

South Dakota State University

# Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange

---

Electronic Theses and Dissertations

---

2020

## *In Situ* Analysis of LINE-1 Promoter Activity using LacZ Transgenic Mice

Partha Sarathi Saha  
South Dakota State University

Follow this and additional works at: <https://openprairie.sdstate.edu/etd>



Part of the [Developmental Biology Commons](#), [Genetics and Genomics Commons](#), and the [Microbiology Commons](#)

---

### Recommended Citation

Saha, Partha Sarathi, "*In Situ* Analysis of LINE-1 Promoter Activity using LacZ Transgenic Mice" (2020). *Electronic Theses and Dissertations*. 3916.  
<https://openprairie.sdstate.edu/etd/3916>

This Dissertation - Open Access is brought to you for free and open access by Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact [michael.biondo@sdstate.edu](mailto:michael.biondo@sdstate.edu).

*IN SITU* ANALYSIS OF LINE-1 PROMOTER ACTIVITY USING LACZ  
TRANSGENIC MICE

BY

PARTHA SARATHI SAHA

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Pharmaceutical Sciences

South Dakota State University

2020

## DISSERTATION ACCEPTANCE PAGE

Partha Sarathi Saha

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Wenfeng An

Advisor

Date

Omathanu Perumal

Department Head

Date

Dean, Graduate School

Date

This dissertation is dedicated  
to  
*my family*

## ACKNOWLEDGEMENTS

It was indeed a long journey to pursue this degree. Over time, I came across many individuals, from whom I received help in different ways. I appreciate them here for their cooperation. First of all, I would like to thank Dr. Wenfeng An to give me an excellent opportunity to work on this project of such a great significance. I always regarded him as a perfect embodiment of both passion and patience. He always wanted me to grow my scientific curiosity and build better scientific communication. He always used to take a keen interest in how the experiments progressed. I cannot remember any time when anything related to my experiment was set to wait for the next morning. He was always attentive for my projects so that I could make better progression every day. He always put extra care so that I could troubleshoot my experiments earlier and could save my time and efforts. I always found him dedicated to establishing me as an independent researcher. He is one of them I met in my life so far, who always believed in my abilities and supported all the time to discover new ones. In addition to this project, he also kindly allowed me to work on some other projects. All these gave me great opportunities to be more confident in handling different research projects in parallel. Overall, it was an honor to get him as my doctoral mentor. I also appreciate my doctoral committee members, Drs. Feng Li, Reineke Joshua, Alan Erickson for guiding me over the years. I appreciate their time and efforts. I am thankful to Trenton LaCanne, an undergraduate student, whom I trained to make my right arm later. It was amazing help that he rendered with the QuPath quantification aspects. Kelly Graber, a Senior Research Specialist at Sanford Research in Sioux Falls, did a wonderful job by doing whole-slide scanning of the tissue sections consistently over a long period of time. I thank Karly Ackermann to characterize some of

the single-copy mouse lines before I began working on this project. I am also thankful to Dr. Simon Newkirk for helping me daily in lab with the necessary skills and guidance. I also appreciate Yuchi Hu to help me with plotting various graphs with R. Additionally, it was wonderful to receive help with the transfection from Lingqi Kong. Most of the antibodies I used in the fourth chapter of this dissertation was generously gifted by Professor Chun-Li Zhang at UTSW Medical Center, TX, and I would like to give my hearty thanks for his kindness. I also want to take the opportunity to thank Dr. Michele Mucciante and Zubke Amanda at Animal Resource Wing, SDSU for their wonderful, daily cooperation to take care of the mouse colonies. Finally, I would like to thank Dr. Omathanu Perumal, the Head of the department, and Dr. Jayrama Gunaje to keep a positive impact on my grad-life. Truly, this project was impossible without assistance from the mentioned individuals.

## CONTENTS

	Page#
ABSTRACT.....	viii
<b>Chapter 1</b> .....	<b>1</b>
<b>General introduction</b>	
1. Transposable elements.....	1
2. LINE-1 elements.....	2
3. Structure of LINE-1 elements and their encoded proteins.....	3
4. LINE-1 retrotransposition: how does it work? .....	4
5. LINE-1 Promoter.....	4
6. Regulation of LINE-1 expression.....	6
7. LINE-1 activity in the development.....	10
8. LINE-1 in the embryo, somatic tissues, and disease contexts.....	12
9. Tools for studying mammalian LINE-1 retrotransposition.....	20
10. Detection of LINE-1 expression.....	24
11. Concluding remarks.....	25
12 Objective of the current project.....	26
13. References .....	28
<b>Chapter 2</b> .....	<b>36</b>
<b>Locus-specific LINE-1 promoter activity</b>	
1. Abstract .....	36
2. Introduction .....	37
3. Materials and methods.....	38
4. Results .....	51
5. Discussion .....	93
6. References.....	115

**Chapter 3 .....121****Specific locus- and orientation-dependent LINE-1 promoter activity**

1. Abstract.....	121
2. Introduction.....	121
3. Materials and methods.....	124
4. Results .....	126
5. Discussion .....	134
6. References.....	142

**Chapter 4 .....145****Cell-specific LINE-1 promoter activity**

1. Abstract .....	145
2. Introduction .....	145
3. Materials and methods.....	147
4. Results.....	150
5. Discussion .....	161
6. References.....	162

**Chapter 5 .....164****General discussion and future directions**

1. References.....	166
--------------------	-----



## ABSTRACT

*IN SITU* ANALYSIS OF LINE-1 PROMOTER ACTIVITY USING LACZ  
TRANSGENIC MICE

PARTHA SARATHI SAHA

2020

Apart from an evolutionary role, transposable elements have been implicated in animal development and also in pathophysiology. Non-LTR retrotransposons– LINE-1, Alu and SVA - are responsible for over 120 cases of human genetic diseases as heritable insertions, and are emerging as an important etiological factor for cancer and neurological disorders as somatic mutations. It is estimated that among the total number of 500,000 LINE-1s presents in the human genome, 80-100 LINE-1s remain competent for retrotransposition. Retrotransposition is only possible when LINE-1 is expressed. Because LINE-1 transcription is regulated by its 5'UTR promoter, it is essential to understand the spatiotemporal control of LINE-1 promoter activity. The huge abundance, repetitive nature and complex expression patterns of LINE-1s in the human and mouse genomes necessitate the development of innovative approaches and the careful design of experimental procedures used to study these elements. The primary objective of this dissertation was to develop and validate a mouse model, which can be utilized for studying LINE-1 promoter activity *in vivo*. Here, we utilized an *in situ* staining technique to quantify the endogenous LINE-1 promoter activity in different organs of the mouse model - to understand any organ-specific regulation of the mouse endogenous LINE-1 promoter activity. Moreover, by integrating the transgene into random or specific genomic loci in different orientations, we characterized the locus-dependent as well as the orientation-dependent expression

patterns. In all these aspects, we attempted to understand LINE-1 promoter regulation during different periods of mouse development. Lastly, we also attempted to understand the cell-specific regulation, especially in the brain. We reported here organ-specific, age-linked, locus-associated, and orientation-dependent LINE-1 promoter activities in the mouse genome. Our study provides novel insights into LINE-1 biology and the new mouse model will serve as an invaluable tool to the LINE-1 field.

# Chapter 1

## General Introduction

### 1.1 Transposable elements

Transposable elements describe a unique form of DNA that is mobile in the genome. That is, transposable elements are sequences that are capable of “jumping” and inserting themselves into new genomic contexts. The movement of sequences within the genome can generate a genetic variation by creating new sequences, but transposable elements can also be disruptive to normal gene function by interrupting gene sequences and producing mutations.

In humans, the majority of transposable elements have lost their ability to jump around the genome over evolutionary time, but a few have retained the ability to do so. These transposable elements that can still insert themselves elsewhere are often referred to as “transposition competent” (Faulkner and Billon, 2018). Transposable elements those have lost their ability to jump around the genome have usually done so through acquiring mutations that disrupt their activity (Bodak *et al.*, 2014).

The functions of these transposable elements are sometimes unclear, they were initially assumed to be evolutionary remnants of parasitic infection. But more recent evidence implicates them in a wide variety of processes, including regulating gene expression (Elbarbary *et al.*, 2016), cell identity (Percharde *et al.*, 2018), and promoting genome

variation (Richardson *et al.*, 2015). Transposable elements represent almost half of mammalian genomes (Belancio *et al.*, 2008).

Transposable elements move to new locations in the genome by either by transposition (in the case of DNA transposons) or retrotransposition (in the case of retrotransposons). Transposition is a “cut and paste” mechanism, where the element moves from one region to another, and retrotransposition is a “copy and paste” mechanism, where the element first makes a copy of itself by reverse transcription and then integrates elsewhere in the genome (Faulkner and Billon, 2018). This introductory chapter will discuss recent advances in our understanding of the expression and control of LINE-1 elements, an important class of retrotransposons and only actively mobile elements in the human genome.

## **1.2 LINE-1 elements**

LINE-1 elements are a member of the family of Long INterspersed Elements (LINEs or L1s), a family of retrotransposons that are widespread in the mammalian genome, and present in the genome of other eukaryotes. Retrotransposons require an RNA intermediate to function as transposable elements, and first make a copy of themselves before jumping elsewhere in the genome (Boeke *et al.*, 1985). LINE-1 retrotransposons represent a group of retrotransposons termed non-long terminal repeat (LTR) retrotransposons, reflecting the genetic structure of the elements (Xiong and Eickbush, 1990).

LINE elements are abundant in eukaryotes and makeup approximately 17.5% and 19.9% in the human and mouse genomes, respectively (Lander *et al.*, 2001; Waterston *et al.*, 2002). While most transposable elements in the mammalian genome are inactive, a small percentage (less than 1%) of LINE-1 elements remain capable of mobilization and

generating variation in human and mouse genomes (Beck *et al.*, 2011; Faulkner and Garcia-Perez, 2017; Goodier *et al.*, 2001). This makes LINE-1 elements a unique and important class of transposable elements.

The movement and integration of LINE-1 elements in the genome continue to drive evolution (Ostertag and Kazazian, 2001), but to prevent inappropriate LINE-1 expression from driving mutagenesis, cells have developed mechanisms of diminishing LINE-1 insertions in the majority of temporal and spatial contexts. Some of these mechanisms involve interference with the transcription of LINE-1, and other mechanisms occur at a later stage, regulating LINE-1 at the RNA or protein level (Bodak *et al.*, 2014).

### **1.3 Structure of LINE-1 elements and their encoded proteins**

Transposition-competent LINE-1 elements are roughly 6-7 kb in length (Scott *et al.*, 1987), containing a 5'UTR with an internal promoter (Minakami *et al.*, 1992), two open reading frames (ORFs), and a 3'UTR containing a poly (A) tail (Dombroski *et al.*, 1991). The LINE-1 ORFs encode two proteins: ORF1p and ORF2p (Scott *et al.*, 1987). These proteins are required for efficient mobilization of LINE-1 elements (Moran *et al.*, 1996). Despite their similar names, the encoded proteins are quite different from one another.

ORF1p is an approximately 40kDa protein possessing RNA binding and chaperoning activity (Kolosha and Martin, 2003; Martin and Bushman, 2001). The protein binding of LINE-1 RNA to ORF1p is a necessary step of retrotransposition (Kulpa and Moran, 2005). ORF2p is a larger protein (150kDa) that possesses both endonucleases (Feng *et al.*, 1996) and reverse transcriptase activities (Martin, 2010; Mathias *et al.*, 1991). The activities of both proteins are required for LINE-1 mobility.

#### **1.4 LINE-1 retrotransposition: how does it work?**

Active LINE-1 retrotransposition requires the transcription of a full-length RNA molecule from the 5' internal promoter, which is translated into the proteins ORF1p and ORF2p in the cytoplasm (Alisch *et al.*, 2006; Moran *et al.*, 1996). The LINE-1 mRNA is then capable of binding the translated proteins, forming a ribonucleoprotein (RNP) complex with their respective proteins. This RNP is then imported into the nucleus, where it initiates reverse transcription at the target site with a process called target-primed reverse transcription (TPRT) (Luan *et al.*, 1993). As previously eluded to, most LINE-1 elements present in the mammalian genome are immobilized. This is because of poor processivity during the reverse transcription step or potentially because of post-transcriptional and post-translational LINE-1 RNA degradation.

