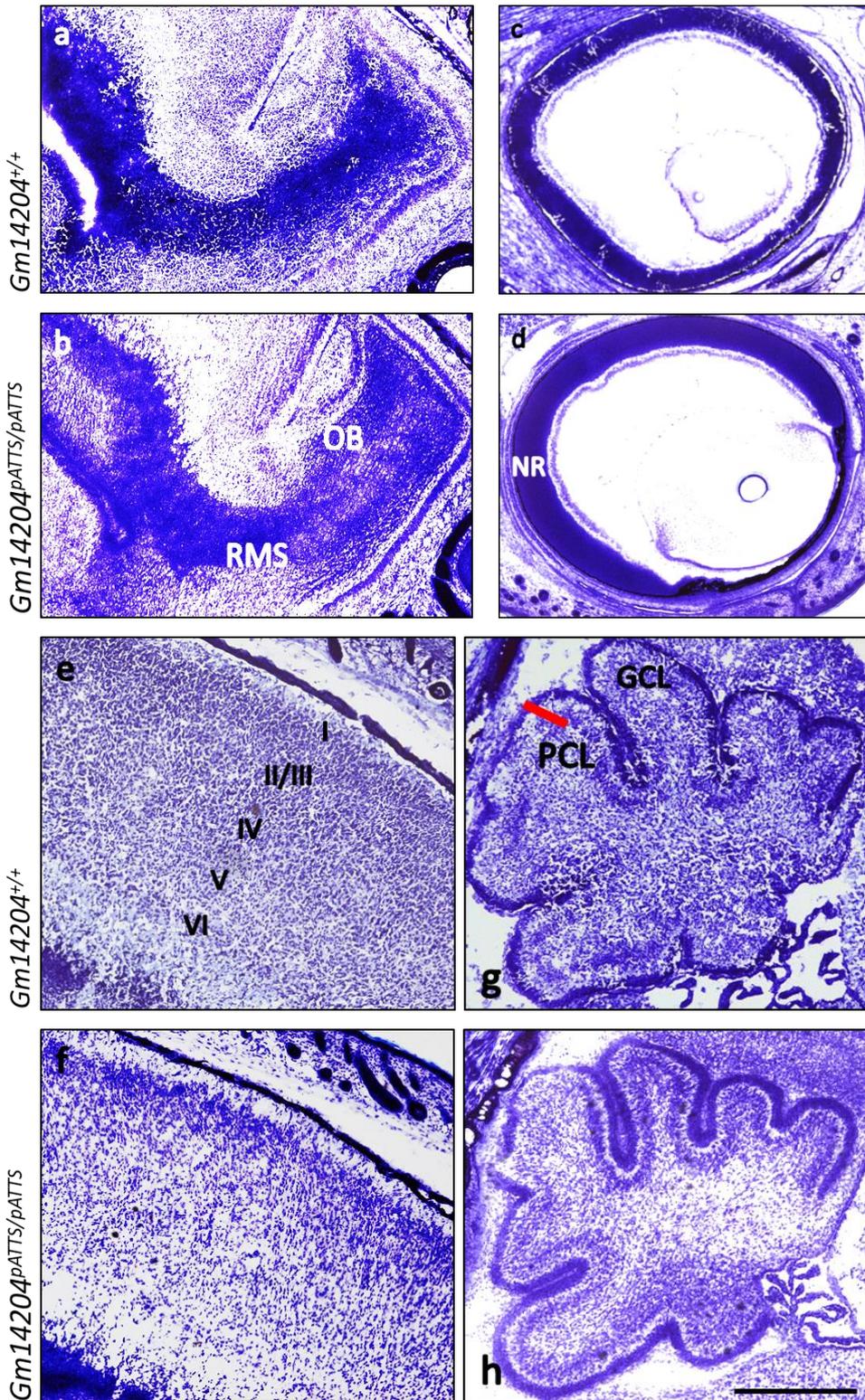
Figure 31: Thionine acetate staining on mouse P2 heads of *Gm14204^{pATTS/pATTS}* sagittal (20µm) sections did not show tissue-specific abnormalitites as compared to *Gm14204^{+/+}*. a,b: OB: olfactory bulb, RMS: rostral migratory stream; c,d: Neural retina (NR); e,f: Cortex (Layers I – VI); g,h: Cerebellum (PCL: purkinje cell layer (red line), GCL: granule cell layer). Scale bar: 500 µm.

a $Gm14204^{+/+}$

b $Gm14204^{pATTS/pATTS}$

c $+/+$
I
II/III
IV
V
VI

d $pATTS/pATTS$
I
II/III
IV
V
VI

e $+/+$
CA1
CA2
CA3
DG

f $pATTS/pATTS$

g $Gm14204^{+/+}$
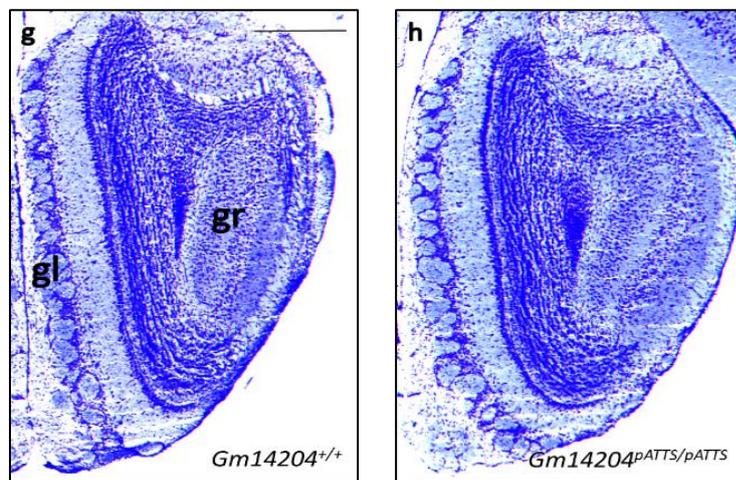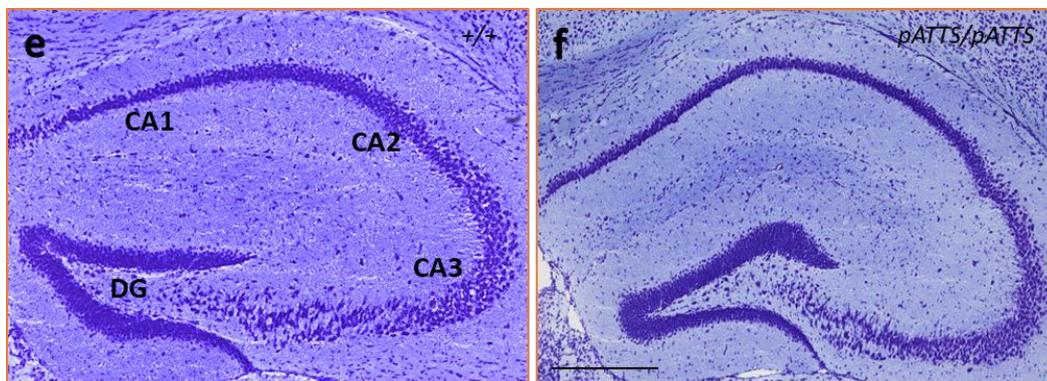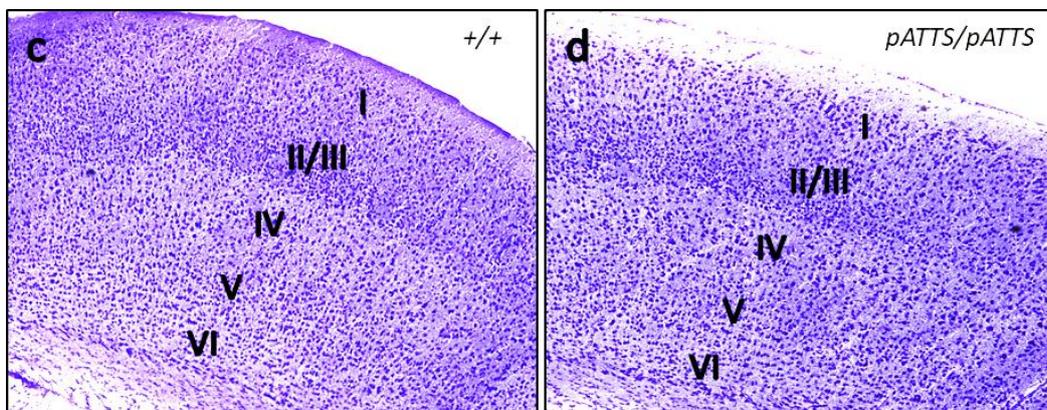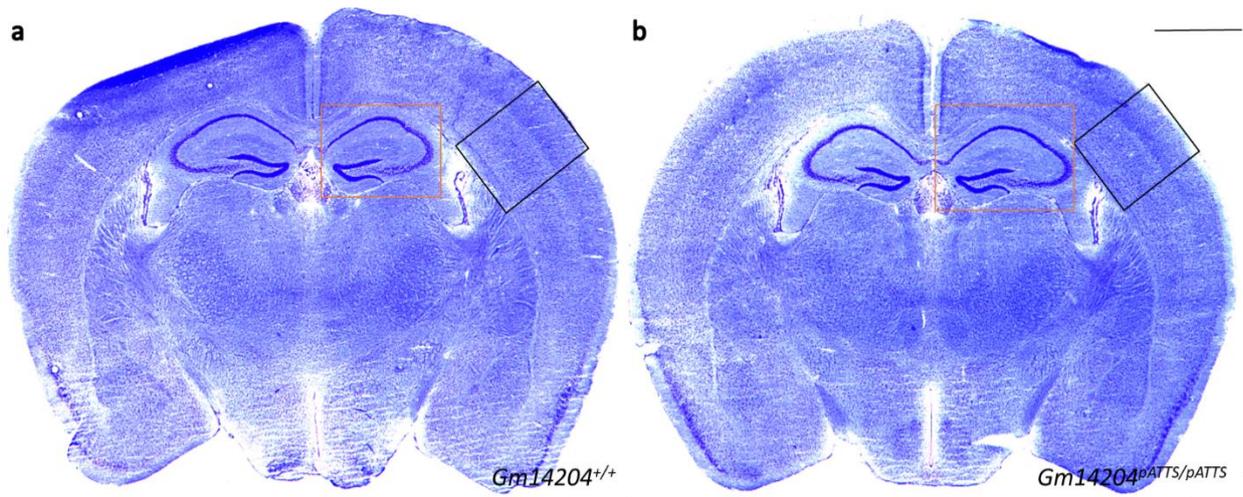gl
gr

h $Gm14204^{pATTS/pATTS}$

Figure 32: Thionine acetate staining on mouse four-week-old (P28) brains of *Gm14204^pATTS/pATTS* (20µm) sections did not show tissue-specific abnormalitites as compared to *Gm14204^+/+*. a,b: Overall view of coronal section of mouse brain; c,d: Cortical layers (I-VI); e,f: hippocampus (DG: dentate gyrus); g,h: olfactory bulb (gr: granule layer, gl: glomerular layer). Scale bar: a,b: 2 mm; c-f: 500 µm; g-h: 200 µm.

Figure panels:
- a: *Gm14204^{+/+}*
- b: *Gm14204^{pATTS/pATTS}*
- c: +/+ (cortical layers I, II/III, IV, V, VI)
- d: *pATTS/pATTS* (cortical layers I, II/III, IV, V, VI)
- e: +/+ (CA1, CA2, CA3, DG)
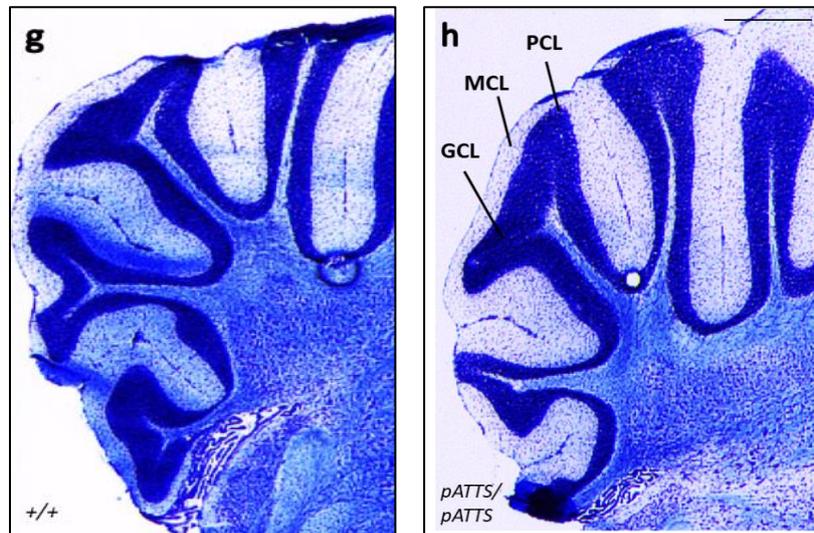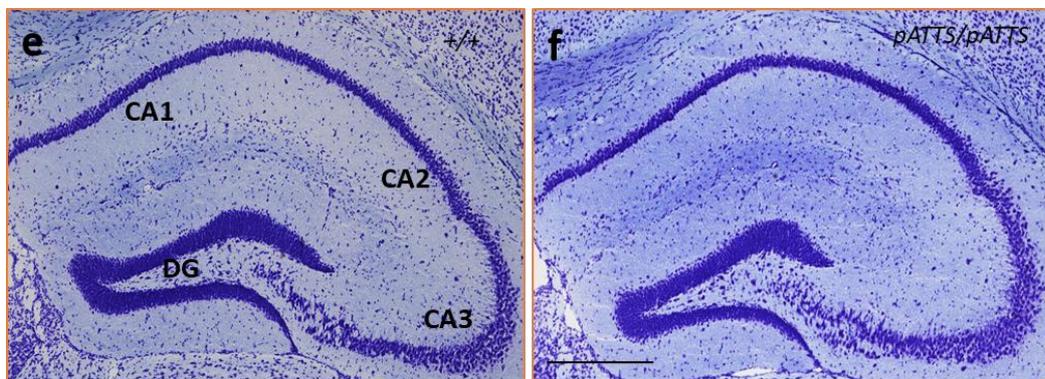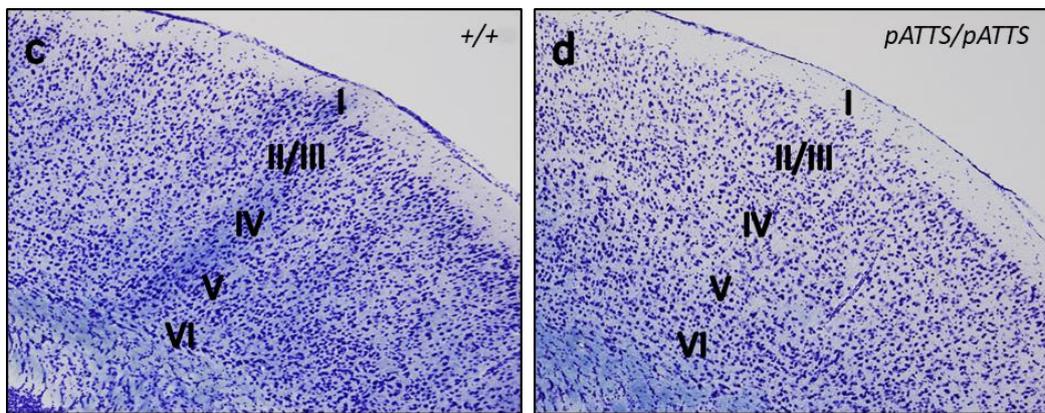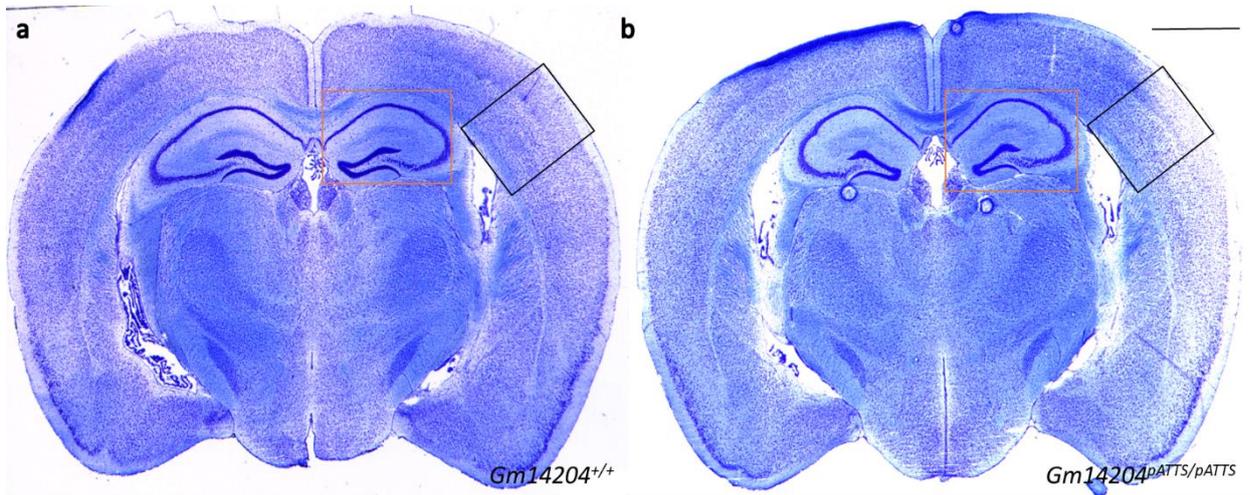- f: *pATTS/pATTS*
- g: +/+
- h: *pATTS/pATTS* (PCL, MCL, GCL)

Figure 33: Thionine acetate staining on mouse ten-week-old (P70) brains of *Gm14204^pATTS/pATTS* (20µm) sections did not show tissue-specific abnormalitites as compared to *Gm14204^+/+*. a,b: Overall view of coronal section of mouse brain; c,d: Cortical layers (I-VI); e,f: hippocampus (DG: dentate gyrus); g,h: cerebellum (GCL: granule cell layer, PCL: purkinje cell layer, MCL: molecular cell layer). Scale bar: a,b: 2 mm; c-f: 500 µm; g-h: 200 µm.
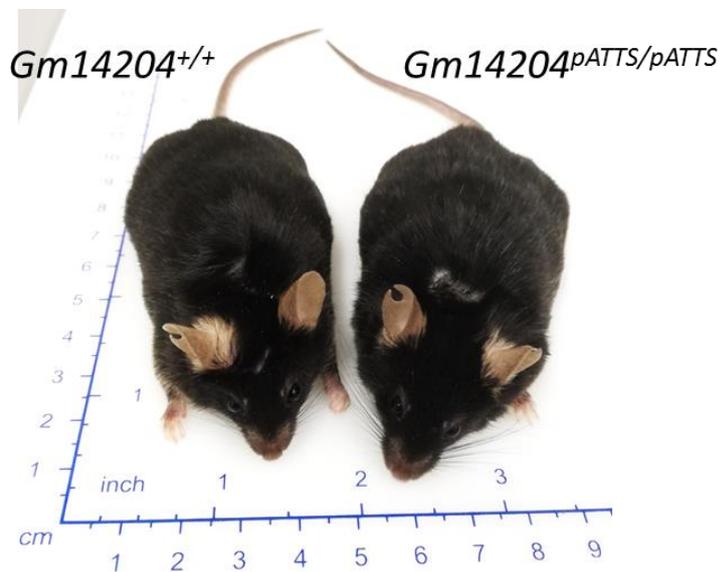

Figure 34: *Gm14204^pATTS/pATTS* mice viable and healthy similar to *Gm14204^+/+* mice. No outer gross morphological abnormalities were observed in adult lncRNA mutants (above shown 10 months old mice). They looked normal and healthy based on their physical appearance (n = 7-9 each).

3.0 Discussion

Before a decade, the significance of long noncoding RNAs in the mammalian genome was quite controversial (Ponjavic et al., 2007; Wilusz, Sunwoo, & Spector, 2009). But now it is clear that a large portion of the mammalian genome generates lncRNAs, many of which have already been implicated to have roles in mouse development and organogenesis (reviewed in Grote & Herrmann, 2015; Schmitz, Grote, & Herrmann, 2016). In this study first, we have performed RNA-seq of E14.5 whole mouse embryos and generated a comprehensive atlas for nearly 7000 lncRNAs (RPKM ≥ 0.1) expressed during E14.5 stage of mouse embryogenesis and *de novo* assembled them to the mouse genome. E14.5 developmental stage was preferred because at this time point organogenesis is nearly complete and the majority of key brain events start to occur and neurogenesis has already reached its peak in the brain (Molyneaux et al., 2007). To date, our sequencing data is the first in-depth characterization of total lncRNAs expressed at E14.5 developmental stage. Although we have mapped nearly 7000 lncRNAs to the mouse genome, it is challenging to determine how many of these transcripts are functional, and how many of them are just transcriptional noise resulting from transcription of protein-coding genes. We have classified these lncRNAs into sub-classes such as intergenic, intronic, sense and antisense, yet our attention was primarily on understanding the functional mechanism of divergent lncRNAs, as this category of lncRNA have shown to regulate the expression of their adjacent protein-coding gene pairs (Grote et al., 2013; Liu et al., 2017; Sigova et al., 2013; Vance et al., 2014; Wu & Sharp, 2013). Our deep sequencing data revealed nearly 700 lncRNA transcripts of divergent class. Based on their expression values (RPKM), conservation scores, important gene regulatory loci, and an unbiased approach, we filtered out 10 lncRNAs, which were used for RNA *in situ* hybridization (ISH). RNA-ISH for six of them revealed similar expression patterns with their neighboring coding gene pairs. This sharing of similar expression patterns with their coding

gene mRNAs might be either due to the same transcriptional machinery used by the two genes or such gene pairs might share common promoter sequences, meaning both genes are transcribed from a bi-directional promoter.

3.1 Mechanism of divergent lncRNA *Gm14204* in *cis*

We have explored the spatiotemporal expression profile of a previously uncharacterized lncRNA *Gm14204* in the developing and adult mouse. *Gm14204* lncRNA resides on an important developmental gene locus and is transcribed divergently to mammalian inhibitory neurotransporter gene *Slc32a1*. Interestingly, both these genes share an identical expression profile in the embryonic mouse nervous system, but not in adult mouse brain. As per our hypothesis, we had assumed that *Gm14204* could regulate the expression of *Slc32a1* gene at a transcriptional level. To validate our hypothesis, we generated a first knockout mouse model for this lncRNA under B6N background. Many strategies have been used in the field to knockout lncRNAs (Bassett et al., 2014). Other groups have concluded that the knock-in of single or triple polyA successfully terminates lncRNA transcription. However, one study found that even after knock-in of four times polyA in the *Haunt* lncRNA gene locus only abrogated 65% of Haunt transcripts (Yin et al., 2015). Therefore, to make our strategy more effective and to avoid the risk of leaky transcription, we decided to knock-in a strong polyA signal followed by four transcription termination sequences (TTS) (Schwalb et al., 2016) just at the beginning of the first intron of *Gm14204* lncRNA gene. Our novel strategy successfully abrogated complete lncRNA transcription. Since the insertion in the first intron, the first exon of lncRNA that is almost 50 bp will still be transcribed, but due to its instability, it will be prone to and degraded by RNA exonucleases. Thus, in the end, there is no *Gm14204* lncRNA transcript been generated as a side product. Quantitative RT-PCR analysis confirmed this results. Next, as per our hypothesis, we determined the expression levels of *Slc32a1* mRNA in

E12.5-E13.0 lncRNA knockout mice brains ($Gm14204^{pATTS/pATTS}$) and found the mRNA level comparable to wild-type ($Gm14204^{+/+}$). This led us to prove that *Gm14204* is not regulating *Slc32a1* gene at a transcriptional level. Therefore, the co-localization of *Gm14204* and *Slc32a1* in Fig. 22 might result only due to the close proximity of both genes and not essentially because of *Gm14204* lncRNA acting on *Slc32a1* gene locus.

Since lncRNAs function in *cis* to regulate the expression of genes on the same chromosome, we checked four other regulatory important genes apart from *Slc32a1*, which are located in close proximity of the *Gm14204*. We observed no significant changes in three genes (*Ralgapb*, *Fam83d,* and *Dlx35*), but *Actr5* (Actin-related protein member 5) was significantly decreased in $Gm14204^{pATTS/pATTS}$ embryonic mouse brains.