#### **1.5. LINE-1 Promoter**

Although the mouse and humans LINE-1s share similar ORFs, they differ markedly within the 5' UTR sequence. This difference is believed to be responsible for the differences in transcriptional activities in these two species (Severynse *et al.*, 1992). Particularly, the 5'-UTRs of the full-length human LINE-1s carry two types of internal promoters, namely sense and antisense (Hancks & Kazazian, 2012). In contrast, 5'-UTRs of the full-length mouse LINE-1s contains a sequence of 200 bp. These tandemly repeated sequences are known as monomers (Severynse *et al.*, 1992). The number of the monomer repeats may vary among the individual mouse LINE-1 families, where the copy number is linked with the LINE-1 transcriptional activity (Severynse *et al.*, 1992; DeBerardinis *et al.*, 1999).

Moreover, the monomers are used to subclassify LINE1 into different subfamilies, and their presence or absence generally determine which transcription factors (TFs) would regulate the transcription of these elements. For instance, the promoter of the murine Tf subfamily binds to the YY1 (YY1 Transcription Factor) factor, whereas the TA subfamily does not conserve this region (Severynse *et al.*, 1992). That again suggests that the promoter activity pattern may vary in between the subfamilies and species.

### **1.5.1 Sense promoter**

The first 100 base pairs of the 5'- UTR carries an internal sense promoter. This is essential for transcription initiation (Swergold 1990). However, this region shows high variability. In other words, this is very poorly preserved regions in mammals (Zimmerman, 1997; Eppig *et al.*, 2012). Additionally, this exhibits a frequent stop codons and a high GC content (>50%) in comparison to the rest of the LINE1 sequence (Aporntewan & Mutirangura, 2011, Eppig *et al.*, 2012). These features suggest that this region's function could be controlling the LINE1 transcription (Aporntewan & Mutirangura, 2011; Eppig *et al.*, 2012).

### **1.5.2 Antisense promoter**

Besides the transcription of the forward LINE-1 RNA, LINE-1 elements also have antisense promoters that can drive the expression of nearby genes. The antisense promoter is found in the 5'UTR (Speck, 2001) of a LINE-1 element. The transcribed mRNA transcripts (initially discovered in 2001) were chimeric, containing a 5'UTR from LINE-1, and exons from the nearby genes. Subsequently, almost 1000 antisense transcripts were identified in a comprehensive computational study (Criscione *et al.*, 2016). Some of these chimeric transcripts were also found to be unique to (or, up-regulated in) cancer cell lines

(Cruickshanks and Tufarelli, 2009; Weber *et al.*, 2010). More recent studies have implicated LINE-1 antisense promoter in regulating the tissue-specific expression of long non-coding RNAs (Chishima *et al.*, 2018). A large number of these chimeric transcripts present in the mammalian genome could reflect a global mechanism for regulating lineage-specific transcription programs. This is consistent with other studies suggesting LINE-1 elements can have regulatory functions in the mammalian genome (Elbarbary *et al.*, 2016).

## **1.6 Regulation of LINE-1 expression**

The expression of LINE-1 elements in mammals is tightly controlled, protecting their genomes from deleterious effects of random integrations. Initial work suggested that the transcription of LINE-1 was restricted to the pluripotent cells of the primordial germline and in embryonic development (Garcia-Perez *et al.*, 2007; Martin and Branciforte, 1993; Ostertag *et al.*, 2002; Packer *et al.*, 1993; Trelogan and Martin, 1995) where a global wave of epigenetic remodelling that occurs during embryogenesis allows them to become upregulated. However, more recent evidence discussed above suggests that their regulation is complex, multifaceted and that they can be expressed in somatic cells.

### **1.6.1 LINE-1 and chromatin landscape**

An important layer that governs the regulation of mammalian gene expression is epigenetics and chromatin structure. Epigenetics refers to covalent modifications to the DNA and the associated proteins (chromatin) that do not modify the actual DNA sequence. Despite the lack of changes to the underlying DNA sequence or transcription factors, the effects of these modifications on gene activity and function can be profound (Mazzio and Soliman, 2012).



LINE-1 elements are no exception to this rule: a powerful tool for regulation of LINE-1 expression is through modulating the surrounding chromatin environment (Jachowicz and Torres-Padilla, 2016). One of the most studied mechanisms by which a repressive chromatin landscape can silence LINE-1 elements is through DNA methylation.

### **1.6.2 DNA methylation**

DNA methylation is the direct modification to the DNA sequence, through the deposition of a methyl group at the 5' carbon on the cytosine base (5mC). In the mammalian genome, this mark is primarily found at cytosine bases in the context of a CpG dinucleotide (Klose and Bird, 2006). DNA methylation is enzymatically catalyzed by a class of enzymes called DNA methyltransferases. These enzymes either act upon an unmethylated DNA strand, depositing de novo DNA methylation or act upon a hemimethylated transcript to maintain DNA methylation patterns after DNA replication. DNMT3A and 3B are the de novo methyltransferases, DNMT1 is the maintenance methyltransferase that faithfully recapitulates DNA methylation following semi-conservative replication (Edwards *et al.*, 2017).

DNA methylation is broadly associated with the repression of transcription; however, its precise relationship with gene activity is more complex than initially recognized. One genomic context in which its repressive role is less controversial is in the silencing of transposable elements (Arand *et al.*, 2012; Jones and Takai, 2001; Walsh *et al.*, 1998).

DNA is typically hypermethylated at the promoters of LINE-1 elements in mammalian cells (Meissner *et al.*, 2008). This DNA methylation is thought to have important functional

consequences because in mouse embryonic stem cells lacking any active DNA methyltransferase enzymes, the transcription of LINE-1 is elevated (Tsumura *et al.*, 2006).

### **1.6.2.1 DNA methylation in development**

The role of DNA methylation in silencing LINE-1 elements is certainly significant in somatic tissues (Arand *et al.*, 2012). The findings of various studies have indicated that DNA methylation has a complex relationship with LINE-1 activity, with cell-type-specific regulatory mechanisms. Further complicating this, it is the evidence that the regulation of LINE-1 elements by DNA methylation could be locus-specific (Philippe *et al.*, 2016; Vafadar-Isfahani *et al.*, 2017). Historically, studying individual LINE-1 elements has been technically challenging, but it has recently been shown that in some cases, DNA methylation at LINE-1 promoters does not correlate with their expression (Vafadar-Isfahani *et al.*, 2017). The locus-specific activity of LINE-1 elements was also analyzed in cancer cells in a report published in 2016. Researchers observed that the majority of transpositionally active LINE-1 elements (the youngest and human-specific LINE-1 subfamily, LINE-1HS-Ta) were transcribed from a few cell-type-specific loci (Philippe *et al.*, 2016). The researchers found that in their cancer cell lines where they detected LINE-1HS-Ta, the general mechanisms regulating LINE-1 activity were not perturbed. This suggests that locus-specific and cell-type-specific regulatory mechanisms for LINE-1 repression exist in distinct cellular contexts, and that heterogeneity can exist in an already tightly controlled and multifaceted system.

### 1.6.3 Histone modifications

Another mechanism of epigenetic regulation of transcription is through the post-translational modification of histone tails, a powerful regulator of chromatin structure and gene expression (Bannister and Kouzarides, 2011). Histone tails can be modified at several amino acid residues by the addition of small chemical groups, with the most common being methylation, acetylation and phosphorylation. The resulting landscape of histone modifications regulates the structure and function of chromatin, fine-tuning gene expression profiles in a different cell type- and locus-specific contexts. The trimethylation of lysine 9 on histone H3 (H3K9me3) is most commonly associated with the silencing of repetitive sequences and transposable elements (Karimi *et al.*, 2011).

Interestingly, H3K9me3-mediated LINE-1 silencing was found to be restricted to mouse embryonic stem cells. In lineage-committed cell types, DNA methylation was observed at the H3K9me3-repressed LINE-1 loci, suggesting a shift from histone methylation-dependent silencing to DNA methylation-mediated silencing (Walter *et al.*, 2016).

A very recent and comprehensive study aimed to explore the interplay of various chromatin marks at transposable elements in mouse embryonic stem cells (He *et al.*, 2019). To analyze the effects of histone modifications on LINE-1 expression, various histone-modifying enzymes were knocked down and changes in chromosome accessibility and transposable element expression were assessed. The results revealed that a complex landscape of histone modifications exists at LINE-1 elements. While the authors acknowledge that definitively mapping chromatin marks at LINE-1 sequences is difficult owing to their abundance and sequence identity, they find various histone marks enriched at LINE-1 elements including H4R3me2, H3K4me1, H3K27ac, H3k9me3 and others (He *et al.*, 2019). These histone

marks were overlapping and associated with changes to chromatin accessibility. This can be assayed by a technique called as Assay for Transposase-Accessible Chromatin followed by sequencing (ATAC-seq), and gene expression programs. This study highlights the complexity of the epigenetic regulation of these elements and suggests that much of the role of epigenetic systems in the cell is to manage the expression of mobile DNA.

Besides the transcriptional regulation of LINE-1 with the epigenetic regulation, several posttranscriptional, translational, and even posttranslational regulations of LINE-1 elements are also involved to restrict LINE-1 mobilization in the mammalian genome.

### **1.7 LINE-1 activity in the development**

LINE-1 expression during development has been a particularly important area of research. The dynamic remodeling of the transcriptional and epigenomic landscapes during the transition from pluripotency to the acquisition of lineage commitment affords LINE-1 elements an opportunity for upregulation and activity.

Initial understanding of LINE-1 activity led to the belief that LINE-1 elements were only expressed *in vivo* in the early stages of mammalian embryogenesis, or during the specification of primordial germ cells (precursors of the gametes). Later, retrotransposition assays showed LINE-1 expression and retrotransposition can occur in a variety of somatic cell types, and also *in vitro* culture models, including neural progenitor cells, cancer cell lines and other mammalian cell culture contexts (Muotri, 2016). LINE-1 expression and mobilization have been studied extensively both *in vivo* using mouse models and tissue samples, and *in vitro* using cultured cell assays.

### 1.7.1 Evidence of LINE-1 expression *in vivo* during development

The historical view of LINE-1 retrotransposition is that LINE-1 elements are only active during embryogenesis and malignant transformation. This was challenged by the finding that LINE-1 retrotransposition can occur in various cells of the brain, and LINE-1 mRNA and proteins have been detected in various adult somatic tissues (Belancio *et al.*, 2010; Richardson *et al.*, 2014). The activity of LINE-1 in somatic tissues and disease contexts will be discussed in detail in later sections.

LINE-1 activity during development *in vivo* must be tightly regulated to prevent genomic instability and the accumulation of mutations. This is particularly important during development, where the genome undergoes rapid and dynamic remodeling, associated with more permissive chromatin states (Seisenberger *et al.*, 2013). These permissive states are associated with genomic activation, and LINE-1 expression must be kept in check to ensure the cell retains control of its activity. The control of LINE-1 is particularly important in germ cells, where any new LINE-1 insertions will be passed on to the next generation. LINE-1 expression must be controlled in the pre-implantation embryo, which represents another crucial period of epigenomic remodeling.

Endogenous LINE-1 elements are expressed at various points in the mouse embryo (Jachowicz *et al.*, 2017; Packer *et al.*, 1993; Veselovska *et al.*, 2015; Watanabe *et al.*, 2008). LINE-1 expression has also been shown to be abundant in the two-cell stage (Fadloun *et al.*, 2013), and decrease over embryonic time between the 2 and 8 cell stage. The presence of these transcripts has been shown by Northern blot (Packer *et al.*, 1993), RNA Fluorescence in Situ Hybridization (FISH), and reverse transcriptase-polymerase chain reaction (RT-PCR) (Fadloun *et al.*, 2013).

LINE-1 mRNA and protein are expressed in the germline in human and mouse cells (Peaston *et al.*, 2004; Trelogan and Martin, 1995). However, to prevent their over-activation and the disruption of genome integrity by these elements, they are tightly controlled by various means. In the germline, specific proteins exist to post-transcriptionally regulate LINE- activity. These proteins (termed PIWI proteins) are discussed in detail in a later section. Without the regulatory effects of these proteins, LINE-1 transcripts accumulate in the testes of adult male mice, but not in wild type mice. This was shown using in situ hybridization as well as RT-PCR detecting mRNA transcripts (Aravin *et al.*, 2007; Carmell *et al.*, 2007).