3.2 Regulation of development-related genes by *Gm14204* lncRNA

LncRNAs show high tissue and cell-specific expression profiles, more specific than protein-coding genes (M. Cabili et al., 2011; Mercer et al., 2008). This extensive amount of tissue specificity has been linked to important regulatory roles of lncRNAs in diverse cellular processes. Mammalian brain, which is considered to be highly complex in terms of both structure and function, is known to be the richest source of lncRNAs in the body (Mattick, 2011). In our study, we have observed an enriched expression of *Gm14204* lncRNA in the developing central nervous system of the mouse. Moreover, in the embryonic mouse brain, *Gm14204* lncRNA shows a highly patterned expression profile that is comparable to several protein-coding mRNAs. For instance, the lncRNA has an overlapping expression with that of Glutamate acid decarboxylase isoforms (*Gad1/2*) (source: GenePaint database) in various brain compartments. *Gad1/2* are GABA synthesizing enzymes that are required for neurotransmitter production in the mammalian brain. They are also responsible for the

production of GABAergic interneuron precursors in the ganglionic eminences, where *Gm14204* has a high expression. Double knockout of *Gad1/2* genes results in several tissue abnormalities and loss of GABA synthesis in the brain (Kakizaki, Oriuchi, & Yanagawa, 2015). The transcriptional regulation of *Gad* gene isoforms is still unclear. We tested whether or not *Gm14204* has a role in regulating *Gad* genes. Quantitative RT-PCR analysis from our work showed that the expression levels of *Gad* isoforms remained unaffected in *Gm14204$^{pATTS/pATTS}$* mice, suggesting that this lncRNA does not regulate *Gad* genes.

Several *Distal-less* homeobox genes (*Dlx*) such as *Dlx1, Dlx2, Dlx5,* and *Dlx6* are necessary for the differentiation of GABAergic interneurons and mutation in these genes results in a massive reduction of these interneurons in the brain. Specifically, knockout of *Dlx6* in mice affects interneuron migration and specification at E13.0 day (Fukumoto et al., 2018; Panganiban & Rubenstein, 2002; Y. Wang et al., 2010). More recently, it was shown that a lncRNA called *Evf-2* transcribed from an ultra-conserved region between *Dlx5* and *Dlx6* genes in mice regulates *Dlx5*/*Dlx6* negatively and also interneuron synthesis is decreased in *Evf2* mutant mouse brains (Bond et al., 2009). Interestingly, *Gm14204* lncRNA shares similar expression patterns with all *Dlx* genes and *Evf-2* lncRNA in the developing forebrain (Source: GenePaint database). This led us to inquire if *Gm14204* lncRNA regulated *Dlx* genes. Quantitative RT-PCR analysis of E13.0 *Gm14204$^{pATTS/pATTS}$* mouse brains showed no significant changes in *Dlx1*, *Dlx2*, *Dlx5* mRNAs or *Evf-2* lncRNA as compared to WT. However, the expression of *Dlx6* mRNA was significantly decreased in *Gm14204$^{pATTS/pATTS}$* brains. An important question now is whether the lncRNA regulates *Dlx6* directly or indirectly. More questions arise, whether or not *Gm14204* regulates other genes in the genome. Our smFISH results clearly indicate that *Gm14204* lncRNA has multiple localization centers inside the cell nucleus (**Fig. 23**), indicating a *trans* function of this lncRNA.

3.3 *Gm14204^{pATTS/pATTS}* female mice might suffer from a lactation phenotype

While breeding *Gm14204* homozygous pairs, we observed an unusual phenotype shown by the *Gm14204^{pATTS/pATTS}* mothers. We found that the *Gm14204^{pATTS/pATTS}* mothers showed an aggressive phenotype and killed all their own newborns within a week of birth. We observed this behavior in three out of four homozygous matings performed so far. While looking at the new-born pups closely we found that they apparently lacked a milk spot in their stomach. Milk spots are visualized to confirm that the pups are feeding properly. Perhaps because of this reason, the mothers harmed and tried to kill them. To overcome this situation and obtain more *Gm14204^{pATTS/pATTS}* mice for analysis, we transferred the pups to the foster mothers as soon as they were born. Now, the pups fed well and remained healthy. No deaths were found when the foster mother fed the mice. This allowed us to conclude that the *Gm14204^{pATTS/pATTS}* mothers were unable to lactate after giving birth and presumably killed the newborns. Additionally, we observed that the heterozygous mothers also showed a mild phenotype since we found dead cadavers of the pups inside the cages of these mice. Genotyping (data not shown) revealed that these heterozygous mothers killed the newborns independent of the offspring genotype, i.e. not necessarily the mutants. Since only three homozygous matings have confirmed the cannibalism behavior by mothers, more such matings need to performed in order to validate these findings.

Histology staining of one out of three homozygous mutant female mice showed the absence of major thalamic-hypothalamic areas in two weeks old brain (**Fig. 35**). These areas coincidently include structures responsible for lactation in female mice. The magnocellular neurons of the hypothalamus arising from the paraventricular nucleus (PVN) and the supraoptic nucleus (SON) are required for the production of the lactating hormone oxytocin. This hormone is released into the posterior lobe of pituitary gland from where it is transported

into the blood-stream (reviewed in Jurek & Neumann, 2018). We assume that the absence of these neurons might lead to a deficiency of oxytocin and subsequently lack of milk production in *Gm14204^{pATTS/pATTS}* mothers. Although it is too early to say that the thalamic-hypothalamic brain structure is absent in all lncRNA mutants and additional histology staining needs to be carried out to determine the penetrance of this phenotype.
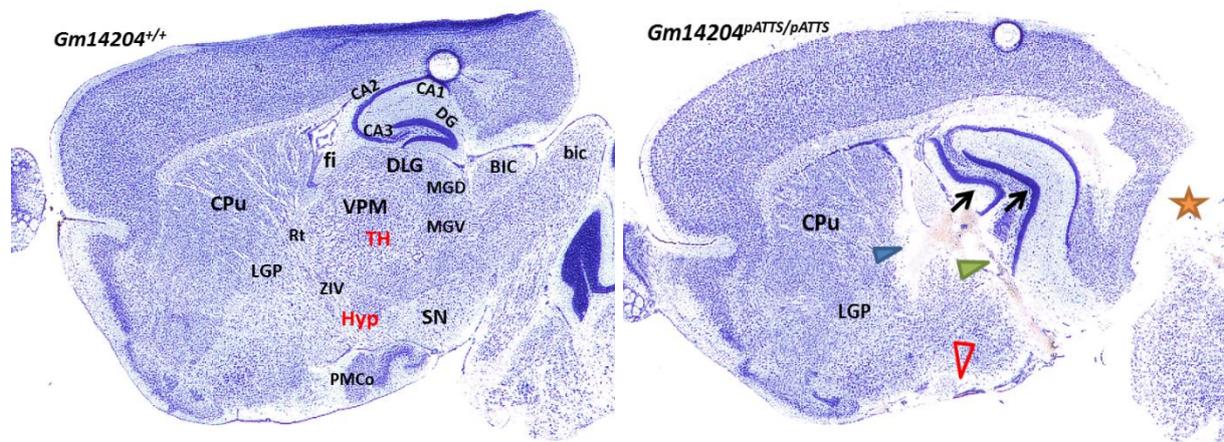

Figure 35: Two weeks old brain of *Gm14204^{pATTS/pATTS}* female mouse shows the absence of important structures and major deformities. Missing structures in the mutant brain include ventral posteromedial thalamic nucleus (VPN), Rt (reticular nucleus), ZIV (zona incerta, ventral part) (blue arrowhead); dorsal lateral geniculate nucleus (DLG), medial genic nucleus, dorsal part (MGD) and ventral part (MGV) (green arrowhead), posteromedial cortical amygdala nucleus (PMCo) (red arrowhead); and superior and inferior colliculus (BIC/bic) (orange star). Hippocampus showed deformities in the CA3 layer and the dentate gyrus (DG) layer (black arrows). CPu: caudate putamen, LGP: lateral globus pallidus, SN: substantia nigra, TH: thalamus, Hyp: hypothalamus.

3.4 *Gm14204* lncRNA knockout is not lethal in mice

Gm14204 is a highly expressed and evolutionary conserved lncRNA. To understand its biological function, we have developed a successful knockout mouse model through polyA and termination sequences knock-in strategy. This model led us to conclude that *Gm14204* is dispensable for mouse viability and development. It is surprising that although the lncRNA has a tissue-specific expression during embryonic stages, its knockout is not lethal for mouse development. *Gm14204* is located on an important genomic regulatory locus and is transcribed divergently to *Slc32a1* gene, whose knockout leads to mouse lethality. From our study, although we validate that this lncRNA has no effect on *Slc32a1* gene transcription, it can still have subtle effects on regulating post-transcriptional gene activity.

Many lncRNA knockout mouse models derived since the last decade have displayed very subtle phenotypes (Bond et al., 2009; Eißmann et al., 2012; Oliver et al., 2015; Sauvageau et al., 2013), with a few exceptions where lncRNA knockouts led to mouse lethality (Grote et al., 2013; Sauvageau et al., 2013). We predict that *Gm14204* lncRNA might also have a very subtle phenotype as seen for *Evf-2* lncRNA. Similar to *Gm14204*, *Evf-2* lncRNA is also broadly expressed in the GABAergic interneurons and has similar expression pattern in the ventral embryonic forebrain. *Evf-2* mutant mice were viable but showed a decreasing number of GABAergic interneurons compared to the wild-type mice (Bond et al., 2009). Such subtle phenotype can be expected, although more experiments might be necessary to check the GABAergic cell population in our lncRNA knockouts.

4.0 Materials and Methods

4.1 Chemicals:

Agarose	BioZYM, Germany
Tris-(hydroxymethyl)-aminomethane	VWR chemicals, Belgium
Sodium Chloride	Merck, Germany
Potassium Chloride	Merck, Germany
Ethylene di-amine Tetra Acetate	Sigma Aldrich, Germany
RNAlater solution	Ambion, USA
Gibco DPBS (1X)	Life Technologies, Germany
Sodium Acetate Buffer solution	Sigma Aldrich, Germany
NP-40 Surfact-Amps Detergent Solution	ThermoScientific, Switzerland
Gel loading Dye Purple (6X)	New England Biolabs, Germany
DNA ladder 100 bp	New England Biolabs, Germany
RNaseZAP	Sigma Aldrich, Germany
Ethanol	Merck, Germany
RNase A solution	Sigma Aldrich, Germany
IGEPAL CA-630	Sigma Aldrich, Germany
Tween 20	Sigma Aldrich, Germany
Triton-X 100	Merck, Germany
Chloroform	Merck, Germany
2-Propanol	Merck, Germany
Xylol	Merck, Germany
Acetic Acid	Merck, Germany
TRIzol Reagent	Ambion, USA
Antigen Unmasking solution	Vector Laboratories, UK
Lithium Chloride Precipitate Solution	Ambion, USA
Ammonium acetate solution	Sigma Aldrich, Germany
Dimethyl Sulphoxide (DMSO)	Sigma Aldrich, Germany
Dithiothreitol (DTT)	Sigma Aldrich, Germany
Hydrochloric Acid (HCl)	Merck, Germany

ß-Mercaptoethanol   Sigma Aldrich, Germany

4.2 Commercial Kits:

RNeasy Plus Mini Kit   Qiagen, Germany

Single molecule FISH   ACDBio, USA

QIAquick PCR purification kit   Qiagen, Germany

QIAquick gel extraction kit   Qiagen, Germany

iQ SYBR Green Supermix   Bio-Rad, Germany

SuperScript III First-Strand Synthesis System   Thermo Scientific, Switzerland

4.3 Enzymes:

Proteinase K   New England Biolabs, Germany

Complete protease inhibitor-EDTA free   Roche, Germany

T7 RNA polymerase   New England Biolabs, Germany

SP6 RNA polymerase   New England Biolabs, Germany

Taq DNA polymerase   New England Biolabs, Germany

SUPERase-In RNase inhibitor   Invitrogen, USA

RNasin Ribonuclease inhibitor   Promega, USA

DNAseI recombinant RNase-free   Roche, Germany

Superscript III   Thermo Scientific, Switzerland

RNase OUT   Thermo Scientific, Switzerland

E.coli RNase H   Thermo Scientific, Switzerland

4.4 Buffers/Solutions:

PBND lysis buffer (Genotyping) (provided by Technical Staff)**:**

10mM Tris (pH 8.3)

50mM KCl

2.5mM $MgCl_2$

0.1mg/ml Gelatine

0.45 % (v/v) NP40

0.45 % (v/v) Tween20

TAE (50x, 1L):

242 g Tris-Base

57.1 ml acetic acid

100 ml 0.5 M EDTA (pH 8.0)

Solutions for RNA-ISH:

All the solutions for RNA-ISH were prepared by the technical staff as per Eichele & Diez-Roux, (2011).

4.5 Tissue isolation:

Mice were killed by cervical dislocation or by CO_2 if required. For RNA isolation, brains or other organs were quickly removed and stored in RNAlater solution at 4°C overnight to stabilize the RNA and prevent it from degradation. For histology, the brains and embryos were dissected quickly, placed in pre-cool OCT medium for 15-20 min before embedding them into the freezing chambers. The exact procedure for the freezing of mouse embryos or brains was followed as per the protocol from the Eurexpress database. (http://www.eurexpress.org/ee/technology/resources/SOP_Dissection_291104.pdf).

4.6 RNA isolation:

RNA from cells and tissues was isolated using the RNeasy Plus Mini Kit (Qiagen) according to the manufacturer's instructions. All centrifugations were performed at maximum speed (13,000 rpm) at RT. Briefly, cells ($< 10^7$) were harvested as cell pellet, incubated with 600 µL of RLT buffer containing ß-mercaptoethanol and vortexed for 30s to 1 min. Tissues (< 30 mg) either stored in RNAlater at -20°C or collected fresh were incubated with 600 µL of RLT buffer with ß-mercaptoethanol and transferred to 2 mL screw-capped tubes with ceramic balls. Tissues were homogenized using Beadruptor for 3x 30s at speed 5 and a 30s pause step in-between. The homogenate was spun down at maximum speed for 3 min and the supernatant was transferred to a gDNA Eliminator spin column placed in a 2 mL collection tube. The samples were centrifuged for 30s at maximum speed, the columns were discarded and the flow through was saved. One volume of 70% ethanol was added, mixed well and 700 µL was immediately loaded on to the RNA collection spin column. The spin columns were centrifuged for 15s and flow-through discarded. This step was repeated in order to spin down the total 1200 µL of sample. Spin columns were washed with 700 µL of Buffer RW1, centrifuged as above and flow through discarded. 500 µL of Buffer RPE was added to the

columns and centrifuged for 15s. This step was repeated and the centrifugation was performed for 2 min. The columns were transferred to a fresh 2 mL collection tube and centrifuged for 1 min to remove any residual RPE solution. Next, the columns were placed in fresh RNase free eppendorfs and 20-30 µL of RNase free water was added. Centrifugation was performed for 1 min to elute pure RNA. The RNA concentration was quantified using Nanodrop 1000. RNA was either used immediately for cDNA synthesis or stored at -80ºC.

4.7 RNA-sequencing and assembly of lncRNAs:

Total RNA was extracted by TRIZOL method (as per manufacturer's instructions) from three mouse embryos (E14.5) and was used for strand-specific library construction and RNA sequencing with single ended 100 bp reads, which was then used for *de novo* lncRNA assembly. The RNA-seq and analysis were performed at the Functional Genomics Group, PICB, Shanghai, China. LncRNAs were assembled *de novo* and combined with annotated lncRNAs from public databases of Ensembl, RefSeq, and UCSC. The discovery pipeline of lncRNAs was adapted from the study of (M. Cabili et al., 2011) and the transcripts were assembled by Cufflinks (Trapnell et al., 2010). To alleviate the mapping of reads that span splice site junctions, Ensembl gene annotation, mRNAs and ESTs of mouse mm9 genome from UCSC genome browser were used to guide the recognition of splice junctions. This was performed majorly by using software like TopHat and Cufflinks. All the genes/transcripts which overlapped with the protein-coding genes, tRNAs, rRNAs and pseudogenes on the same strand were removed. Transcripts containing more than one read were retained to gain higher confidence in the data. All single-exonic transcripts were also removed and only multi-exonic transcripts were retained. Protein coding potential for all lncRNA transcripts was evaluated using the Coding Potential Calculator tool (Kong et al., 2007) and transcripts with no coding capability were retained (CPC score<0). Additionally, redundant transcripts were

also removed from our data. The final screen of lncRNA transcripts was done based on RPKM (Reads Per Kilobase of transcript per Million mapped read) values. All transcripts with RPKM > 0.1 were retained. Using this screening procedure, we were left with ~7000 lncRNA transcripts (expressed in at least one biological replicate) in E14.5 mouse embryos.

4.8 cDNA synthesis:

cDNA was synthesized from RNA isolated from various tissues using Invitrogen Superscript III First-strand Synthesis System for RT-PCR kit according to Manufacturer's instructions with some modifications. The first mix was prepared using 50 µM oligo(dT)$_{20}$, 10 mM dNTP mix, 0.1 µg – 1 µg RNA and DEPC treated water made up to 10 µL. The reaction was incubated at 65°C for 5 min and then placed on ice. The second cDNA synthesis mix was prepared using 10X RT (reverse transcriptase) buffer, 25 mM MgCl$_2$, 0.1 M DTT, 40 U/µL RNaseOUT^{TM} and 200 U/µL Superscript III RT. 10 µL of this mix was added to the RNA mixture, mixed gently and collected by brief centrifugation. The mix was incubated first at 50°C for 50 min and then the reaction was terminated at 85°C for 5 min, followed by chilling on ice. 1 µL of 2 U /µL *E. coli* RNase H was added to each tube and the reaction was further incubated for 20 min at 37°C. The resulting cDNA was stored forever at -20°C.

4.9 Template synthesis:

DNA templates were synthesized according to the protocol described in (Yaylaoglu et al., 2005) with small modifications. DNA templates (< 1 kilobase) were synthesized using PCR with gene-specific forward and reverse primers having T7 and Sp6 promoters attached at their ends respectively (T7- . Sp6-). These extended ends would facilitate RNA polymerase binding during riboprobe synthesis. All the primers were commercially ordered from Eurofins

MWG Biotech. To check an exact annealing temperature of every primer set, a temperature gradient PCR was performed (Temp: 49° to 62°C). The reaction mix was prepared using 10x B Buffer (Qiagen), 5x Q enhancer (Qiagen), 2 mM dNTP mix (Roche), 0.5 µL *Taq* DNA polymerase (5 U/µL, Qiagen), 1 µL cDNA, 5 pmol of each forward and reverse primers and water to final required volume. The PCR conditions were as follows: Initial denaturation at 94°C for 2 min, followed by 35 cycles of [25 seconds denaturation (94°C), 25 seconds of annealing (49-62°C) and 1 min extension (72°C)] and a final extension at 72°C for 9 min. 2 µL was analyzed on 1% Agarose gel. 100 bp ladder (New England Biolabs) was used as a reference marker.

To amplify the product at a specific temperature, a second PCR was carried out using similar reaction buffers and conditions (except, the annealing temperature was kept constant). PCR products were analyzed on 1% Agarose gel and purified using QIAquick PCR purification spin column kit (Qiagen). If the non-specific band was observed on an agarose gel, then the correct PCR band was gel excised and purified using Qiagen gel extraction kit. The products were quantified using Nanodrop spectrophotometer and sequence verified at Eurofins MWG Biotech. Primers used for template synthesis are listed in 4.17.

4.10 Riboprobe synthesis:

Riboprobes were synthesized according to the protocol described in (Yaylaoglu et al., 2005) with small modifications. RNA probes were synthesized by *in vitro* transcription (IVT). IVT was carried out using 10X RNA transcription buffer (New England Biolabs), 10x DIG (Digoxigenin) or FITC (Fluorescein) mix (Roche), 20,000 U/mL Sp6 RNA polymerase (New England Biolabs), 40 U/µL Rioboloc RNase Inhibitor (ThermoScientific), 1-1.5 µg DNA template and DEPC water to required final volume. The reaction mix was incubated at 37oC for 2.5 hours. To remove any residual DNA contamination, 10,000 U/mL of DNase I was

added to each mix and further incubated at 37oC for 15 min. After incubation, 108 µL of chilled 4 M ammonium acetate and 705 µL of chilled 100% Ethanol were added to each reaction mix and incubated overnight at -80°C. The following day, samples were centrifuged at 13,000 rpm for 20 min at 4°C. The supernatant was discarded and the pellet was washed with 70% Ethanol by centrifugation as above. The supernatant was discarded again, the pellet was air dried and dissolved in 44 µL DEPC treated water. 2 µL was used for quantification and remaining 42 µL was diluted in HYB-Buffer (Ambion) to a final probe concentration of 100 ng/µL and stored at -20°C until use.

4.11 Cryosectioning:

Before sectioning, the OCT blocks of embedded brains/embryos were removed from the freezer and were mounted in a Leica Cryostat and allowed to equilibrate to the chamber's temperature. Sectioning was mostly performed as sagittal sections 25 µm (or 20 µm if needed) divided into 6 sets for embryos and 20 µm divided into 8 sets for brains. SuperFrost glass slides (ThermoScientific) were used for sections. Each slide had 4 sections. Sections were fixed, acetylated, air-dried and stored in a glass box (sealed) at -20°C. All the sectioning was performed by technical staff.

4.12 RNA *in situ* hybridization (RNA ISH):

RNA *in situ* hybridization was carried out by technical staff, using TECAN Genesis liquid handling platform. Details about the procedure have been described in the protocol (Eichele & Diez-Roux, 2011).

4.13 Quantitative RT-PCR (qRT-PCR):

cDNA samples were used for qPCR at a concentration of 5 ng/µL. 10 ng cDNA was used for each reaction in a total reaction volume of 10 µL. Reaction volumes are mentioned in Table 1. 8 µL of master mix {containing iQ SYBR Green Supermix (Bio-Rad), forward and reverse primers (custom designed) and dist. water} was added to each well, with 2 µL of cDNA template. CFX96 Bio-Rad machine was used for qPCR. The program used is listed in Table 2. Data analysis was performed using Bio-Rad CFX96 3.0 software. *Gapdh* and *Rpl* house-keeping genes were used as positive controls. Dist. water was used as negative control. Primers used for qPCR are listed in 4.17.

Reagent	Volume
iQ SYBR mix	5 µL
Primer mix (6 pmol)	0.5 µL
cDNA	2 µL
Dist. H$_2$O	2.5 µL
Total	10 µL

Table 1: Reaction volumes for single qPCR reaction

Temperature	Time	Options	Cycles
95ºC	7 min		
95ºC	10 s		
60ºC	25 s		40x
72ºC	20 s	+ Plate read	
65ºC - 95ºC, incr. 0.5ºC	5 s	+ Plate read	
4ºC	Pause		

Table 2: qPCR program used by CFX96 Bio-Rad machine

4.14 Constitutive ablation of *Gm14204* lncRNA in mouse

The generation of *Gm14204* lncRNA mouse knockout was performed at Polygene Transgenetics, Switzerland. The targeting strategy and was planned by me with additional help from Polygene. All other steps starting from vector construction until the generation of heterozygous mutants were performed at Polygene. The method used by them to generate knockout mice has been mentioned below. The breeding of *Gm14204* heterozygous mice and the generation of *Gm14204* mouse colony was performed in the BTL facility of our institute (Max Planck Institute for Biophysical Chemistry, Göttingen) by me and with the help of animal house technical staff.

4.14.1 Targeting vector construction:

For the constitutive knock-in of the STOP cassette in the lncRNA gene locus, a targeting vector was generated, containing two homology arms: a short arm of 2.34 kb (SA) and a long arm of 5.11 kb (LA). The homology arms were amplified by conventional PCR from the RP23-392P11 BAC DNA template. The STOP cassette was amplified from plasmid 'pCMV-Gluc-2-8IV' (provided generously by Dept. of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Göttingen).
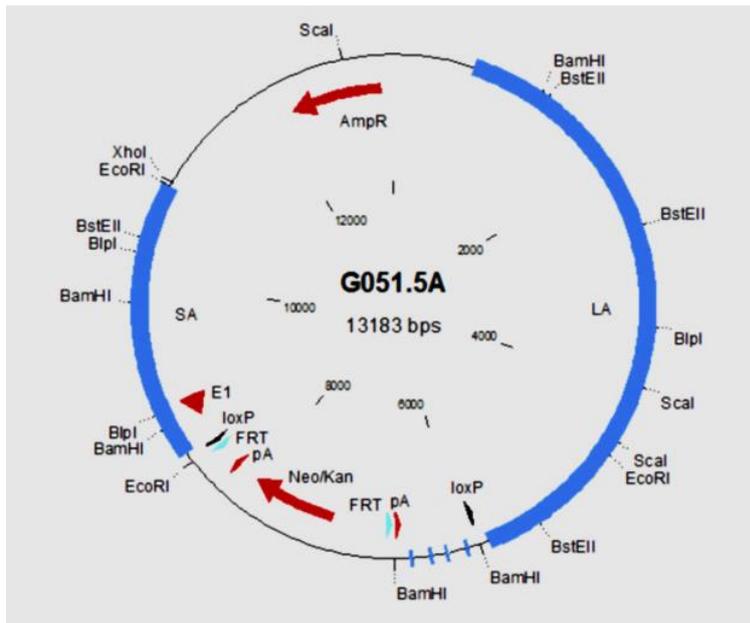
Figure 36: Map of the targeting vector used for the *Gm14204* knockout. The FRT flanked (blue arrow-heads) neomycin resistance cassette and the STOP cassette is inserted directly downstream of *Gm14204* exon 1 (red marked arrow at E1). The STOP cassette is shown as polyA site (red arrow) and a series of TTS sites (rectangle blue boxes). The homology arms are shown as LA (long arm) and SA (short arm). The loxP sites are shown as black arrow-heads. Restriction enzyme sites are shown as required for confirmation. The vector will be linearized with *XhoI* before electroporation.