## **1.8 LINE-1 in the embryo, somatic tissues, and disease contexts**

### **1.8.1 Expression of LINE-1 in the embryo**

Endogenous LINE-1 elements are highly expressed in the mouse embryo (Jachowicz *et al.*, 2017). LINE-1 elements are transcribed soon after fertilization (Fadloun *et al.*, 2013). Transcription is highest at the two-cell stage and is rapidly reduced after the blastocyst stage of development. Interestingly, this is (at least initially) primarily attributable to a loss of activating marks, rather than a gain of repressive marks (Fadloun *et al.*, 2013). LINE-1 expression is higher in the in vivo embryo than in cultured mouse embryonic stem cells (Jachowicz *et al.*, 2017). This could potentially be explained by the increase in DNA methylation in cultured cells relative to the cells of the inner cell mass (Ficz *et al.*, 2013).

Importantly, activation of LINE-1 elements during development does not appear to merely be a consequence of spurious activation because of epigenome remodeling. It has been shown that inhibiting LINE-1 expression before the two-cell stage interferes with the

development of the embryo, reducing developmental progression (Jachowicz *et al.*, 2017). LINE-1 elements have been implicated in modulating the wide-spread changes in chromatin organization observed during early mouse development (Jachowicz *et al.*, 2017). This suggests that LINE-1 expression has genuine roles in normal mammalian biology and the mRNAs are not simply problematic transcripts that are only activated when silencing mechanisms are perturbed. Recently, Percharde *et al.* revealed that in mouse embryonic stem cells (ESCs) and also in pre-implantation embryos, LINE1s have a critical role in directing the self-renewal capability, transcriptional regulations and also overall developmental potency (Percharde *et al.*, 2018).

### **1.8.2 Germ cell expression of LINE-1**

An important finding in the past let us believe that mutagenesis mediated by transposable elements may occur in the germline of humans. In that event, CYBB gene was inserted by LINE-1 most likely during meiosis I of maternal primary oocyte genome of a male patient with a chronic granulomatous disease (Boruha *et al.*, 2002). From then onwards, it is well established now that transpositional events in the germline might be a major source of genomic variations and diseases in the human population. In that case, the most active transposable elements: Alus, LINE-1s and SVAs are carrying out the events of reshaping the genomic landscape of haploid cells and at the same time causing different rare genetic diseases. It is noteworthy here that these reshaping events not only include insertional events but also include the deletions of the host DNA sequence (Gilbert *et al.*, 2002).

The exact timing of the endogenous *de novo* retrotransposition in human germline remained unclear. A recent study with mouse model indicates that heritable *de novo* insertions might start to take place as early as prior germline specification in mammals. In

that study, a new LINE-1 insertion was found in 3-5% of offspring, whose both parents had no such insertion in somatic cells, but the male parent had <1 copy insertion per cell in both testicles (Richardson *et al.*, 2017). That result indicated that new LINE-1 insertion in germline-restricted mosaic male parent took place, most likely, in early primordial germ cells (PGCs) before they colonize genital ridges to form testicles anytime during E10.5-E12.5 (Ewen *et al.*, 2010). Due to ethical considerations, the option of proving this notion of human PGC development is limited.

### **1.8.3 LINE-1 in somatic cells**

Initially, human somatic LINE-1 retrotransposition events were thought to be relatively rare. However, more recent studies have shown that LINE-1 expression and mobilization occurs in a variety of somatic cells, for example in tumors (Burns, 2017), and interestingly, in various tissues of the brain (Goodier, 2014; Suarez *et al.*, 2018).

### **1.8.4 LINE-1 in the brain**

Among the somatic organ systems, the brain has emerged as one of the most active sites for retrotransposition during development. It is thought that a certain level of retrotransposition might be advantageous for neuronal development by promoting genomic diversity. On the other hand, excessive expression or retrotransposition could have deleterious effects on neural functions.

Retrotransposition in the mammalian brain was first demonstrated in the laboratory mouse. In 2005, the Gage lab identified new retrotransposition events in many regions of the mouse brain (Muotri *et al.*, 2005). The insertions were initiated from a human LINE-1-based transgene and seemed to occur only in neurons but not in oligodendrocytes and astrocytes



during embryonic and adult neurogenesis. LINE-1 ORF1p was also detected in several regions of the brain, including the ventricular zone and the dentate gyrus of the hippocampus. In 2009, the same group showed variable but significant levels of increase in endogenous LINE-1 copies in multiple regions of human brain samples (Coufal *et al.*, 2009). Subsequently, two separate groups were able to confirm somatic LINE-1 insertions in the human brain at the sequence level (Baillie *et al.*, 2011; Evrony *et al.*, 2012). Together these studies established the brain as a hub for active retrotransposition in humans and heralded a new era of investigating the extent of somatic mosaicism in the brain and its functional implications.

How frequently does LINE-1 retrotransposition occur in a human brain? The answer to this question is complicated by both the methodology used and the inherent variation in retrotransposition. Using quantitative PCR on bulk samples the initial report suggested an increase of 80 copies of LINE-1 per cell in the hippocampus (Coufal *et al.*, 2009). The figure was later refined by two studies employing advanced single-neuron sequencing analyses. The Walsh lab estimated up to 1.1 somatic LINE-1 insertions per cortex and caudate neurons (Evrony *et al.*, 2012). The Faulkner lab's estimation was an average of 13.7 new insertions per hippocampus neuron (Upton *et al.*, 2015). The discrepancy between these two studies likely originated from technical variations in sequencing approaches, data analytics and validation methods (Evrony *et al.*, 2016). Indeed, factoring the most stringent validation criteria into the calculation, the frequency of unique LINE-1 insertions may be as low as 0.04 or 1 in every 25 neurons (Evrony *et al.*, 2012). The variation in estimated retrotransposition frequencies does not necessarily diminish the potential functional impact of such insertions. If we assume 0.04 insertions per neuron as

the tangible minimum, there will still be approximately 3.4 billion unique somatic insertions among 86 billion neurons in a typical adult human brain! Also, somatic retrotransposition is variable in different individuals as well as in different regions of the brain.

Information about the developmental timing of LINE-1 retrotransposition in the brain remains scarce. Nevertheless, important insights have been gained from lineage tracing analysis of two somatic LINE-1 insertions in a normal human brain (Evrony *et al.*, 2015). One insertion was distributed over the entire left hemisphere and present not only in neurons but also in non-neuronal cells, suggesting that it arose in one of the earliest progenitor cells of the central nervous system. In contrast, the second insertion was restricted to neurons at the left middle frontal gyrus, suggesting that it occurred relatively late during cortical development in the embryo. Unlike the original mouse study, single-cell analyses have also identified somatic LINE-1 insertions in glial cells (Evrony *et al.*, 2015; Upton *et al.*, 2015; Erwin *et al.*, 2016). Whether differentiated glial cells support *de novo* retrotransposition is unknown as many such insertions could have originated in progenitors common to neurons and glia (Evrony *et al.*, 2015; Upton *et al.*, 2015). Although it has not been established *in vivo*, *in vitro* cell culture experiments have provided a comparison of neural stem cells, neuronal progenitor cells (NPCs) and terminally differentiated neurons in their capabilities of supporting retrotransposition. The overwhelming evidence pinpointed neuronal progenitor cells as the hub for LINE-1 insertional events. Besides, both human and rodent LINE-1 promoters possess overlapping SOX2/WNT binding sites. In these cells, *Sox2* is downregulated as *Wnt* is upregulated. The latter activates LINE-1 promoter activity and transcription (Kuwabara *et al.*, 2009).

Whether LINE-1 retrotransposition plays a functional role in normal brain physiology is not yet understood. Given the connectivity of brain cells and the estimated collective mutational burden in the entire brain, somatic retrotransposition has the potential to exert a significant impact on neuronal functions. Of relevance, in both rat NPCs and human hippocampal neurons, somatic LINE-1 insertions can occur in neuronally expressed genes, including those that are involved in different synaptic processes (Muotri *et al.*, 2005; Upton *et al.*, 2015). Notably, besides insertional mutagenesis, LINE-1s can also remodel the genomic landscape of neurons by inducing large genomic DNA deletions, a process that is retrotransposition independent (Erwin *et al.*, 2016). In this context, genomic diversity may beget a functional diversity within the human brain.

On the other hand, excessive LINE-1 mobilization in the brain has been linked to many neurodevelopmental and neurodegenerative disorders. For example, the rate of LINE-1 retrotransposition was found to be higher in NPCs derived from human tissue of a patient with Rett syndrome (RTT), a neurodevelopmental disorder due to a mutation in the X-linked *MECP2* gene (Muotri *et al.*, 2010).

### **1.8.5 LINE-1 in other somatic tissues**

LINE-1 mRNA has been detected, although mostly at low levels, in a variety of cell types. Interestingly, a human retrotransposition assay in mouse models suggests there is no inherent barrier to LINE-1 protein expression and activity in somatic cells (Ostertag *et al.*, 2002). A comprehensive study from 2010 studied an array of tissues, finding LINE-1 RNA, protein, and de novo insertions in most human (somatic) tissues (Belancio *et al.*, 2010). Whilst specific reports of LINE-1 expression and activity in healthy somatic tissues are

rare, expression and activity have been noted in cells of the gastrointestinal system as well as in the esophagus.

In a study investigating Barrets Esophagus and esophageal cancer found that although LINE-1 retrotransposition events were seldom found in normal tissue, LINE-1 protein expression was detected in all tissues examined (Doucet-O'Hare *et al.*, 2015). LINE-1 protein expression has also been found in cells closely associated with male germ cells, such as Sertoli cells and vascular endothelial cells, likely linked to transposition in the gametes (Ergun *et al.*, 2004).

These reports suggest that whilst the expression of LINE-1 occurs outside of the germline and neuronal tissues, the level of activity is highly heterogeneous between tissues. Active retrotransposition is likely inhibited by other means in these cells to prevent the accumulation of DNA damage.

### **1.8.6 LINE-1 in disease states**

LINE-1 is associated with various disease states (Hancks and Kazazian, 2016). This includes in the initiation and progression of cancers, autoimmune disorders and Mendelian diseases.

In the germline, LINE-1 can act as a mutagenic agent through insertional mutagenesis – disrupting exons and inducing double-stranded breaks (Belancio *et al.*, 2008). This has been shown as the causative mutation in cases of haemophilia A (Kazazian *et al.*, 1988), choroideremia (Van den Hurk *et al.*, 2007),  $\beta$  thalassemia (Lanikova *et al.*, 2013) and various other diseases reviewed (Beck *et al.*, 2011). Interestingly, a high number of these

disease-causing mutations are found on the X chromosome, potentially implicating recombination in LINE-1 insertions (Belancio *et al.*, 2008).

LINE-1 insertions can also occur in somatic cells, although these effects are not inherited by the next generation. LINE-1 over-expression and mobilization has also been associated with multiple neuropathologies (Bundo *et al.*, 2014; Liu *et al.*, 2016; Suarez *et al.*, 2018; Tan *et al.*, 2018) and is linked with genomic instability and malignancy in cancers (Burns, 2017).

### **1.8.7 LINE-1 expression in cancer**

LINE-1 protein expression is a hallmark of malignancy (Rodic *et al.*, 2014). LINE-1 hypomethylation, activation and integration is associated with many cancers, often correlated with poor prognosis. LINE-1 integration can be a source of genome instability through inducing DNA damage, insertional mutagenesis and chromosomal rearrangement. The dysregulation of LINE-1 elements, often by hypomethylation, can contribute to the pathogenicity of tumors (Briggs *et al.*, 2018; Burns, 2017; Carreira *et al.*, 2014; Schulz, 2006; Doucet-O'Hare *et al.*, 2015; Miki *et al.*, 1992; Kerachian and Kerachian, 2019; Rodic *et al.*, 2014). Certain tumor types are more prone to LINE-1 retrotransposition (**Table 1.1**). For example, tumors of the gastrointestinal tract (Lee *et al.*, 2012; Solyom *et al.*, 2012), as well as hepatocellular carcinomas, prostate and ovary cancers. These retrotransposition events usually correlate with LINE-1 mRNA and protein expression, suggesting that mechanisms to diminish LINE-1 expression have been compromised (Burns, 2017; Rodic *et al.*, 2014). The activity of LINE-1 in cancer pathogenesis is heterogeneous; sometimes LINE-1 expression and insertions are early events in tumorigenesis (Scott *et al.*, 2016;

Tubio *et al.*, 2014), in other cases, they accumulate in downstream events, likely as an effect of better conditions for LINE-1 expression to occur (Burns, 2017).

Cancer group	Criteria	Cancers types or origins
Very-high activity	>10 insertions per sample	Oesophageal adenocarcinoma Lung squamous cell carcinoma Head and neck squamous cell carcinoma Colorectal adenocarcinoma
High activity	5–10 insertions per sample	Gastric carcinoma Ovarian adenocarcinoma Uterine adenocarcinoma Cervical squamous cell carcinoma
Medium activity	1–5 insertions per sample	Pancreatic adenocarcinoma Prostate adenocarcinoma Breast adenocarcinoma
Low activity	<1 insertion per sample	Lung adenocarcinoma Liver hepatocellular carcinoma Bone osteosarcoma
Very-low activity	~0 insertion per sample	Brain and hematolymphoid systems

**Table 1.1: Classification of cancers in human based on somatic retrotransposition activity** (adopted from Saha & An, 2019).

## 1.9. Tools for studying mammalian LINE-1 retrotransposition

Studying mobile elements has been historically challenging. This is due to their abundance in the genome, repetitive sequences, and the accumulation of polymorphisms. However, the development of cleverly designed reporter systems as well as an adaptation of classic molecular biology techniques and advances in sequencing technology have produced a range of tools for use in studying LINE-1 biology. Some of these tools and methods are discussed below.

### 1.9.1 LINE-1 retrotransposition reporter constructs

The first published cultured cell LINE-1 retrotransposition assay in 1996 represented a significant advance in the field, as it allowed retrotransposition to be studied in real-time

(Kopera *et al.*, 2016; Moran *et al.*, 1996). The rationale behind this assay is the integration of a reporter construct that is only detectable when a LINE-1 element is transcribed, reverse transcribed, and integrated elsewhere in the genome. A retrotransposition indicator cassette is integrated into 3'-UTR of LINE-1, in the opposite direction of LINE-1 transcription. This cassette consists of a reporter gene sequence, which is interrupted by an intron which is transcribed in the same direction as the LINE-1 mRNA. The reporter construct can only be expressed when transcription occurs from the LINE-1 promoter, which splices out the intron from the reporter cassette, resulting in reverse transcription of LINE-1 RNA and the integration of new copies of the LINE-1 sequence as well as the reporter cassette into the genome. The cells possessing a successful integration can be selected based on the presence of the reporter construct, and researchers now have a set of tools to analyze LINE-1 activity (Kopera *et al.*, 2016; Moran *et al.*, 1996).

Several adaptations of this assay have permitted various studies on LINE-1 activity in a range of systems (Rangwala and Kazazian, 2009). This includes studies of LINE-1 retrotransposition in the neural progenitor cells (Coufal *et al.*, 2011; Coufal *et al.*, 2009), non-dividing primary human cells (Kubo *et al.*, 2006) and the generation of a mouse model (Ostertag *et al.*, 2002). Important regulatory questions have also been investigated using derived assays, including the epigenetic silencing of the LINE-1 retrotransposition (Garcia-Perez *et al.*, 2010), and the cellular kinetics of retrotransposition (Ostertag *et al.*, 2000). However, concerns are there regarding the robustness of these tools which may bring about variables between the assays to cause the misinterpretation of the results (Cook and Tabor, 2016).

### 1.9.2 Studies in cultured cells

Embryonic stem cells are a frequently used model for studying the regulation of molecular events and the role of different regulatory factors. This is a result of the plastic genome of pluripotent cells and the ability of these cells to transition into a variety of different cell types in culture. Embryonic stem cells are of particular interest in the field of LINE-1 biology, as LINE-1 transcripts, as well as active LINE-1 transposition, are frequently detected in these cells (Garcia-Perez *et al.*, 2007).

Using induced pluripotent stem cells (iPSCs) from human neonatal dermal fibroblasts, it has also been shown that LINE-1 expression is elevated during reprogramming from the somatic cell to the induced pluripotent state. This activity resulted in low-level insertions of LINE-1 elements (Arokium *et al.*, 2014). This study could be of importance to the iPSC field, as it cautioned researchers about potential genotoxic effects that occur during somatic cell reprogramming.

### 1.9.3 Mouse models

To better understand human LINE-1 retrotransposition, Ostertag and colleagues generated a transgenic mouse model, in which eGFP is conditionally expressed in the spermatocytes through a spermatozoa-specific preproacrosin promoter (Ostertag *et al.*, 2002). A functional eGFP is only produced when a retrotransposition event has occurred (Moran *et al.*, 1996). This is because the eGFP cassette contained an antisense  $\gamma$ -globin intron that cannot be spliced out. However, when the cassette is cloned into the LINE-1 3'UTR, in an antisense orientation, retrotransposition can remove the antisense intron and produce a functional eGFP (Moran *et al.*, 1996; Ostertag *et al.*, 2002; Ostertag *et al.*, 2000). Using



this system, researchers discovered that one *de novo* insertion was occurring in every 70 sperm for mice in the germ cells of mice. Although, later An et al found that the mutagenic effect of LINE-1 insertions was relatively high as 1 insertion per 3 sperms using CAG-ORFeus transgenic mouse line (An *et al.*, 2006). Mouse models have also been utilized to show that integration events occur more often in embryogenesis than in the germ cells, generating somatic mosaicism (Kano *et al.*, 2009).

#### **1.9.4 High throughput sequencing**

Technical challenges are put forward by the sequence abundance of LINE-1 and their polymorphism in the genome for their detection. To circumvent these issues, more advanced methods of high throughput sequencing have been implemented to study the expression, regulation and activity of LINE-1 elements in the genome (Xing *et al.*, 2013). For example, Retrotransposition Capture Sequencing (RC-seq) (Baillie *et al.*, 2011; Sanchez-Luque *et al.*, 2016), which is a method that enriches sequencing libraries for retrotransposon insertions. RC-seq achieves this using biotinylated capture probes, which target the 5' and the 3' end of the LINE-1 consensus sequence. This reduces the level of PCR amplification required and limits biases associated with normal genome-wide sequencing of rare genomic elements, such as heterogeneous retrotransposition events.

#### **1.9.5 *In vitro* biochemical assays**

*In vitro* biochemical assays have also been valuable in providing insights into LINE-1 function (Viollet *et al.*, 2016). These have been focused on detecting the retrotransposition activity of ORF2p.

The LINE-1 Element amplification Protocol (LEAP) enables researchers to assess the ability of ORF2p to reverse transcribe LINE-1 mRNA in vitro using the purified LINE-1 RNP from human cells harboring LINE-1 expression constructs (Kopera *et al.*, 2016; Kulpa and Moran, 2006; Viollet *et al.*, 2016). The assay involves the transfection of cells with constructs expressing differentially tagged ORF1p and ORF2p. The RNP complexes are purified from cells by centrifugation or immunoprecipitation. The RNP is then incubated with an oligonucleotide (termed as LEAP adapter) to prime cDNA synthesis. The LINE-1 cDNAs are the PCR-amplified primers complementary to the adapters and the LINE-1 construct. The PCR products can then be visualized and characterized (Kopera *et al.*, 2016). This assay has been used to investigate the process of target-primed reverse transcription. The direct LINE-1 extension assay (DLEA) is a similar assay with an alternative design to detect reverse transcription of LINE-1 mRNAs. DLEA involves the incorporation of a radiolabeled nucleotide before primer elongation (Monot *et al.*, 2013; Viollet *et al.*, 2016).

ORF1p, is one of the proteins essential for retrotransposition, is expressed in large extents in the cellular cytoplasm. These proteins can also be targeted with monoclonal antibodies as a robust indicator of LINE-1 expression in cells or tissues (Sharma *et al.*, 2016).

## **1.10 Detection of LINE-1 expression**

Studies of retrotransposons have been largely hindered by their repetitive and abundant nature. This makes amplification, detection and sequencing of these regions challenging.

**1.10.1 Protein-based detection:** One way to study LINE-1 activity in terms of its expression and localization is through classical detection methods, such as immuno-based

assays. This typically involves using an antibody against ORFP2. Whilst this is a robust method, it will not detect LINE-1 elements that are transcribed into RNA but not translated into proteins. Therefore, it does not take into account the post-transcriptional regulation of LINE-1 transcripts. (Sharma *et al.*, 2016). However, it is a useful method for detecting protein levels as well as localization at a single nucleus level.

**1.10.2 Nucleic acid-based detection:** First, RNA-FISH is a technology that is useful for detecting LINE-1 expression, as it also benefits from single nucleus resolution (Jachowicz *et al.*, 2017). However, the advantage of RNA-FISH over immune-detection methods is that nascent transcripts are assayed, meaning transcription is more directly measured. On the other hand, LINE-1 insertion events have also been tracked with DNA FISH, detecting retrotransposition patterns using a LINE-1/neomycin vector, which is only detectable using FISH probes when the Neomycin gene has been reverse transcribed as a result of LINE-1 activity (Bojang and Ramos, 2016). Microscopy can also be implemented finally to both of these approaches to analyze the LINE-1 expression.

Northern blot analysis remains the most traditionally used and reliable technique to detect LINE-1 transcription which uses probes complementary to LINE-1 RNA (Deininger and Belancio, 2016). The probes can be designed to detect particular sub-types of LINE-1 elements, such as full-length LINE-1 elements or the discrimination between sense and antisense transcripts (Deininger and Belancio, 2016).

## **1.11 Concluding remarks**

While initially transposable elements were thought merely to be marks of parasitic infection, it is becoming increasingly clear that they play fundamental roles in mammalian

biology. LINE-1 elements, in particular, are a crucial part of the mammalian genome. Understanding the tight control of their expression patterns as well as their function in normal and diseased cells is a question of utmost importance in the fields of epigenetics, transposon biology and developmental biology.

The huge abundance of LINE-1 elements in the human and mouse genomes, repetitive nature and complex expression patterns requires the development of innovative technologies and the careful design of experimental procedures used to study these elements. With the advancement of these tools, we have learned of the multi-layered regulation of these elements at all layers of their expression and mobility, and we have discovered novel functions of transposable elements. Aside from their clear roles in shaping the landscape of the mammalian genome, it appears that cells have evolved ways to utilize LINE-1 expression and retrotransposition.

The mechanism and purpose of fine-tuned LINE-1 expression during embryogenesis has been a frequently asked question. Recently, researchers have made progress in answering such questions; LINE-1 elements appear not to be an unfortunate side-effect of epigenetic remodeling but are required for the normal development of the embryo (Jachowicz *et al.*, 2017). The implication of LINE-1 elements in such crucial processes along with the emerging theme of their context- and locus-dependent expression (He *et al.*, 2019) patterns are likely to have a huge impact on the future of the field.

### **1.12 Objective of the current project**

Retrotransposons belong to a class of mobile genetic elements that comprise 43% of the human genome (Lander *et al.*, 2001). Long interspersed elements type 1 (LINE-1s) are the

most abundant retrotransposon, accounting for 17% of the human genome. The human genome is impacted by retrotransposons in multiple ways (Goodier and Kazazian, 2008). Insertional mutagenesis is the most noticeable form of alteration and has been observed in more than 100 cases of human diseases, including cancer and birth defects (Hancks and Kazazian, 2016). The majority of LINE-1s became immobile during the course of evolution. Although LINE-1s more retrotranspose in the germline cycle, some retrotransposons may also be active in somatic tissues (Belancio et al. 2010; Ergun et al. 2004). It is estimated that among the total number of 500,000 LINE-1s presents in the human genome, 80-100 LINE-1s remain competent for retrotransposition (Brouha *et al.*, 2003). LINE-1 insertions can also impact the genome by altering gene expression. A full-length LINE-1 is typically 6-7 kb and has its promoter located in the 5' untranslated region (UTR) (Swergold, 1990). Retrotransposition is only possible when LINE-1 is expressed, and because LINE-1 transcription is regulated by its 5'UTR promoter, it is essential to understand the spatiotemporal control of LINE-1 promoter activity. Besides the small set of retrotransposition-competent LINE-1s, an additional set of 7000 immobile LINE-1s still carry active promoters, which are capable of producing transcripts (Khan *et al.*, 2006). The vast number of intact LINE-1 promoters when active may control the expression of protein-coding genes and also can produce chimeric transcripts that might lead to pathogenic conditions, like cancer. It has also been proposed that activated LINE-1 promoters may trigger initiation of cancer through epigenetic changes (Wilkins, 2010). Nevertheless, till today no attempt has been made to profile LINE-1 promoter activities in a locus dependent manner *in vivo*. Therefore, this project will employ transgenic mouse model to profile

locus-dependent LINE-1 promoter activities in various somatic tissues as well as in gonads throughout different developmental time points.