4.14.2 Transfection of embryonic stem cells:

For modification of the *Gm14204* gene locus, 20 µg of the targeting vector G051.5A was linearized using *XhoI* and electroporated into 1×10^7 C57Bl/6-derived ES cells. G418 (0.2 mg/mL) was used to select for stable transfection. After 8 days of selection, a total of 4 times 96 clones were picked and analyzed *via* PCR. G418 was used for screening and 4x 96 clones were selected and isolated. Genomic DNA was isolated and lysates were tested *via*. screening PCR analysis. Long range PCR was designed and performed to screen and confirm that correct homologous recombination occurred at *Gm14204* locus. PCR analysis yielded 16 positive clones that were selected for further analysis (data not shown).

4.14.3 Southern blot analysis:

To confirm the correct homologous recombination at the *Gm14204* locus, Southern blot analysis was performed using *BclI* digested DNA on the selected 16 clones.
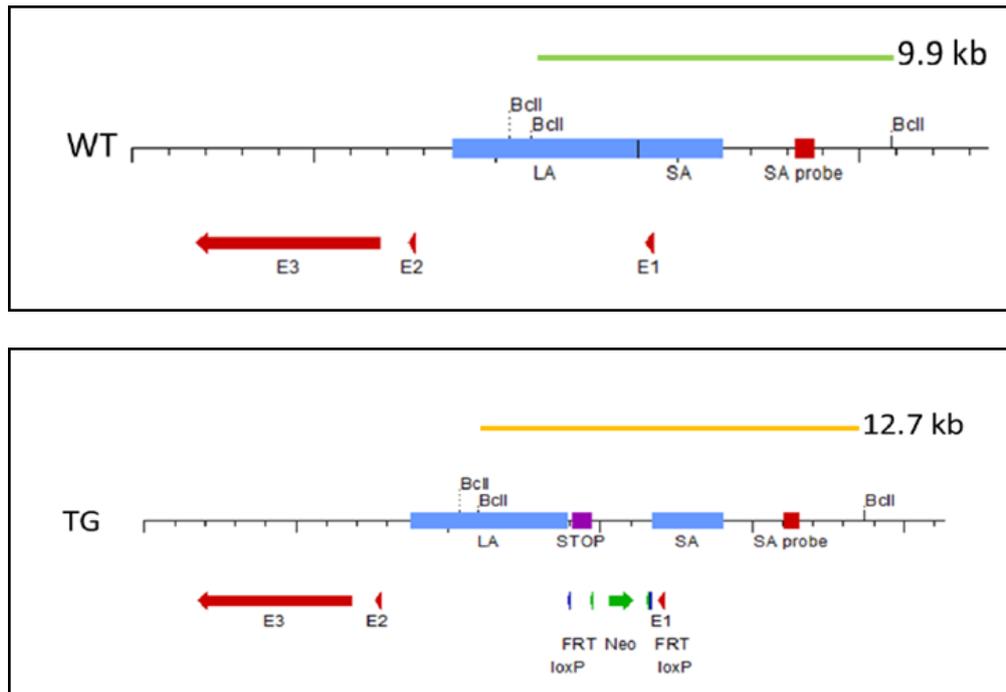
Figure 37: Schematic diagram of wild-type (WT) and targeted alleles (TG). The targeted allele (TG) contains an FRT-flanked neomycin cassette (green arrow, below the line) and the desired stop cassette (magenta box). Exons (E1, E2, and E3) are marked as red arrows. The restriction enzyme *BclI* and an upstream probe (SA probe, marked as a red box) were used for Southern blot. The expected DNA fragments and their sizes are shown as green (WT) and orange (TG) bars. The homologous regions (*SA:* short arm of homology; *LA:* long arm of homology) used for recombination are depicted as blue boxes.

Hybridization resulted in a 9.9 kb signal for the wild-type allele and 12.7 kb signal for the correctly targeted allele. Southern blot analysis confirmed correct homologous

recombination for 12 clones, which were then used for blastocyst injection. The positive clones are shown in red (**Fig. 38**).
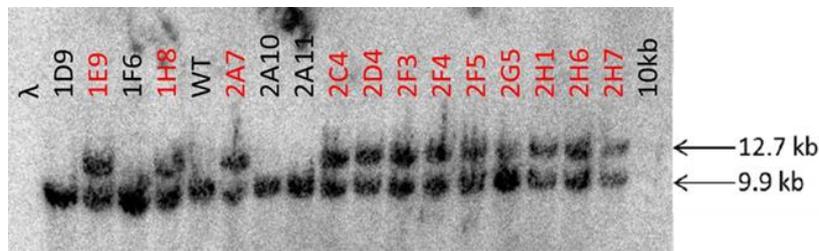
Figure 38: Confirmation of candidate ES clones with Southern blot analysis. DNA from 16 potentially positive ES cell clones was digested using the restriction enzyme *BclI*. A 503 bp upstream external probe was used for detection of the modified *Gm14204* allele. The analysis confirmed correct homologous recombination within 12 of 16 clones tested, marked in red. This is indicated by a signal of 12.7 kb representing the correctly targeted allele. The wild-type signal has a size of 9.9 kb. Wild-type DNA served as control (WT), the positions of λ and 10 kb markers used during the electrophoresis are also marked.

4.14.5 Generating chimeric mice:

Three ES clones (2D4, 2F5 and 2H7) were injected into 51 blastocysts from grey C57Bl/6 mice. Out of these 48 surviving blastocysts were transferred to two CD-1 foster mice. Four out of five pups born from clone 2D4 showed chimerism (80%, 50%, and 2x 30%) based on fur color estimation. From clone 2F5, three pups were chimeric (80%, 50%, and 30%). All the chimeric mice were males.

4.14.6 Screening strategy of F1 mice:

Chimeric mice were set up with heterozygous grey Flp 'deleter' mice. Germline (B6N) offspring were screened using PCR and confirmed by sequencing (data not shown). Primers used are shown in **Fig. 39** and the sequences are mentioned in Table 3. Two major procedures were used for screening: a) three primer multiplex approach since it has a low sensitivity to

neo contaminations. Primer Neo.MP1 binds to a region of neomycin cassette and to a region on mouse chromosome 3. Neo.MP6 binds to neomycin cassette, resulting in a fragment of 512 bp with Neo.MP1. Neo.MP5 binds to the mouse genome region on chromosome 3 and yields a PCR product of 380 bp (control fragment), b) used primers G051.1 and A300.3 to screen for deletion of neomycin cassette ('Del-Neo' genotype). This screening yielded a 370 bp amplicon if the transgene is present and the Neo cassette was cut out. The screening was also performed for the Flp gene that comes from Flp deleter strain using SD24 and SD25 (Flp genotype).
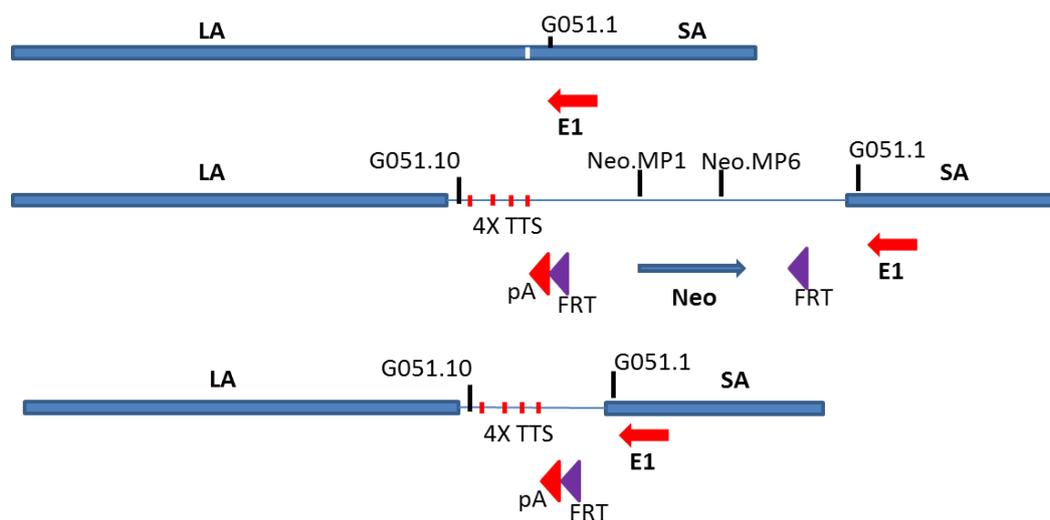
Figure 39: Possible alleles of the targeted locus. Wild-type (WT), targeted allele (Neo) before Flp recombination and after Flp recombination (Del-Neo); Arms of homology (LA and SA) are shown as long blue rectangles. Exon 1 (E1) is shown as a red arrow. FRT sites are shown as purple arrowheads. 4xTTS are shown in small red vertical lines. PolyA (pA) site is shown in red arrowhead. Primers at respective positions are shown as G051.1, G051.10, Neo.MP1 and Neo.MP6.

Primer ID	Sequence	Purpose
Neo.MP1	GCTGTGCTCCACGTTGTCAC	
Neo.MP5	GGAAAGCTGGGCTTGCATCTC	
Neo.MP6	GGAGCGGCGATACCGTAAAG	Screening for Neo genotype
SD24	CTAATGTTGTGGGAAATTGGAGC	
SD25	CTCGAGGATAACTTGTTTATTGC	Screening for Flp genotype
G051.1	GTGTCTCTGCCATCTATCAC	Screening for Del-Neo genotype
A300.3	CAATGTATCTTAATCAACGGTTACGAAGTTCC	

Table 3: List of primers used for genotype screening. These set of primers were used by Polygene Transgenetics for their screening purposes. The primers used by us for are mentioned below in the Genotyping section.

4.14.7 Genotyping:

Tissue digestion:

Ear biopsies were collected and stored at -20°C until use for genotyping. 100 µL of PBND buffer and 1.5 µL of Proteinase K (20 mg/mL; Thermo Scientific) solutions were added to the tissues and incubated at 55°C for 3 hours with 1000 rpm shaking. Next, the temperature was increased to 85°C and samples were incubated again for 1 hour without shaking to inactivate the enzyme. The samples were then centrifuged at maximum speed for 10 min and the supernatant was stored at -20°C until use.

4.14.8 PCR:

Genomic DNA PCR was performed to determine the genotype of each animal. 1 µL of tissue biopsy lysate was taken for a 50 µL PCR reaction volume (Table 4) and the PCR was performed as the protocol given in Table 5. The primers used for genotyping are mentioned in Table 6. Their respective positions are shown in **Fig 40**.

Buffers	Volume
10x PCR buffer (cont. 15 mM MgCl$_2$; Qiagen)	5 µL
5x Q solution (Qiagen)	10 µL
2 mM dNTPs (Roche)	5 µL
250 U *Taq* DNA polymerase (Qiagen)	0.5 µL
WT_F primer (5 pmol)	6 µL
WT_R primer (5 pmol)	3 µL
Het_R2 primer (5 pmol)	6 µL
Biopsy DNA	1 µL
Dist. H$_2$O	13.5 µL
Total	50 µL

Table 4: PCR reactions and program used for genotyping

Temp.	Time	Cycles
94ºC	Pause	
94ºC	3 min	
94ºC	30 s	
59ºC	30 s	35x
72ºC	45 s	
72ºC	10 min	
4ºC	Pause	

Table 5: PCR reactions and program used for genotyping

Possibilities	Primers	Product(s)
WT alleles	For: 5'-GCGTACCCATGTTTCTGA-3'	
	Rev: 5'-GGTTGGACCATCTGAGC-3'	354 bp
Het. alleles	For: 5'-GCGTACCCATGTTTCTGA-3'	
	Rev: 5'-CCCTGATAGACGGTTTTT-3'	354 bp + 574 bp
Mut. alleles	For: 5'-GCGTACCCATGTTTCTGA-3'	
	Rev: 5'-CCCTGATAGACGGTTTTT-3'	574 bp

Table 6: Primers used for genotyping
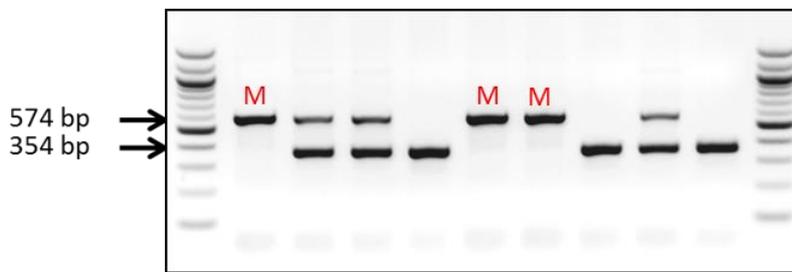
Figure 40: Genotyping results for *Gm14204* WT +/+, Het +/- and Mut (*pATTS/pATTS*) alleles. The WT allele shows a band of 354 bp, the mutant (M) alleles show a band of 574 bp, whereas the heterozygous alleles show both 354 bp and 574 bp bands on 2% Agarose gel.