It is extremely technically challenging to monitor transcriptional activities of individual endogenous LINE-1s due to high sequence homology. Toward this goal, we generated single-copy 5'UTR-LacZ and 5'UTR-LacG mouse models. Both LacZ and LacG encode functional  $\beta$ -galactosidase, which can be visualized by X-gal staining. The LacG reporter gene lacks CpG dinucleotides, preventing transcriptional silencing via DNA methylation of the transgene body. After the mapping of the transgene locus in each line, different tissues from these single-copy transgenic mice were stained with X-gal to visualize LINE-1 promoter activity in these tissues. In addition to that, we checked the influence of different orientations, sense and antisense, of the same endogenous promoter in a specific locus, Rosa26. Also, we attempted to identify the brain cells holding the transgene expression.

We found that transgenic expression from two independent transgenic constructs varied significantly, with LacG lines having high expression compared to LacZ line. Besides interline, intraline variation was also observed in these two broad classifications. The kidney and thalamus of the brain were found to be a preferential hub of high promoter activity in most of the LacG lines. Also, an extreme contrast was observed between these two gene-targeted sense and antisense lines.

### 1.13 References

1. Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, and Moran JV. (2006) Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev.*; 20: 210-224.

2. An W, Han JS, Wheelan SJ, Davis ES, Coombes CE, Ye P, Triplett C, Boeke JD. (2006) Active retrotransposition by a synthetic LINE-1 element in mice. *Proc. Natl. Acad. Sci. USA*; 103(49):18662-7.
3. Aporn Dewan C, Mutirangura A. (2011) Connection up-and down-regulation expression analysis of microarrays. *Asian Biomed.*; 5: 257–262.
4. Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, Jenuwein T, Xu G, Leonhardt H, Wolf V, et al. (2012) In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.*; 8: e1002750.
5. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, and Hannon GJ. (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*; 316: 744-747.
6. Arokium H, Kamata M, Kim S, Kim N, Liang M, Presson AP, and Chen IS. (2014) Deep sequencing reveals low incidence of endogenous LINE-1 retrotransposition in human induced pluripotent stem cells. *PLoS One*; 9: e108682.
7. Baillie JK, Barnett MW, Upton KR, et al. (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*; 479: 534–537.
8. Bannister AJ, and Kouzarides T. (2011) Regulation of chromatin by histone modifications. *Cell Res.*; 381-395.
9. Beck CR, Garcia-Perez JL, Badge RM, and Moran JV. (2011) LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.*; 12: 187-215.
10. Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P. (2010) Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.*; 38:3909–3922.
11. Belancio VP, Hedges DJ, and Deininger P. (2008) Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res.*; 18: 343-358.
12. Bodak M, Yu J, and Ciaudo C. (2014) Regulation of LINE-1 in mammals. *Biomol. Concepts*; 5: 409-428.
13. Boeke JD, Garfinkel DJ, Styles CA, Fink GR. (1985) Ty elements transpose through an RNA intermediate. *Cell*; 40: 491-500.
14. Bojang P, and Ramos KS. (2016) Analysis of LINE-1 Retrotransposition at the Single Nucleus Level. *J. Vis. Exp.*;(110). doi: 10.3791/53753
15. Briggs EM, Ha S, Mita P, Brittingham G, Sciamanna I, Spadafora C, and Logan SK. (2018) Long interspersed nuclear element-1 expression and retrotransposition in prostate cancer cells. *Mob. DNA*; 9:1.
16. Brouha B, Meischl C, Ostertag E, de Boer M, Zhang Y, Neijens H, Roos D, Kazazian HH Jr. 2002 Evidence consistent with human LINE-1 retrotransposition in maternal meiosis I. *Am. J. Hum. Genet.*; 71: 327–336.
17. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. (2003) Hot LINE-1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. USA*; 100(9):5280–5285.
18. Bundo M, Toyoshima M, Okada Y, Akamatsu W, Ueda J, Nemoto-Miyauchi T, Sunaga F, Toritsuka M, Ikawa D, Kakita A, et al. (2014) Increased LINE-1 retrotransposition in the neuronal genome in schizophrenia. *Neuron*; 81: 306-313.
19. Burns KH. (2017) Transposable elements in cancer. *Nat. Rev. Cancer*; 17: 415-424.
20. Carmell MA, Girard A, Van K HJ, Bourc'his D, Bestor TH, de Rooij DG, and Hannon GJ. (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell*; 12: 503-514.
21. Carreira PE, Richardson SR, and Faulkner GJ. (2014) LINE-1 retrotransposons, cancer stem cells and oncogenesis; *FEBS J.*; 281(1):63-73.
22. Chishima T, Iwakiri J, and Hamada M. (2018) Identification of Transposable Elements Contributing to Tissue-Specific Expression of Long Non-Coding RNAs. *Genes (Basel)*; 9(1): 23.

23. Cook PR, and Tabor GT. (2016) Deciphering fact from artifact when using reporter assays to investigate the roles of host factors on LINE-1 retrotransposition. *Mob DNA*; 7:23.
24. Coufal NG, Garcia-Perez JL, Peng GE, et al. (2009) LINE-1 retrotransposition in human neural progenitor cells *Nature*; 460: 1127–1131.
25. Coufal NG, Garcia-Perez JL, Peng GE, Marchetto MC, Muotri AR, Mu Y, Carson CT, Macia A, Moran JV, and Gage FH. (2011) Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (LINE-1) retrotransposition in human neural stem cells. *Proc. Natl. Acad. Sci. USA*; 108: 20382-20387.
26. Criscione SW, Theodosakis N, Micevic G, Cornish TC, Burns KH, Neretti N, and Rodic N. (2016) Genome-wide characterization of human LINE-1 antisense promoter-driven transcripts. *BMC Genomics*; 17: 463.
27. Cruickshanks HA, and Tufarelli C. (2009) Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. *Genomics*; 94: 397-406.
28. DeBerardinis RJ, Kazazian HH. (1999) Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics*; 56: 317–323.
29. Deininger P, and Belancio VP. (2016) Detection of LINE-1 RNAs by Northern Blot. *Methods Mol. Biol.*; 1400: 223-236.
30. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, and Kazazian HH. (1991) Isolation of an active human transposable element. *Science*; 254: 1805-1808.
31. Doucet-O'Hare TT, Rodic N, Sharma R, Darbari I, Abril G, Choi JA, Young Ahn J, Cheng Y, Anders RA, Burns KH, et al. (2015) LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc. Natl. Acad. Sci. USA*; 112: E4894-4900.
32. Edwards JR, Yarychivska O, Boulard M, and Bestor TH. (2017) DNA methylation and DNA methyltransferases. *Epigenetics & Chromatin*; 10: 23.
33. Elbarbary RA, Lucas BA, and Maquat LE. (2016) Retrotransposons as regulators of gene expression. *Science*; 351: aac7247.
34. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE. (2012) The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucl. Acids Res.*; 40(1): D881–D886.
35. Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H et al. (2004) Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J. Biol. Chem.*; 279:27753–27763.
36. Erwin JA, Paquola AC, Singer T, et al. (2016) L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nature Neuroscience*; 19: 1583–1591.
37. Evrony GD, Cai X, Lee E, et al. (2012) Single-neuron sequencing analysis of LINE-1 retrotransposition and somatic mutation in the human brain. *Cell*; 151: 483–496.
38. Evrony GD, Lee E, Mehta BK, et al. (2015) Cell lineage analysis in human brain using endogenous retroelements. *Neuron*; 85: 49–59.
39. Evrony GD, Lee E, Park PJ and Walsh CA. (2016) Resolving rates of mutation in the brain using single-neuron genomics. *eLife* ;5. pii: e12966. doi: 10.7554/eLife.12966
40. Ewen KA, Koopman P. (2010) Mouse germ cell development: from specification to sex determination. *Mol. Cell Endocrinol.*; 323(1):76-93.
41. Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahashi H, Gorab E, Carninci P, and Torres-Padilla ME. (2013) Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.*; 20: 332-338.
42. Faulkner GJ, and Billon V. (2018) LINE-1 retrotransposition in the soma: a field jumping ahead. *Mob. DNA*; 9: 22.
43. Faulkner GJ, and Garcia-Perez JL. (2017) LINE-1 Mosaicism in Mammals: Extent, Effects, and Evolution. *Trends Genet.*; 33: 802-816.



44. Feng Q, Moran JV, Kazazian HH Jr, and Boeke JD. (1996) Human LINE-1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*; 87: 905-916.
45. Ficz G, Hore T, Santos F, Lee H, Dean W, Arand J, Krueger F, Oxley D, Paul YL, Walter J, et al. (2013) FGF signaling inhibition in escs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency cell. *Stem Cell*; 351-359.
46. Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gage FH, O'Shea KS, and Moran JV. (2007) LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.*; 16: 1569-1577.
47. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, Hammer GD, Collins KL, O'Shea KS, Menendez P, et al. (2010) Epigenetic silencing of engineered LINE-1 retrotransposition events in human embryonic carcinoma cells. *Nature*; 466: 769-773.
48. Gilbert N, Lutz-Prigge S, & Moran JV. (2002) Genomic deletions created upon LINE-1 retrotransposition. *Cell*; 110: 315-325.
49. Goodier JL, Kazazian HH Jr. (2008) Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*; 135(1):23-35.
50. Goodier, JL. (2014) Retrotransposition in tumors and brains. *Mob DNA*; 5: 11.
51. Goodier JL, Ostertag EM, Du K, and Kazazian HH Jr. (2001) A novel active LINE-1 retrotransposon subfamily in the mouse. *Genome Res.*; 11: 1677-1685.
52. Hancks DC, Kazazian HH Jr. (2012) Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.*; 22(3): 191-203.
53. Hancks DC, Kazazian HH Jr. (2016) Roles for retrotransposon insertions in human disease. *Mobile DNA*; 7:9.
54. He J, Fu X, Zhang M, He F, Li W, Abdul MM, Zhou J, Sun L, Chang C, Li Y, et al. (2019) Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat. Commun.*; 10: 34.
55. Jachowicz JW, and Torres-Padilla, ME. (2016) LINEs in mice: features, families, and potential roles in early development. *Chromosoma.*; 125: 29-39.
56. Jachowicz JW, Bing X, Pontabry J, Boskovic A, Rando OJ, and Torres-Padilla ME. (2017) LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.*; 49: 1502-1510.
57. Jones PA, and Takai D. (2001) The role of DNA methylation in mammalian epigenetics. *Science*; 293: 1068-1070.
58. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, and Kazazian HH, Jr. (2009) LINE-1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.*; 23: 1303-1312.
59. Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M, et al. (2011) DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs cell. *Stem Cell*; 8: 676-687.
60. Kazazian, HH Jr, Wong C, Yousoufian H, Scott AF, Phillips DG, and Antonarakis SE. (1988) Haemophilia A resulting from de novo insertion of LINE-1 sequences represents a novel mechanism for mutation in man. *Nature*; 332: 164-166.
61. Kerachian MA, and Kerachian M. (2019) Long interspersed nucleotide element-1 (LINE-1) methylation in colorectal cancer. *Clin. Chim. Acta*; 488: 209-214.
62. Khan H Smit A, Boissinot S. (2006) Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Research*; 16:78-87.
63. Klose RJ, and Bird AP. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*; 31: 89-97.

64. Kolosha VO, and Martin SL. (2003) High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *J. Biol. Chem.*; 278: 8112-8117.
65. Kopera HC, Larson PA, Moldovan JB, Richardson SR, Liu Y, and Moran JV. (2016) LINE-1 Cultured Cell Retrotransposition Assay Methods. *Mol. Biol.*; 1400: 139-156.
66. Kubo S, Seleme MC, Soifer HS, Perez JL, Moran JV, Kazazian HH Jr, and Kasahara N. (2006) LINE-1 retrotransposition in nondividing and primary human somatic cells. *Proc. Natl. Acad. Sci. USA*; 103: 8036-8041.
67. Kulpa DA, and Moran JV. (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.*; 14: 3237-3248.
68. Kulpa DA, and Moran JV. (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.*; 13: 655-660.
69. Kuwabara T, Hsieh J, Muotri A, et al. (2009) Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nature Neuroscience*; 12: 1097-1105.
70. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J et al. (2001) Initial sequencing and analysis of the human genome. *Nature*; 409:860-921.
71. Lanikova L, Kucerova J, Indrak K, Divoka M, Issa JP, Papayannopoulou T, Prchal JT, and Divoky, V. (2013) beta-Thalassemia due to intronic LINE-1 insertion in the beta-globin gene (HBB): molecular mechanisms underlying reduced transcript levels of the beta-globin (LINE-1) allele. *Hum. Mutat.*; 34: 1361-1365.
72. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ, Lohr JG, Harris CC, Ding L, Wilson RK, et al. (2012) Landscape of somatic retrotransposition in human cancers. *Science*; 337: 967-971.
73. Liu S, Du T, Liu Z, Shen Y, Xiu J, and Xu Q. (2016) Inverse changes in LINE-1 retrotransposons between blood and brain in major depressive disorder. *Sci. Rep.*; 6: 37530.
74. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*; 72: 595-605.
75. Martin SL. (2010) Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA Biol.*; 7: 706-711.
76. Martin SL, and Branciforte D. (1993) Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. *Mol. Cell Biol.*; 13: 5383-5392.
77. Martin SL, and Bushman FD. (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell Biol.*; 21: 467-475.
78. Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, and Gabriel A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*; 254: 1808-1810.
79. Mazzi EA, and Soliman KF. (2012) Basic concepts of epigenetics: impact of environmental signals on gene expression. *Epigenetics*; 7: 119-130.
80. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*; 454: 766-770.
81. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, and Nakamura Y. (1992) Disruption of the APC gene by a retrotransposal insertion of LINE-1 sequence in a colon cancer. *Cancer Res.*; 52: 643-645.
82. Minakami R, Kurose K, Etoh K, Furuhashi Y, Hattori M, and Sakaki Y. (1992) Identification of an internal cis-element essential for the human LINE-1 transcription and a nuclear factor(s) binding to the element. *Nucleic Acids Res.*; 20: 3139-3145.
83. Monot C, Kuciak M, Viollet S, Mir AA, Gabus C, Darlix JL, and Cristofari G. (2013) The specificity and flexibility of LINE-1 reverse transcription priming at imperfect T-tracts. *PLoS Genet*; 9: e1003499.

84. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, and Kazazian HH, Jr. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*; 87, 917-927.
85. Muotri AR, Marchetto MC, Coufal NG, et al. (2010) LINE-1 retrotransposition in neurons is modulated by MeCP2. *Nature*; 468: 443–446.
86. Muotri AR. (2016) LINE-1 retrotransposition in neural progenitor cells. *Methods Mol. Biol.*; 1400: 157-163.
87. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, and Gage FH. (2005) Somatic mosaicism in neuronal precursor cells mediated by LINE-1 retrotransposition. *Nature*; 435: 903-910.
88. Ostertag EM, and Kazazian HH, Jr. (2001) Biology of mammalian LINE-1 retrotransposons. *Annu Rev. Genet*; 35: 501-538.
89. Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, and Kazazian HH, Jr. (2002) A mouse model of human LINE-1 retrotransposition. *Nat. Genet*; 32: 655-660.
90. Ostertag EM, Prak ET, DeBerardinis RJ, Moran JV, and Kazazian HH, Jr. (2000) Determination of LINE-1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.*; 28: 1418-1423.
91. Packer AI, Manova K, and Bachvarova RF. (1993) A discrete LINE-1 transcript in mouse blastocysts. *Dev Biol.*; 157: 281-283.
92. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, and Knowles BB. (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell*; 7: 597-606.
93. Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, Biechele S, Huang B, Shen X, Ramalho-Santos M. (2018) A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*; 174(2):391-405e19.
94. Philippe C, Vargas-Landin DB, Doucet AJ, van Essen D, Vera-Otarola J, Kuciak M, Corbin A, Nigumann P, and Cristofari G. (2016) Activation of individual LINE-1 retrotransposon instances is restricted to cell-type dependent permissive loci. *eLife* ;5. pii: e13926. doi: 10.7554/eLife.13926
95. Rangwala SH, and Kazazian HH, Jr. (2009) The LINE-1 retrotransposition assay: a retrospective and toolkit. *Methods*; 49, 219-226.
96. Richardson SR, Gerdes P, Gerhardt DJ, Sanchez-Luque FJ, Bodea GO, Muñoz-Lopez M, Jesuadian JS, Kempen MHC, Carreira PE, Jeddelloh JA, Garcia-Perez JL, Kazazian HH Jr, Ewing AD, Faulkner GJ. (2017) Heritable LINE-1 retrotransposition in the mouse primordial germline and early embryo. *Genome Res*; 27(8): 1395–1405.
97. Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Perez JL, and Moran, JV. (2015) The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiol. Spectr.*; 3(2):MDNA3-0061-2014.
98. Richardson SR, Morell S, and Faulkner GJ. (2014) LINE-1 retrotransposons and somatic mosaicism in the brain. *Annu. Rev. Genet*; 48: 1-27.
99. Rodic N, Sharma R, Zampella J, Dai L, Taylor MS, Hruban RH, Iacobuzio-Donahue CA, Maitra A, Torbenson MS, Goggins M, et al. (2014) Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.*; 184: 1280-1286.
100. Saha PS and An W. (2019) Recently Mobilised Transposons in the Human Genome In: eLS; John Wiley & Sons, Ltd: Chichester.
101. Sanchez-Luque FJ, Richardson SR, and Faulkner GJ. (2016) Retrotransposon Capture Sequencing (RC-Seq): A targeted, high-throughput approach to resolve somatic LINE-1 retrotransposition in humans. *Methods Mol. Biol.*; 1400: 47-77.
102. Schulz WA. (2006) LINE-1 retrotransposons in human cancers. *J Biomed Biotechnol*; 2006: 83672.

103. Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP, Cooley T, Heath P, Smith KD, and Margolet L. (1987) Origin of the human LINE-1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics*; 1: 113-125.
104. Scott EC, Gardner EJ, Masood A, Chuang NT, Vertino PM, and Devine SE. (2016) A hot LINE-1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res*; 26: 745-755.
105. Seisenberger S, Peat JR, Hore TA, Santos F, Dean W, and Reik W. (2013) Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos Trans R Soc Lond B Biol Sci*; 368(1609):20110330.
106. Severynse DM, Hutchison CA 3rd, Edgell MH. (1992) Identification of transcriptional regulatory activity within the 5' A-typemonomer sequence of the mouse LINE-1 retroposon. *Mamm. Genome.*; 2(1): 41-50.
107. Sharma R, Rodic N, Burns KH, and Taylor MS. (2016) Immunodetection of Human LINE-1 Expression in Cultured Cells and Human Tissues. *Methods Mol. Biol.*; 1400: 261-280.
108. Solyom S, Ewing AD, Rahrman EP, Doucet T, Nelson HH, Burns MB, Harris RS, Sigmon DF, Casella A, Erlanger B, et al. (2012) Extensive somatic LINE-1 retrotransposition in colorectal tumors. *Genome Res.*; 22: 2328-2338.
109. Speek M. (2001) Antisense promoter of human LINE-1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell Biol.*; 21: 1973-1985.
110. Suarez NA, Macia A, and Muotri AR. (2018) LINE-1 retrotransposons in healthy and diseased human brain. *Dev. Neurobiol.*; 78: 434-455.
111. Swergold GD. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Molecular Cellular Biology*; 10(12):6718-6729.
112. Tan H, Wu C, and Jin L. (2018) A possible role for long interspersed nuclear elements-1 (line-1) in huntington's disease progression. *Med. Sci. Monit.*; 24: 3644-3652.
113. Trelogan SA, and Martin SL. (1995) Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc. Natl. Acad. Sci. USA*; 92:1520-1524.
114. Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR, et al. (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes. Cells*; 11: 805-814.
115. Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas CP, Zamora J, Raine K, et al. (2014) Mobile DNA in cancer Extensive transduction of nonrepetitive DNA mediated by LINE-1 retrotransposition in cancer genomes. *Science*; 345: 1251343.
116. Upton KR, Gerhardt DJ, Jesuadian JS, et al. (2015) Ubiquitous LINE-1 mosaicism in hippocampal neurons. *Cell*; 161: 228-239.
117. Vafadar-Isfahani N, Parr C, McMillan LE, Sanner J, Yeo Z, Saddington S, Peacock O, Cruickshanks HA, Meehan RR, Lund JN, et al. (2017) Decoupling of DNA methylation and activity of intergenic LINE-1 promoters in colorectal cancer. *Epigenetics*; 12: 465-475.
118. Van den Hurk JA, Meij IC, Seleme MC, Kano H, Nikopoulos K, Hoefsloot LH, Sistermans EA, de Wijs IJ, Mukhopadhyay A, Plomp AS, et al. (2007) LINE-1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.*; 16: 1587-1592.
119. Veselovska L, Smallwood SA, Saadeh H, Stewart KR, Krueger F, Maupetit-Mehouas S, Arnaud P, Tomizawa S, Andrews S, and Kelsey G. (2015) Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol.*; 16: 209.
120. Viollet S, Doucet AJ, and Cristofari G. (2016) Biochemical approaches to study line-1 reverse transcriptase activity in vitro. *Methods Mol. Biol.*; 1400: 357-376.

121. Walsh CP, Chaillet JR, and Bestor TH. (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.*; 20: 116-117.
122. Walter M, Teissandier A, Perez-Palacios R, and Bourc'his D. (2016) An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* ;5. pii: e11418. doi: 10.7554/eLife.11418
123. Wang GS, Wang MW, Wu BY, Yang XY, Wang WH, and You WD. (2008) LINE-1 family member GCRG123 gene is up-regulated in human gastric signet-ring cell carcinoma. *World J. Gastroenterol.*; 14: 758-763.
124. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*; 453: 539-543.
125. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*; 420: 520-562.
126. Weber B, Kimhi S, Howard G, Eden A, and Lyko F. (2010) Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. *Oncogene*; 29: 5775-5784.
127. Wilkins AS. (2010) The enemy within: an epigenetic role of retrotransposons in cancer initiation *Bioessays*; 32(10):856–865.
128. Xing J, Witherspoon DJ, and Jorde LB. (2013) Mobile element biology: new possibilities with high-throughput sequencing. *Trends Genet.*; 29: 280-289.
129. Xiong Y, and Eickbush TH. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *Embo. J.*; 9: 3353-3362.
130. Zimmerman DW. (1997) A Note on Interpretation of the Paired-Samples T-Test. *Journal of Educational and Behavioral Statistics*; 22(3): 349–360.

## Chapter 2

### Locus-specific LINE-1 promoter activity

#### 2.1 Abstract

LINE-1s are tightly regulated with a different host-defense mechanism in a mammalian organ system. It is always intriguing how the LINE-1 promoters are regulated in a wide range of mammalian organs, how active they are through different phases of development, and what makes them mobilize smoothly or even repressed. It is extremely challenging to address these question with alive or post-mortem human samples. Therefore, to learn more about the regulation of the promoter activity, we generated two main transgenic mouse lines: LacZ and LacG, and also generated their sublimes by mobilizing the transgene to different loci. In combination with an *in situ* histochemical detection technique (X-gal staining) aided by an automatic signal detection technique, we screened the main organs from the animals belonging to three different developmental time points. We found that LacZ sublimes had very fewer signals in comparison to the LacG sublimes. LacG071, a subline of LacG line, showed higher promoter activity in many organs than other sublimes. A closer analysis of the expression pattern of the LacG sublimes further elucidated a locus-specific, organ-dependent, and age-linked regulation of LINE-1 promoter activity. Furthermore, kidney and thalamus were especially found to be as the most preferential organ and brain region for high promoter activity, respectively.