4.15 RNAscope Multiplex Fluorescent Assay for single molecule FISH (smFISH):

The RNAscope Multiplex Fluorescent Assay uses a novel method for RNA *in situ* hybridization to visualize the specific location of RNA transcripts inside a cell. This assay is based on signal amplification reaction and low background strategy that enables is to investigate gene expression in a cellular context. The assay first permeabilizes the tissue/cells using protease enzyme followed by hybridizing with ~20 double 'ZZ' target-specific probes. A signal is generated only if the two 'ZZ' probes bind in close proximity to each other. Single 'Z' probes bound non-specifically will be removed in the washing steps. In the next step, the signals are amplifies using RNAscope detection reagents and amplifiers, which are detected as single bright spots under a microscope (scheme shown in **Fig. 41**). We designed a *Gm14204* lncRNA specific probe targeting the longer isoform (NR_040358.1) using ~15 'ZZ' specific probes covering the entire range between 309-1521 nucleotides of lncRNA transcript.
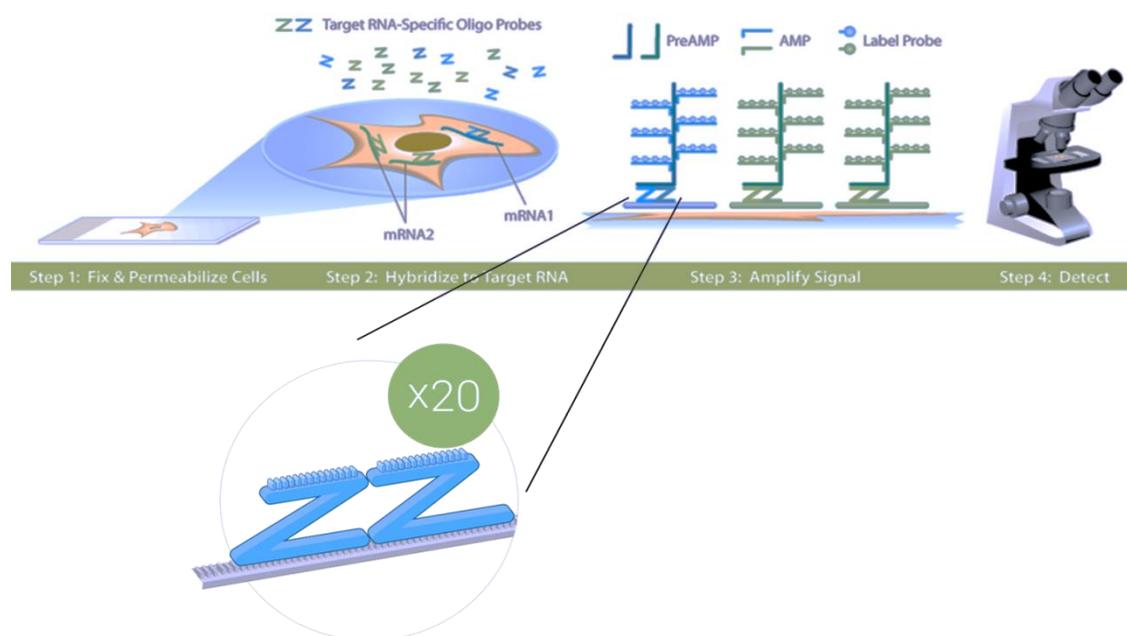
Figure 41: **Principle of Single molecule FISH** as adapted from ACDBio RNAscope technology.

Wash Buffer (1X):

- Add 1 bottle (60 mL) of 50X wash buffer to 2.94 L of dist. H_2O to prepare 1X working wash buffer solution.

- If precipitation occurs in stock solution, warm it up to 40°C for 10-20 min before dilution. Working solution can be stored at room temperature for one month.

Probe preparation:

- Take out the probes from 4°C, warm at 40°C for 10 min and let them cool at room temperature.

- Spin C2 and C3 probes to collect them at the bottom of the tubes.

- Mix 1:1:50 ratios of C2:C3:C1 probes by pipetting 1 volume of C2, 1 volume of C3 and 50 volume of C1 probe in an Eppendorf. Mix the probes several times.

- The mixed probes can be stored at 4°C for up to a month.

Equilibrate reagents:

- Place Amp 1-4 FL reagents at RT.

- HybEZ oven and humidity control tray should be at 40°C.

Run the Assay:

IMP: Do not let the sections dry out between steps. Work quickly.

Hybridize the probe:

- Tap or flick to remove excess liquid from the slides and place them in the HybEZ slide rack.

- Add 4 drops (or enough drops) of the required probe to cover the entire section.

- Keep the slide rack in the HybEZ humidity control tray and place in the HybEZ oven at 40°C for 2 hours.

- After incubation, remove the tray and one slide at a time, quickly remove excess liquid by decanting and place the slide in any slide rack.

- Wash the slides in 1X wash buffer by moving slide rack up and down occasionally for 1 min at RT.

- Repeat above step with fresh 1X wash buffer.

Hybridize Amp 1-FL:

- One at a time, take each slide from the slide rack and tap/flick to remove excess liquid from the slide. Place the slides back in the HybEZ slide rack.

- Add 4 drops or enough to cover the section of Amp 1-FL.

- Place the slide rack in the humidity control tray removed from the HybEZ oven. Seal tray and insert back into the oven for 30 min at 40°C.

- Remove the HybEZ control tray and HybEZ slide rack from the oven.

- One at a time, quickly remove excess liquid and place the slides in 1X wash buffer.

- Wash slides for 1 min at RT with occasional agitation.

- Repeat above step.

Hybridize Amp 2-FL:

- One at a time, take each slide from the slide rack and tap/flick to remove excess liquid from the slide. Place the slides back in the HybEZ slide rack.

- Add 4 drops or enough to cover the section of Amp 2-FL.

- Place the slide rack in the humidity control tray removed from the HybEZ oven. Seal tray and insert back into the oven for 15 min at 40°C.

- Remove the HybEZ control tray and HybEZ slide rack from the oven.

- One at a time, quickly remove excess liquid and place the slides in 1X wash buffer.

- Wash slides for 1 min at RT with occasional agitation.

- Repeat above step.

Hybridize Amp 3-FL:

- One at a time, take each slide from the slide rack and tap/flick to remove excess liquid from the slide. Place the slides back in the HybEZ slide rack.

- Add 4 drops or enough to cover the section of Amp 3-FL.

- Place the slide rack in the humidity control tray removed from the HybEZ oven. Seal tray and insert back into the oven for 30 min at 40°C.

- Remove the HybEZ control tray and HybEZ slide rack from the oven.

- One at a time, quickly remove excess liquid and place the slides in 1X wash buffer.

- Wash slides for 1 min at RT with occasional agitation.

- Repeat above step.

Hybridize Amp 4-FL:

- One at a time, take each slide from the slide rack and tap/flick to remove excess liquid from the slide. Place the slides back in the HybEZ slide rack.

- Add 4 drops or enough to cover the section of Amp 4-FL.

Note: At this point, there are 3 options for alternate fluorescent color modules. Any fluorescent color combination (Amp 4-FL, A, B or C) can be selected.

- Place the slide rack in the humidity control tray removed from the HybEZ oven. Seal tray and insert back into the oven for 15 min at 40°C.

- Remove the HybEZ control tray and HybEZ slide rack from the oven.

DAPI staining:

- Add 2-3 drops of VectaShield Mounting medium containing DAPI onto the slide and place a coverslip on top.

- Use the slides for imaging or store them at 4°C.

4.16 Thionine Acetate (Nissl) Staining:

Remove the slides from freezer and thaw them to room temperature in 1x PBS solution. Immerse the slides in different solutions as follows:

1. Xylol   15 min
2. 100% Ethanol   1 min
3. 100% Ethanol   1 min
4. 95% Ethanol   30 sec
5. 70% Ethanol   1 min
6. 50% Ethanol   30 sec
7. 30% Ethanol   30 sec
8. Distilled water   30 sec
9. 0.25% (w/v) Thionine acetate   (2-5) min
10. Distilled water   30 sec
11. 70% Isopropanol+0.1% Acetic acid   30 sec
12. 70% Isopropanol   30 sec
13. 100% Isopropanol   30 sec
14. 100% Isopropanol   30 sec
15. Xylol   15 min

4.17 List of Primers

(i)   For RNA probe synthesis (RNA-ISH):

Gene name	Forward primer 5'-> 3'	Reverse primer 5'-> 3'
Gm14204_1	AGGCAGAGCAAACAAAGAGC	GCCTTTACCATGTGCCAAGT
Gm14204_2	CAGGAGGTGAGGAAGCAGAG	CAGGCAACTAAACCCCAGAA
Gm14204_3	CCCATCTGGTGTGTTGAGTG	GTTGTGTGTTGCCTGTGGTC
Gm14204_4	CCATGTAGGTGCTGGGAACT	GGACCCGTGAAAATTAAGCA
Gm14204_5	GACCACAGGCAACACACAAC	GGCCAGTATCAGTGGGAAGA
CUFF.557225	GTGTGGCAGAGGCTGAAAAT	GACCAGTCAGGCCTTGGATA
9130024F11Rik	AGCCGCAGAGAGAAAGACAC	ATCCGGGTCAAAACTACAGG
Fendrr	CAGCACCCACTTGAGAGTCA	CCCTAGGGGAGAAACTACCG
Gm11266	CAGGACGGATGTTTCCAAGT	TCTGCCTTTCTTCTGGTTGG

(ii)   For qRT-PCR:

Gene name	Forward primer 5'-> 3'	Reverse primer 5'-> 3'
Gm14204_1	GCCATAGACTGGCTCTCAGG	TCAACACACCAGATGGGAAA
Gm14204_2	CATAGGCGTGTGTGTTCCAC	GGATCCACGTAGCAGAAGGA
Gm14204_3	AAGAAACGGAGGGGGACTAA	TAATGCAACGTGATGGATGG
Gm14204_4	GGATCCAAGATGCAAGCCTA	CCCCTCCTGGAATTCTGAAA
Actr5	GCTGCTGGACTACAGCTTCC	GACGGGAGTAGAGGGGGTTA
Dhx35	CCCTACGCTGCCCTTTCTAT	CAACAATCACCACGGTCTGA
Fam83d	CATCGACATCTTCCGAGACC	AAAAAGTGGGGGAGCAGAGT
Ralgapb	AATGGGTGGTGGAGAAAACA	TCGCTATCAGGGGTTGTAGG
Gad67	CAGAAAACTGGGCCTGAAGA	CAGGAAAGCAGGTTCTTGGA
Gad65	CCAAATCAGCCCTAGCAAAG	ATCTGGGAGGAAGCCATTCT
Lhx6	GGACAAGGACGAAGGTAGAGC	CCGCAACTGGAGCAGATATT
Dlx2	GTGCAGATTTCCCCACATTT	TACGTCGCAGCTTTCACAAC
Dlx1	ACTTTGGGCAGCAAGAGAGA	TTAAAGCGACCTGGAATTGG
Dlx5	GGCCGCTTTACAGAGAAGGT	TCACCTGTGTTTGCGTCAGT

Gene name	Forward primer 5'-> 3'	Reverse primer 5'-> 3'
Dlx6	AAAACGGGGAAATCAGGTTC	AGTCTGCTGAAAGCGATGGT
Htr3a	ACTTCCCCTTTGATGTGCAG	CTCACTTCTTCCGGTGATCG
Sst	TGAAGGAGACGCTACCGAAG	GAGACGGCAGGACAGCAT
Evf2	GCCACAGTTTCCCCTCTACA	TAAGAGGTCAGCGCTCCAAT
Slc32a1	AACGTGACAAATGCCATTCA	TGAGGAACAACCCCAGGTAG
Gapdh	CATGGCCTTCCGTGTTCCTA	CCTGCTTCACCACCTTCTTGA
Rpl	AACGGACTCCTGGTGTGAAC	ACAGGAGCAGTGCCTAAGGA

List of Tables

List of Figures

Abbreviations

aa	amino acids
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
bp	base pair
CA	*cornu ammonis*
CARD	Catalyzed Reporter Deposition
CNS	Central Nervous System
CPAT	Coding Potential Assessment Tool
CP	Cortical Plate
DAPI	4',6-diamidino-2-phenylindole
DIG	Digoxigenin
DG	Dentate Gyrus
DNA	Deoxyribonucleic Acid
E14.5	Embryonic day 14.5
EDTA	Ethylenediamietetraacetic acid
ENCODE	Encyclopedia of DNA elements
ESC	Embryonic Stem Cells
EtOH	Ethanol
Fc	Fold Change
GCL	Granule Cell Layer
FANTOM	Functional Annotation of Mouse
FPKM	Fragment Per Kilobase of transcript per Million Mapped Reads
FRT	FLP Recognition Target
ISH	*In situ* hybridization
IZ	Intermediate Zone
Kan	Kanamycin
LGE	Lateral Ganglionic Eminence
lncRNA	Long noncoding RNA
lincRNA	Long Intergenic noncoding RNA
LoxP	Locus of X-over P
mRNA	Messenger RNA
MCL	Molecular Cell Layer

MGE	Medial Ganglionic Eminence
Myr	Million Years
Neo	Neomycin
NBT	Nitro Blue Tetrazolium
NR	Neural Retina
OB	Olfactory Bulb
OCT	Optimal Cutting Temperature
ORF	Open Reading Frame
P1/2/7/56	Postnatal day 1/2/7/56
PBND	PCR Buffer with Nonionic Detergents
PBS	Phosphate Buffered Saline
PCL	Purkinje Cell Layer
PFA	Paraformaldehyde
Pol-II	Polymerase II
PRC2	Polycomb Repressive Complex 2
PTh	Pre-Thalamus
qRT-PCR	Quantitative Reverse Transcriptase-Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
RMS	Rostral Migratory Stream
RPKM	Read Per Kilobase of transcript per Million Mapped Read
TAE	Tris-Acetate-EDTA
TH	Thalamus
UCSC	University of California Santa Cruz
UTR	Untranslated Region
VZ	Ventricular Zone

References

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., … Olson, E. N. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*. https://doi.org/10.1016/j.cell.2015.01.009

Anderson, S. A., Eisenstat, D. D., Shi, L., & Rubenstein, J. L. R. (1997). Interneuron migration from basal forebrain to neocortex: Dependence on Dlx genes. *Science*. https://doi.org/10.1126/science.278.5337.474

Bartolomei, M. S., Webber, A. L., Brunkow, M. E., & Tilghman, S. M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes and Development*. https://doi.org/10.1101/gad.7.9.1663

Bassett, A. R., Akhtar, A., Barlow, D. P., Bird, A. P., Brockdorff, N., Duboule, D., … Ponting, C. P. (2014). Considerations when investigating lncRNA function in vivo. *ELife*. https://doi.org/10.7554/eLife.03058

Bellucci, M., Agostini, F., Masin, M., & Tartaglia, G. G. (2011). Predicting protein associations with long noncoding RNAs. *Nature Methods*. https://doi.org/10.1038/nmeth.1611

Berghoff, E. G., Clark, M. F., Chen, S., Cajigas, I., Leib, D. E., & Kohtz, J. D. (2013). Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development*. https://doi.org/10.1242/dev.099390

Bernstein, E., & Allis, C. D. (2005). RNA meets chromatin. *Genes and Development*. https://doi.org/10.1101/gad.1324305

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., … De Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. https://doi.org/10.1038/nature05874

Bogu, G. K., Vizán, P., Stanton, L. W., Beato, M., Di Croce, L., & Marti-Renom, M. A. (2016). Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Molecular and Cellular Biology*. https://doi.org/10.1128/MCB.00955-15

Bonasio, R., & Shiekhattar, R. (2014). Regulation of Transcription by Long Noncoding RNAs. *Annual Review of Genetics*. https://doi.org/10.1146/annurev-genet-120213-092323

Bond, A. M., Vangompel, M. J. W., Sametsky, E. A., Clark, M. F., Savage, J. C., Disterhoft, J. F., & Kohtz, J. D. (2009). Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nature Neuroscience*. https://doi.org/10.1038/nn.2371

Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L., & Barry, G. (2015). Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. *Neuron*. https://doi.org/10.1016/j.neuron.2015.09.045

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. *RNA*. https://doi.org/10.1261/rna.037598.112

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., … Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*. https://doi.org/10.1016/0092-8674(92)90519-I

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., & Willard, H. F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*. https://doi.org/10.1038/349038a0

Bulfone, A., Puelles, L., Porteus, M. H., Frohman, M. A., Martin, G. R., & Rubenstein, J. L. R. (1993). Spatially restricted expression of Dlx-1, Dlx-2 (Tes-1), Gbx-2, and Wnt-3 in the embryonic day 12.5 mouse forebrain defines potential transverse and longitudinal segmental boundaries. *The Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.2750-07.2007

Cabili, M. N., Dunagin, M. C., McClanahan, P. D., Biaesch, A., Padovan-Merhar, O., Regev, A., … Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biology*. https://doi.org/10.1186/s13059-015-0586-4

Cabili, M., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes and Development*. https://doi.org/10.1101/gad.17446611

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., … Nishiguchi, S. (2005). The transcriptional landscape of the mammalian genome. *Science*. https://doi.org/10.1126/science.1112014

Castelo-Branco, G., Amaral, P. P., Engström, P. G., Robson, S. C., Marques, S. C., Bertone, P., & Kouzarides, T. (2013). The non-coding snRNA 7SK controls transcriptional termination, poising, and bidirectionality in embryonic stem cells. *Genome Biology*. https://doi.org/10.1186/gb-2013-14-9-r98

Cech, T. R., & Steitz, J. A. (2014). The noncoding RNA revolution - Trashing old rules to forge new ones. *Cell*. https://doi.org/10.1016/j.cell.2014.03.008

Chen, L. L. (2016). Linking Long Noncoding RNA Localization and Function. *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2016.07.003

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., & Chang, H. Y. (2011). Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2011.08.027

Chu, C., Zhang, Q. C., Da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., … Chang, H. Y. (2015). Systematic discovery of Xist RNA binding proteins. *Cell*. https://doi.org/10.1016/j.cell.2015.03.025

Clemson, C. M., McNeil, J. A., Willard, H. F., & Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: Evidence for a novel RNA involved in nuclear/chromosome structure. *Journal of Cell Biology*. https://doi.org/10.1083/jcb.132.3.259

Crick, F. (1970). Central dogma of molecular biology. *Nature*. https://doi.org/10.1038/227561a0

Davidovich, C., & Cech, T. R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: Lessons from PRC2. *RNA*. https://doi.org/10.1261/rna.053918.115

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., … Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*. https://doi.org/10.1101/gr.132159.111

Dinger, M. E., Pang, K. C., Mercer, T. R., & Mattick, J. S. (2008). Differentiating protein-coding and noncoding RNA: Challenges and ambiguities. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1000176

Eichele, G., & Diez-Roux, G. (2011). High-throughput analysis of gene expression on tissue sections by in situ hybridization. *Methods*. https://doi.org/10.1016/j.ymeth.2010.12.020

Eißmann, M., Gutschner, T., Hämmerle, M., Günther, S., Caudron-Herger, M., Groß, M., … Diederichs, S. (2012). Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biology*. https://doi.org/10.4161/rna.21089

ENCODE, Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., & Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*. https://doi.org/nature11247 [pii]\n10.1038/nature11247

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., … Guttman, M. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. https://doi.org/10.1126/science.1237973

Fukumoto, K., Tamada, K., Toya, T., Nishino, T., Yanagawa, Y., & Takumi, T. (2018). Identification of genes regulating GABAergic interneuron maturation. *Neuroscience Research*. https://doi.org/10.1016/j.neures.2017.11.010

Grote, P., & Herrmann, B. G. (2015). Long noncoding RNAs in organogenesis: Making the difference. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2015.02.002

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., … Herrmann, B. G. (2013). The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Developmental Cell*. https://doi.org/10.1016/j.devcel.2012.12.012

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., … Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. https://doi.org/10.1038/nature07672

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., & Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*. https://doi.org/10.1016/j.cell.2013.06.009

Hacisuleyman, E., Goff, L. A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., … Rinn, J. L. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.2764

Heavner, W., & Pevny, L. (2012). Eye development and retinogenesis. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a008391

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Reports*. https://doi.org/10.1016/j.celrep.2015.04.023

Hoki, Y., Kimura, N., Kanbayashi, M., Amakawa, Y., Ohhata, T., Sasaki, H., & Sado, T. (2009). A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development*. https://doi.org/10.1242/dev.026427

Hung, T., & Chang, H. Y. (2010). Long noncoding RNA in genome regulation. *RNA Biology*. https://doi.org/10.4161/rna.7.5.13216

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., … Chinnaiyan, A. M. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics*. https://doi.org/10.1038/ng.3192

Jiang, Y., Wang, X., Bao, S., Guo, R., Johnson, D. G., Shen, X., & Li, L. (2010). INO80 chromatin remodeling complex promotes the removal of UV lesions by the nucleotide excision repair pathway. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1008388107

Jurek, B., & Neumann, I. D. (2018). The Oxytocin Receptor: From Intracellular Signaling to Behavior. *Physiological Reviews*. https://doi.org/10.1152/physrev.00031.2017

Kakizaki, T., Oriuchi, N., & Yanagawa, Y. (2015). GAD65/GAD67 double knockout mice exhibit intermediate severity in both cleft palate and omphalocele compared with GAD67 knockout and VGAT knockout mice. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2014.12.030

Kapranov, P., Cawley, S. E., Drenkow, J., Bekiranov, S., Strausberg, R. L., Fodor, S. P. A., & Gingeras, T. R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science*. https://doi.org/10.1126/science.1068597

Kelsom, C., & Lu, W. (2013). Development and specification of GABAergic cortical interneurons. *Cell and Bioscience*. https://doi.org/10.1186/2045-3701-3-19

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., … Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0904715106

Khraiwesh, B., Arif, M. A., Seumel, G. I., Ossowski, S., Weigel, D., Reski, R., & Frank, W. (2010). Transcriptional Control of Gene Expression by MicroRNAs. *Cell*. https://doi.org/10.1016/j.cell.2009.12.023

Kitayama, K., Kamo, M., Oma, Y., Matsuda, R., Uchida, T., Ikura, T., … Harata, M. (2009). The human actin-related protein hArp5: Nucleo-cytoplasmic shuttling and involvement in DNA repair. *Experimental Cell Research*. https://doi.org/10.1016/j.yexcr.2008.10.028

Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., Fields, P. A., Steinhauser, M. L., … Boyer, L. A. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*. https://doi.org/10.1016/j.cell.2013.01.003

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., & Gao, G. (2007). CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkm391

Kornienko, A. E., Guenzl, P. M., Barlow, D. P., & Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biology*. https://doi.org/10.1186/1741-7007-11-59

Kung, J. T. Y., Colognori, D., & Lee, J. T. (2013). Long noncoding RNAs: Past, present, and future. *Genetics*. https://doi.org/10.1534/genetics.112.146704

Latos, P. A., Pauler, F. M., Koerner, M. V., Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., … Barlow, D. P. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*. https://doi.org/10.1126/science.1228110

Le, T. N., Zhou, Q.-P., Cobos, I., Zhang, S., Zagozewski, J., Japoni, S., … Eisenstat, D. D. (2017). GABAergic Interneuron Differentiation in the Basal Forebrain Is Mediated through

Direct Regulation of Glutamic Acid Decarboxylase Isoforms by *Dlx* Homeobox Transcription Factors. *The Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.2125-16.2017

Lee, J. T. (2009). Lessons from X-chromosome inactivation: Long ncRNA as guides and tethers to the epigenome. *Genes and Development*. https://doi.org/10.1101/gad.1811209

Lee, S., Kopp, F., Chang, T. C., Sataluri, A., Chen, B., Sivakumar, S., … Mendell, J. T. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell*. https://doi.org/10.1016/j.cell.2015.12.017

Li, L., & Chang, H. Y. (2014). Physiological roles of long noncoding RNAs: Insight from knockout mice. *Trends in Cell Biology*. https://doi.org/10.1016/j.tcb.2014.06.003

Liu, B., Ye, B., Yang, L., Zhu, X., Huang, G., Zhu, P., … Fan, Z. (2017). Long noncoding RNA lncKdm2b is required for ILC3 maintenance by initiation of Zfp292 expression. *Nature Immunology*. https://doi.org/10.1038/ni.3712

Ludwig, T., Eggenschwiler, J., Fisher, P., D'Ercole, A. J., Davenport, M. L., & Efstratiadis, A. (1996). Mouse mutants lacking the type 2 IGF receptor (IGF2R) are rescued from perinatal lethality in Igf2 and Igf1r null backgrounds. *Developmental Biology*. https://doi.org/10.1006/dbio.1996.0182

Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biology*. https://doi.org/10.4161/rna.24604

Marchese, F. P., Raimondi, I., & Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biology*. https://doi.org/10.1186/s13059-017-1348-2