## 2.2 Introduction

Our understanding of LINE-1 regulation in our genome is limited from the organ- and locus-specific determinants. The extent of LINE-1 insertional polymorphism and the abundance of sequence identity in the mammalian genome put a challenge for studying these elements. Based on the recent evidence at both the RNA level (Philippe *et al.*, 2016; Deininger *et al.*, 2017) and reinsertion (Tubio *et al.*, 2014), it is clear that only a few numbers of loci are retrotransposition competent. Therefore, it is high time to quantitate the locus-specific LINE-1 activity. Many studies adopted qPCR assay due to the ubiquity of the LINE-1 RNA. However, owing to the repetitive nature of LINE-1 sequences, it is hard to obtain an accurate, quantitative data based on LINE-1 transcripts. Furthermore, it is hard for qPCR assay to distinguish between a handful of the retrotransposition-competent transcripts in a plethora of non-coding RNAs and truncated LINE-1 transcripts. This problem, however, can be addressed with implementing a transgenic approach.

Until today, none of the approaches could quantify locus-dependent LINE-1 expression *in situ* in transgenic mouse models. Therefore, in this study, we examined the role of chromosomal location on the expression of endogenous LINE-1 promoter fused with reporter construct in a wide range of somatic as well as germline organs. To achieve that we generated single-copy transgenic mice by pronuclear injection, where transgenic constructs were mobilized into unique random chromosomal sites. The *in vivo* expression from some specific LINE-1 loci was assayed by histochemical staining for beta-galactosidase activity in a wide range of organs in different mouse developmental time points.

## 2.3 Materials and Methods

### Ethics statement

The study was carried out under the strict accordance with the guidelines provided by the National Institutes of Health in the Guide for the Care and Use of Laboratory Animals. The protocols were thoroughly approved by the Institutional Animal Care and Use Committees (IACUC).

### Plasmids Construction Details

Plasmid pWA370 contains the 5'UTR-LacZ transgene. It is constructed via a three-way ligation of the BglII/NotI fragment from pMD002, the NotI/NcoI fragment from pWA125 (Newkirk et al, 2017), and the NcoI/BamHI fragment from pQUEST-nucLacZ (a gift from Liqun Luo; Addgene plasmid # 24356; <http://n2t.net/addgene:24356>; RRID:Addgene 24356) (Potter et al, 2010).

Plasmid pWA371 contains the 5'UTR-LacG transgene. It is constructed via a three-way ligation of the BglII/NotI fragment from pMD002, the NotI/NcoI fragment from pWA125 (Newkirk et al, 2017), and the NcoI/BglII fragment from pAAVf-EnhCB-lacZnls (a gift from Phillip Zamore; Addgene plasmid # 35642; RRID:Addgene 35642).

### Transfection

The MW ratio between pWA370 (5930151.43 Da) and pWA371 (5324719.19 Da) is 1.1 (pWA370/pWA371). To be fair for expression comparison, we adjusted the plasmids amount to achieve an equal copy number of the plasmids. 880ng pWA370 was mixed with 2.2ul P3000 reagent (Lipofectamine 3000 kit) in 50ul Opti-MEM medium. Similarly, 800ng PWA371 mix with 2.2ul P3000 reagent (Lipofectamine 3000 kit) in 50ul Opti-MEM



medium. Vortexed shortly two vials of DNA/P3000 mixture with 3.3ul L3000 (Lipofectamine 3000 kit, premix with another 50ul Opti-MEM medium) respectively. For both plasmids, incubated for 15 minutes under room temperature and aliquots into 3 different wells following titration manner (60ul, 30ul, 10ul) in 24-well plate. Then, added 500ul 3T3 cell suspension ( $1.2 \times 10^5$ /ml) into corresponding wells which had transfection complex and gently shook the plate to evenly seed cells. We also included a GFP control plasmids which had similar size and show good transfection efficiency. Within 3 different plasmids amount (300ng, 150ng, 50ng), 50 ng was the most the optimal condition which clearly showed pWA371 (GCless) had significantly higher expression profile, while higher plasmids amount might saturate the expression and could not be differentiated obviously.

The transfected cells rinsed with 1XPBS and were fixed for 5 minutes in 4C, which was followed by 2X washing in again 1XPBS at room temperature. Overnight staining was carried out in staining solution. Fixative contained 2% formaldehyde and 0.2% glutaraldehyde. The staining solution contained 5mM K-ferricyanide, 5mM K-ferrocyanide, 2mM MgCl<sub>2</sub>, and 1mg/ml X-gal.

### **Mice**

WT and transgenic mouse lines were maintained in the C57BL/6J (B6) background. Mice were housed in well-ventilated cages. In maximum, up to five adult mice per cage was allowed. They fed on quality-controlled standard pellet chow and pure water *ad libitum* in a regular 12-hour light/dark cycle at room temperature. Wild type variants or C57BL/6J (B6) were initially purchased from Jackson Laboratory. H1t-SB100X mice were generated in C56BL/6N background at Max Delbrück Center for Molecular Medicine, Germany. LacG and LacZ are generated by traditional pronuclear microinjection protocol by Cyagen

Biosciences Inc., US. Rosa26 sense and Rosa26 antisense lines were produced with gene targeting (homologous recombination) method by Cyagen Biosciences Inc., US. Z/EG mice (Novak *et al.*, 2000) (JAX stock 3920) were procured from Jackson laboratory as well. Mov1011<sup>+/-</sup> mice (Zheng *et al.*, 2010) were gifted by P. Jeremy Wang, University of Pennsylvania, PA, through Phillip Zamore at University of Massachusetts Medical School, Worcester, MA. Mov1011<sup>+/-</sup> mice were crossed with transgene positive animals to generate Mov1011 knock out a mouse in two steps of breeding.

### Real-time PCR

SYBR-Green I master mix (Applied Biosystems) was used to perform Real-time PCR or quantitative PCR reactions in triplicate with 10ng cDNA template in 25ul of reaction volume.  $Q = E^{(\text{Min}(\text{mean Ct of all samples}) - \text{Ct})}$  formula was used to calculate relative quantity (Q) of a specific transcript. E and Ct represent mean PCR efficiency and threshold cycles, respectively. Mouse Gapdh gene was used as internal control, and also no template controls were included. Following primer pairs were used (**Table 2.1**).

Transgene	Primers
LacG	WA580
	WA581
LacZ	WA574
	WA575

**Table 2.1** The primer pairs used for the genotyping of the transgenic lines.

### Droplet digital PCR

Bio-Rad protocol was followed for running the Droplet Digital PCR (ddPCR) reactions, containing approximately 60ng of gDNA and fluorescence probe in reaction volumes. Each well of the reactions contained 60ng of gDNA. Droplet generation step was preceded by a step of DNA digestion with 10U/ul of *NcoI* (NEB) at 37C for 15 minutes. An eight-well

Bio-Rad DG8 droplet generator cassette was used for carrying out droplet generation step. Each well of this cassette was using 20ul of sample added with 40ul of droplet oil. The generated droplets underwent through a PCR cycling condition: 95C for 10min, (94C for 30sec, 60C for 1min) x 40 cycles, extension temperature of 98C for 10min. QX200 droplet reader was used to read the fluorescence signals from each droplet. Later, the result was analyzed with the help of Bio-Rad Quantasoft software version 1.3.2.0. The primers are listed below in Table 2.1.

### **Ligation mediated PCR (LM-PCR) & sequencing of the amplicon**

Steps in Ivics, et al. 2011, Nature Protocols was used to get the amplicons, using the PCR reaction protocol described in the table. The DNA sequence was (agarose 1%) get cut and purified, using QIAquick PCR Purification Kit (QIAGEN). The amplicons were Sanger Sequenced by Eurofins Genomics. Each sequencing result was visualized with FinchTV (version 1.4) software for Windows.

### **BLAST**

A basic local alignment search tool (BLAST) or BLAST-like alignment tool (BLAT) was used to search for the DNA sequence directly flanking the transposon, at the University of California, Santa Cruz (UCSC) Genome Bioinformatics website (<http://genome.ucsc.edu/cgi-bin/hgBlat>) or the National Center for Biotechnology Information (NCBI) website (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

### **Genotyping PCR & Locus Specific-PCR (LS-PCR)**

Genotyping of the mice was performed with gDNA mainly from their tail biopsies. A PCR reaction using ExTaq or ExTaq HS polymerase was run for the detection of the presence

or absence of expected transgene with the set of specific primers designed onto the respective transgene sequences (Table 2.2a).

Locus Specific-PCR (LS-PCR) or junction PCR uses the primers designed to amplify the location between the transgene and the genomic DNA. The primers were designed with MacVector software Table 2.2b. It used the following conditions and PCR protocols to get the expected band.

Line	Primer pairs	Expected band	PCR protocol													
LacZ	WA0570, WA0571	385 bp	Genotyping PCR <table border="1"> <thead> <tr> <th>C</th> <th>min/s</th> </tr> </thead> <tbody> <tr> <td><b>94</b></td> <td>3min</td> </tr> <tr> <td><b>94</b></td> <td>15sec</td> </tr> <tr> <td><b>57.5</b></td> <td>30sec</td> </tr> <tr> <td><b>72</b></td> <td>1 min</td> </tr> <tr> <td><b>72</b></td> <td>7min</td> </tr> </tbody> </table> 35cyl		C	min/s	<b>94</b>	3min	<b>94</b>	15sec	<b>57.5</b>	30sec	<b>72</b>	1 min	<b>72</b>	7min
C	min/s															
<b>94</b>	3min															
<b>94</b>	15sec															
<b>57.5</b>	30sec															
<b>72</b>	1 min															
<b>72</b>	7min															
LacG	WA0572, WA0573	311 bp														
H1T	WA0440, WA0441	200 bp														
Z/EG	WA0679, WA0680	173 bp														

**Table 2.2a** Desired length of the amplicon in the genotyping PCR.

Lines	Primer pairs to get the transgenic band	PCR protocol	Expected bands
LacG082	WA0984 - WA0598	LSPCR	510bp
LacG221	WA0988 - WA0598	LSPCR	500bp
LacG061	WA0980 - WA0598	LSPCR (with 60.5C annealing temperature)	1200bp
LacG071	WA0649 - WA0598	LSPCR touchdown	500bp
LacG141	WA1529 – WA0598	LSPCR	508bp

**Table 2.2b** The PCR protocols used and the expected bands for different sublines.

## PCR protocols

Types		Protocols		
<b>LSPCR PCR</b>				
Cycle #	Denature	Anneal	Extend	Hold
1	94°C, 5 min			
2-31	94°C, 1 min	55°C, 30s	72°C, 30s	
32			72°C, 7 min	
33				4°C
<b>LSPCR touchdown</b>				
Cycle #	Denature	Anneal	Extend	Hold
1	94°C, 5 min			
2-11	94°C, 1 min	65°C, -1°/cycle, 30s	72°C, 30s	
12-36	94°C, 1 min	55°C, 30s	72°C, 30s	
37			72°C, 7 min	
38				4°C
<b>LMPCR 1</b>				
Cycle #	Denature	Anneal	Extend	Hold
1	96°C, 2 min			
2-11	92°C, 40s	60°C, -1°/cycle, 40s	72°C, 2 min	
12-36	92°C, 40s	50°C, 40s	72°C, 1 min	
37			72°C, 10 min	
38				4°C
<b>LMPCR 2</b>				
Cycle #	Denature	Anneal	Extend	Hold
1	96°C, 2 min			
2-7	92°C, 40s	60°C, -1°/cycle, 40s	72°C, 1 min	
8-21	92°C, 40s	59°C, 40s	72°C, 1 min	
37			72°C, 10 min	
38				4°C

**Table 2.3** Essential PCR protocols (adopted from Ivics *et al.*, 2011).

## gDNA Isolation

In general, for genotyping, gDNA was isolated from tail biopsies, using the Genra PureGene Tissue Kit (Qiagen). Mouse tissues were collected from animals between 1 week

–18 months of age, and a section of all trimmed organs, from the mice of all ages, was always kept stored in RNAlater (Sigma) for methylation analysis or DNA needed for ddPCR. The gDNA from these tissues (weighing, 10mg) was extracted using either DNeasy Blood & Tissue Kit (Qiagen). In some cases, where extracting gDNA from the paraformaldehyde tissues was essential, after cell lysis step, the lyse was further treated in the high heat of 95C for 30 minutes to break the peptide cross-linkages before using it for further steps of Gentra PureGene Tissue Kit (Qiagen).