Martens, J. A., Laprade, L., & Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature*. https://doi.org/10.1038/nature02538

Mattick, J. S. (2011). The central role of RNA in human development and cognition. *FEBS Letters*. https://doi.org/10.1016/j.febslet.2011.05.001

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., … Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*. https://doi.org/10.1038/nature14443

McIntire, S. L., Reimer, R. J., Schuske, K., Edwards, R. H., & Jorgensen, E. M. (1997). Identification and characterization of the vesicular GABA transporter. *Nature*. https://doi.org/10.1038/39908

Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F., & Mattick, J. S. (2008). Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0706729105

Mercer, T. R., Dinger, M. E., Mariani, J., Kosik, K. S., Mehler, M. F., & Mattick, J. S. (2008). Noncoding RNAs in long-term memory formation. *Neuroscientist*. https://doi.org/10.1177/1073858408319187

Mercer, T. R., & Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.2480

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., … Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. https://doi.org/10.1038/nature06008

Minajigi, A., Froberg, J. E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., … Lee, J. T. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science*. https://doi.org/10.1126/science.aab2276

Mohamed, J. S., Gaughwin, P. M., Lim, B., Robson, P., & Lipovich, L. (2010). Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA*. https://doi.org/10.1261/rna.1441510

Mohammad, F., Mondal, T., Guseva, N., Pandey, G. K., & Kanduri, C. (2010). Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development*. https://doi.org/10.1242/dev.048181

Molyneaux, B. J., Arlotta, P., Menezes, J. R. L., & Macklis, J. D. (2007). Neuronal subtype specification in the cerebral cortex. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn2151

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., … Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*. https://doi.org/10.1038/nature12943

Nelson, B. R., Makarewich, C. A., Anderson, D. M., Winders, B. R., Troupes, C. D., Wu, F., … Olson, E. N. (2016). Muscle physiology: A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*. https://doi.org/10.1126/science.aad4076

Nielsen, M. M., Tehler, D., Vang, S., Sudzina, F., Hedegaard, J., Nordentoft, I., … Pedersen, J. S. (2014). Identification of expressed and conserved human noncoding RNAs. *RNA*. https://doi.org/10.1261/rna.038927.113

Oliver, P. L., Chodroff, R. A., Gosal, A., Edwards, B., Cheung, A. F. P., Gomez-Rodriguez, J., … Ponting, C. P. (2015). Disruption of Visc-2, a brain-expressed conserved long noncoding RNA, does not elicit an overt anatomical or behavioral phenotype. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhu196

Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., … Shiekhattar, R. (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell*. https://doi.org/10.1016/j.cell.2010.09.001

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., … Kanduri, C. (2008). Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2008.08.022

Panganiban, G., & Rubenstein, J. L. R. (2002). Developmental functions of the Distal-less/Dlx homeobox genes. *Development (Cambridge, England)*. https://doi.org/10.1016/s0361-9230(01)00770-5

Personnic, N., Lakisic, G., Gouin, E., Rousseau, A., Gautreau, A., Cossart, P., & Bierne, H. (2014). A role for Ral GTPase-activating protein subunit β in mitotic regulation. *FEBS Journal*. https://doi.org/10.1111/febs.12836

Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M. A., Coppens, S., Pedersen, J. S., … Haussler, D. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. https://doi.org/10.1038/nature05113

Ponjavic, J., Oliver, P. L., Lunter, G., & Ponting, C. P. (2009). Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1000617

Ponjavic, J., Ponting, C. P., & Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Research*. https://doi.org/10.1101/gr.6036807

Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and Functions of Long Noncoding RNAs. *Cell*. https://doi.org/10.1016/j.cell.2009.02.006

Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D., & Grummt, I. (2015). LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2015.10.001

Puelles, L., & Rubenstein, J. L. R. (2003). Forebrain gene expression domains and the evolving prosomeric model. *Trends in Neurosciences*. https://doi.org/10.1016/S0166-2236(03)00234-0

Rashid, F., Shah, A., & Shan, G. (2016). Long Non-coding RNAs in the Cytoplasm. *Genomics, Proteomics and Bioinformatics*. https://doi.org/10.1016/j.gpb.2016.03.005

Rinn, J., & Guttman, M. (2014). RNA and dynamic nuclear organization. *Science*. https://doi.org/10.1126/science.1252966

Rinn, J. L., & Chang, H. Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev-biochem-051410-092902

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., … Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell*. https://doi.org/10.1016/j.cell.2007.05.022

Roberts, T. C., Morris, K. V., & Wood, M. J. A. (2014). The role of long non-coding RNAs in neurodevelopment, brain function and neurological disease. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1055/s-0030-1270848

Sauvageau, M., Goff, L. A., Lodato, S., Bonev, B., Groff, A. F., Gerhardinger, C., … Rinn, J. L. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *ELife*. https://doi.org/10.7554/eLife.01749

Schmitz, K. M., Mayer, C., Postepska, A., & Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes and Development*. https://doi.org/10.1101/gad.590910

Schmitz, S. U., Grote, P., & Herrmann, B. G. (2016). Mechanisms of long noncoding RNA function in development and disease. *Cellular and Molecular Life Sciences*. https://doi.org/10.1007/s00018-016-2174-5

Schwalb, B., Michel, M., Zacher, B., Hauf, K. F., Demel, C., Tresch, A., … Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science*. https://doi.org/10.1126/science.aad9841

Sigova, A. A., Mullen, A. C., Molinie, B., Gupta, S., Orlando, D. A., Guenther, M. G., … Young, R. A. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1221904110

Sleutels, F., Zwart, R., & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*. https://doi.org/10.1038/415810a

Sone, M., Hayashi, T., Tarui, H., Agata, K., Takeichi, M., & Nakagawa, S. (2007). The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *Journal of Cell Science*. https://doi.org/10.1242/jcs.009357

Souquere, S., Beauclair, G., Harper, F., Fox, A., & Pierron, G. (2010). Highly Ordered Spatial Organization of the Structural Long Noncoding NEAT1 RNAs within Paraspeckle Nuclear Bodies. *Molecular Biology of the Cell*. https://doi.org/10.1091/mbc.E10-08-0690

St.Laurent, G., Wahlestedt, C., & Kapranov, P. (2015). The Landscape of long noncoding RNA classification. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2015.03.007

Tian, D., Sun, S., & Lee, J. T. (2010). The long noncoding RNA, Jpx, Is a molecular switch for X chromosome inactivation. *Cell*. https://doi.org/10.1016/j.cell.2010.09.049

Tole, S., Christian, C., & Grove, E. A. (1997). Early specification and autonomous development of cortical fields in the mouse hippocampus. *Development*.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., … Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. https://doi.org/10.1038/nbt.1621

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., … Prasanth, K. V. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2010.08.011

Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., … Chang, H. Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. https://doi.org/10.1126/science.1192002

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., & Bartel, D. P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*. https://doi.org/10.1016/j.cell.2011.11.055

Urbán, N., & Guillemot, F. (2014). Neurogenesis in the embryonic and adult brain: same regulators, different roles. *Frontiers in Cellular Neuroscience*. https://doi.org/10.3389/fncel.2014.00396

Van Heesch, S., Van Iterson, M., Jacobi, J., Boymans, S., Essers, P. B., De Bruijn, E., … Simonis, M. (2014). Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biology*. https://doi.org/10.1186/gb-2014-15-1-r6

Vance, K. W., & Ponting, C. P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2014.06.001

Vance, K. W., Sansom, S. N., Lee, S., Chalei, V., Kong, L., Cooper, S. E., … Ponting, C. P. (2014). The long non-coding RNA paupar regulates the expression of both local and distal genes. *EMBO Journal*. https://doi.org/10.1002/embj.201386225

Wang, F., Flanagan, J., Su, N., Wang, L. C., Bui, S., Nielson, A., … Luo, Y. (2012). RNAscope: A novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *Journal of Molecular Diagnostics*. https://doi.org/10.1016/j.jmoldx.2011.08.002

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., … Chang, H. Y. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. https://doi.org/10.1038/nature09819

Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J. P., & Li, W. (2013). CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt006

Wang, P., Xue, Y., Han, Y., Lin, L., Wu, C., Xu, S., … Cao, X. (2014). The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science*. https://doi.org/10.1126/science.1251456

Wang, Y., Dye, C. A., Sohal, V., Long, J. E., Estrada, R. C., Roztocil, T., … Rubenstein, J. L. R. (2010). Dlx5 and Dlx6 Regulate the Development of Parvalbumin-Expressing Cortical Interneurons. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.5963-09.2010

Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs: Functional surprises from the RNA world. *Genes and Development*. https://doi.org/10.1101/gad.1800909

Wojcik, S. M., Katsurabayashi, S., Guillemin, I., Friauf, E., Rosenmund, C., Brose, N., & Rhee, J. S. (2006). A Shared Vesicular Carrier Allows Synaptic Corelease of GABA and Glycine. *Neuron*. https://doi.org/10.1016/j.neuron.2006.04.016

Wu, L., Murat, P., Matak-Vinkovic, D., Murrell, A., & Balasubramanian, S. (2013). Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins. *Biochemistry*. https://doi.org/10.1021/bi401085h

Wu, X., & Sharp, P. A. (2013). XDivergent transcription: A driving force for new gene origination? *Cell*. https://doi.org/10.1016/j.cell.2013.10.048

Yang, L., Froberg, J. E., & Lee, J. T. (2014). Long noncoding RNAs: Fresh perspectives into the RNA world. *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2013.10.002

Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., … Shen, X. (2015). Opposing roles for the lncRNA haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. *Cell Stem Cell*. https://doi.org/10.1016/j.stem.2015.03.007

Yoon, J.-H., Abdelmohsen, K., & Gorospe, M. (2013). Posttranscriptional Gene Regulation by Long Noncoding RNA. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2012.11.024

Zhang, Y., Zhang, X. O., Chen, T., Xiang, J. F., Yin, Q. F., Xing, Y. H., … Chen, L. L. (2013). Circular Intronic Long Noncoding RNAs. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2013.08.017

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., & Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*. https://doi.org/10.1007/978-3-642-18354-6_20

Acknowledgments

Curriculum Vitae

Joshi, Parth D.

Born: March 02, 1991
            Aliabada, India.

Email: pjoshi@mpibpc.mpg.de
Contact: +49-15217571046

Education:

2015 – 2018	Doctorate in Genes and Development
	Max Planck Institute for Biophysical Chemistry, Göttingen
	GGNB - Genes and Development PhD Program
2012 – 2014	Master of Technology in Genetic Engineering
	SRM University, India
2008 – 2012	Bachelor of Technology in Genetic Engineering
	SRM University, India

Professional Experience:

Oct.'13 – May'14	Master thesis in the Dept. of Genes and Behavior,
	Max Planck Institute for Biophysical Chemistry, Göttingen
Feb.'12 – June'12	Bachelor thesis in the Genomics Laboratory,
	SRM University, India
Dec.'10 – Apr.'11	Research Assistant in the Dept. of Biotechnology,
	SRM University, India
Dec.'09	Student Research Assistant
	Disha Life Sciences, India

Publications:

Fan Z, Zhao M, **Joshi Parth D**…… Yan J. A class of circadian long noncoding RNAs mark enhancers modulating long-range circadian gene regulation. *Nucleic Acids Res.,* 2017, doi: 10.1093/nar/gkx156.

Joshi Parth D. Identification of long noncoding RNA expression during mouse development. *Proceedings of 6th Genetic Engineering Research Symposium,* 2014, SRM University.

Conferences and Symposiums:

Poster Presentation talk at the **Non-Coding Genome** Symposium at the European Molecular Biology Laboratory (EMBL), Heidelberg, Oct.-2015.
Topic: *Identification of long noncoding RNA expression during mouse development.*

Annual Molecular Biology meeting, **Horizons in Molecular Biology,** Max Planck Institute for Biophysical Chemistry, Göttingen. (2014 - 2017).

International Conference on **Next Generation Bioinformatics** at SRM University, India, Feb.-2013.

International Conference on **Bioengineering** at SRM University, India, Aug.-2010.

Poster Presentation at the International Conference on **The Key Role of Life Sciences for Sustainable Development in the World**, Hindustan University, India, Sep.-2010.

Honors and Awards:

Received 'A+' (GPA: 10/10) grade from SRM University, India for my Master's thesis.

Received 'Certificate of Merit' from SRM University, India for my Bachelor's thesis.

Received scholarship of 50,000 INR from SRM University to pursue my Master's thesis from Germany.

Extra-curricular activities:

Representative of the Genes and Development PhD program of GGNB, Göttingen. Jan'15-May'16.

Representative of the PhD & Postdoc community of the Max Planck Institute for Biophysical Chemistry, Göttingen. Dec.'15-Dec.'17.

Joshi, Parth D.