### **Tissue harvesting and preparation**

Animals were euthanized with isoflurane. Immediately after the respiratory arrest, the diaphragm was cut to expose the liver. Later, the diaphragm was first to cut laterally across and again cut on both ends of diaphragm across the ribs and towards the head to expose the heart. A needle of the winged infusion set (19 Ga) connected to a peristaltic pump with a tube inserted the catheter needle into the protrusion to extend up approximately 5mm inside the ascending aorta. The right ventricle of the heart was cut open with a sharp scissor for drainage of blood. First, 1× PBS flowed slowly but constantly until the liver got pale. Next, the fixative used to perfuse the heart. Neonatal animals were perfused with 18ml of either 1XPBS or 4% PFA without additives (duration: 3min; flow rate: 6 ml/min), and the adult and aged mice were perfused with 36ml of either 1XPBS or 4% PFA without additives (duration: 4min; flow rate: 9 ml/min).

Next, the organs were individually collected into cold 1XPBS for clearing of any blood. The organs, except the brain, were then trimmed according to the desired orientations (**Table 2.4**) mentioned in the work by Ruehl-Fehlert *et al.* in 2003 (Ruehl-Fehlert *et al.*, 2003). Next, the tissues were drop-fixed into 2% PFA with additives for neonatal organs

for 4 hours or into 4% PFA with additives for neonatal organs for 6 hours under constant agitation. Later, washed for 40min in 1XPBS for 2 times (each for 20min). Dropped into 15% sucrose solution at 4C until they dropped. When dropped again put into 30% sucrose solution until they dropped again. The organs were soaked for the water out and were equilibrated in OCT not more than 15 - 20 mins. The tissues were then rightly oriented, being in steel made tissue embedding molds, before snap freezing in liquid nitrogen. The snap freezing was carried on an iron/steel base immersed into liquid nitrogen and with the OCT molds placed onto the base. The OCT blocked were packed in zip bags individually and stored at -80C until use.

<b>Organ</b>	<b>Localization</b>	<b>Direction</b>
Brain	Sagittal	Longitudinal
Heart	Through ventricles and atria	Longitudinal
Lung	Left lobe	Longitudinal horizontal
Liver	Left lateral lobe	Transverse
Kidney	Through the tip of the papilla and renal pelvis	One-side longitudinal
Spleen	At largest extension	Transverse
Testis	Close to rete testis	Longitudinal

**Table 2.4** Organ-specific orientations used during the trimming of the tissues.

### **Cryosectioning and X-gal staining of fixed tissue**

Before beginning, it was made sure that the OCT blocks are at -20C at least 20-30 minutes before sectioning. Tissues were sectioned in a thickness of 20um for X-gal staining and 14um for immunostaining. The microscopic slides were dried in air for 15-20 minutes. The slides were washed in 1X PBS for 5min at 4C, incubated in 1% PFA solution for 5 minutes at 4C, and left in 1XPBS at 4C until next step. Washed with freshly prepared LacZ wash

solution for 15-20 minutes on a rocker. Incubated for overnight (14-16 hours) for all lines (except, Rosa26 antisense line and slides for subsequent immunostaining) or at 37C, being immersed into freshly prepared X-gal solution (adjusted to required pH) with a casual covering. Took the slides out of the incubator. Rinsed with water. Washed for 20 min in total in 1XPBS (2 times for each 10 min) under agitation. Post-fixed in 1% PFA for 5 mins in RT. Washed with 1XPBS for 10 mins. Rinsed with water. Counterstained with freshly prepared neutral red for 7 min with intermittent shaking. Rinsed with water. Dehydrated for 3 mins each in 70%, 90%, 100% ethanol. Dried shortly and add EcoMount and coverslip. Stored at room temperature or 4C in a slide holder box. Looked under the microscope.

#### **Preparation of 4% PFA fixative without additive (25ml)**

Added 1g of PFA to 15ml of autoclaved, stirring the water at 60C (waited for 15mins). Added 5ul of 5M NaOH to get it dissolved (waited 15 mins). Added 2.5ml of 10X PBS. Added autoclaved water to volume up to 25ml.

#### **Preparation of 1% PFA fixative with additive (100ml)**

1% PFA	25ml of 4% stock
2mM MgCl	200ul from 1M stock
5mM EGTA pH8.0	4.4ml from 0.1136M stock (solvent: water)
0.2% NP-40	2ml from 10% stock (solvent: water)
10X or 0.1M PBS	10ml
Autoclaved water	Rest of the volume

#### **Preparation of LacZ wash solution (100ml)**

2mM MgCl <sub>2</sub>	200ul from 1M stock (solvent: water)
0.01% sodium deoxycholate	100ul from 10% stock (solvent: water)
0.02% NP-40	200ul from 10% stock (solvent: water)
1XPBS	rest of the volume



**Preparation of LacZ stain solution (or Xgal solution) (10ml)**

Xgal stock	400ul of LacZ stock (25mg/ml)
5mM Potassium ferricyanide	16.5mg
5mM Potassium ferrocyanide	21.1mg
LacZ wash solution	9.6ml

**Detailed information regarding reagents used**

<b>Chemical</b>	<b>Manufacturer</b>	<b>Reference Code(s)</b>
X-gal	Cayman Chemical	Item# 16495; Batch# 0532100-17
PFA	Sigma-Aldrich	Ref.# P6148; Lot# MKCD5278
Neutral red	Sigma-Aldrich	Ref.# 72210; Lot: BCBP6989V
NP40 or IGEPAL	Sigma-Aldrich	Ref.# I3021; Lot:MKBC8185V
Sodium deoxycholate	ACROS	Ref.# 218590250 Lot#A0293327
EGTA	VWR	Ref.#0732; Lot: 18A3056246
Magnesium Chloride Hexahydrate (MgCl <sub>2</sub> )	EMD Chemicals	Ref.# 5980
Potassium ferrocyanide	Sigma-Aldrich	Ref.# P3289
Potassium ferricyanide	Sigma-Aldrich	Ref.# 244023
EcoMount	Biocare Medical	Ref.# EM897L; Lot: 020618
Microscope slides (Superfrost Plus) Size: 25x 75 x 1.0 mm	Fisher Scientific	Cat.# 12-550-15
Cover Glass	Fisher Scientific	Ref.# 12545F; Lot:18838

**Preparation of neutral red counterstain**

500mg of neutral red was stirred in 100 ml of deionized water for overnight.

Later 1 drop of acetic acid was added.

**IF staining followed by X-gal staining**

Fourteen um thickness of the brain samples were sectioned. Dried for 15-20 minutes in the air. Immersed in 1XPBS for 10 minutes. Fixed for 5 mins in 1% PFA. Stored in 1XPBS at 4C until further use. Washed in LacZ wash solution for 15 minutes. Stained with X-gal solution (25mg/ml; pH 7.7) for 5 hours at 37C. Wash with 1XPBS for 15-20 minutes under agitation. Antigen retrieved for 10 minutes at 90C in 1X Na-citrate buffer. Waited for the solution to cool down, and avoided to take the slides out when the solution is hot. Rinsed in water. Incubated for 1 hour in blocking buffer containing serum and 0.2% Tween-20, 1% BSA, 5% donkey serum in 1XTBS at RT. Overnight incubated with blocking buffer with a required ratio of primary antibodies in antibody dilution buffer, containing 0.2% Tween-20 and 1% BSA. Incubated at RT with 1xTBS with only 0.2% Tween-20 for 5 minutes for 3 times. Incubated with the secondary antibody in antibody dilution buffer for 2 hours at RT. Incubated at RT with 1xTBS with only 0.2% for 5 minutes for 3 times. Incubated at RT with DAPI in water for 5 minutes. Dried and added Prolong antifade mounting agent before putting on the coverslips.

<b>Primary Antibodies</b>		
Goat Polyclonal Anti-Lamin B (1:200)	Santa Cruz Biotechnology	Ref.# Sc-6217 Lot# J1311
<b>Secondary antibodies</b>		
4',6-diamidino-2-phenylindole (DAPI) (1:1000)	Thermo Scientific	Ref.: 62248
<b>Other reagents</b>		
Bovine serum albumin (heat-shock treated)	Fisher Bioreagents	CAS 9048-46-8
Donkey serum	Lampire Biological Laboratories	Cat.# 7332100 Lot: 13A29004
Molecular Biology grade water	Hyclone	Cat.# SH30538-02 Lot: AAC200214

### Bisulfite-sequencing analysis

Total gDNA was extracted from tissues of mice, using DNeasy Blood & Tissue Kit (Qiagen) manufacture's instruction. Bisulfite conversion of the gDNA was performed by using the EpiTect Plus DNA Bisulfite Kit (Qiagen). Nested PCR was set up using Ex-Taq Polymerase (Takara). MethPrimer (<http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>) was used to design the bisulfite PCR primers shown in **table 2.5** (Li and Dahiya, 2002). The amplicons were gel-purified using the QIAquick Gel Extraction Kit (Qiagen) followed by cloning into a TA vector (Stratagene). In a blue/white colony screening, the white bacterial colonies were chosen for Sanger Sequencing. To get the DNA methylation status, the sequence data were analyzed with the help of a quantification tool for methylation analysis, QUMA (Kumaki *et al.*, 2008) ([http://quma.cdb.riken.jp/top/quma\\_main\\_j.html](http://quma.cdb.riken.jp/top/quma_main_j.html)). QUMA gave methylation plots, which were next downloaded.

LacZ	WA1299, WA1300 (expected band at 526bp)
LacG	WA1301, WA1302 (expected band at 506bp)

**Table 2.5** The primer pairs used for bisulfite sequencing of the transgenic lines.

### Microscopy and image analysis

In general, Zeiss Axio Imager Upright microscope was used to take regular fluorescence and bright-field images of different magnification of X20 or X40. Aperio VERSA Bright field Fluorescence & FISH Digital Pathology Scanner (Leica, NJ) was used to scan the whole-slide bright-field images of the stained sections at maximum 20X resolution. These whole-slide images (extension .SCN) were navigated with the help of Aperio ImageScope

(version 12.4.0.5043), a pathology slide viewing software, from Leica Biosystem, Leica, NJ.

### **QuPath**

The signals from bright-field or immuno-stained images were quantified with QuPath (version 0.1.2) to quantify the percentage of positive cells with the required signals. The file extension of the original, whole-slide images (.SCN) was not compatible with the QuPath. Therefore, the images were converted to .TIF output (compression mode 'LZW') with the help of image extract option in Aperio ImageScope (version 12.4.0.5043). The interesting area(s) of the tissue sections were manually annotated to let QuPath perform quantification based on the manual or pre-written scripts with the necessary instructions.

### **QuPath: quantification of signals**

Two image types in the program QuPath were used for quantification: brightfield (H-DAB) for X-gal stained tissues and fluorescence for immunofluorescence tissues. For the X-gal stained tissues, a different script was created for each tissue based on its specific characteristics. Each script was modified using a certain number of channels. Each channel denoted a specific command to either denote a cell as either positive, negative, or border based on the intensity of the hematoxylin or DAB. The mean, sum, and max intensity of these two parameters were optimized over a range of several tissues to accurately detect an X-gal positive cell from a negative. A percentage was then derived by taking the number of positive cells divided by the number of negative cells plus positive cells and then multiplied by 100. The cells detected as border were determined to be falsely detected cells and were not included in this equation.

### **QuPath Data Plot**

The QuPath data were plotted using the ggplot2 package of the R software (version 3.6.2). For any tissues with no detected X-gal positive cells, we plotted 1/(total number of cells) since we could better see the data points (there would be too much clutter if we had plotted them all as 0's). To make it clear that these data were referred to tissues with no detected X-gal positive cells, we also created a binary column called Detected, which had a value of "Yes" or "No". To further reduce the clutter, we used the position jitter function so that there was less overlap between data points. We also used the log10 scale for the y-axis since the range of y values (percentage of X-gal positive cells) is very small (0 to about 10 percentage). Thus, using the log10 scale improved the visual by "stretching out" the y-axis and allowed us to better observe the entire range.

### **Statistics**

Statistical analysis and mathematical calculations were performed using either Microsoft excel. Sample means were compared with the help of two-tailed unpaired T-test used and expressed in terms of the *p*-value. Analysis of replication consistency was tested with Coefficient of variation (CV %) was used to analyze replicate consistency. Any CV values with <100% were regarded to have data with less variation.

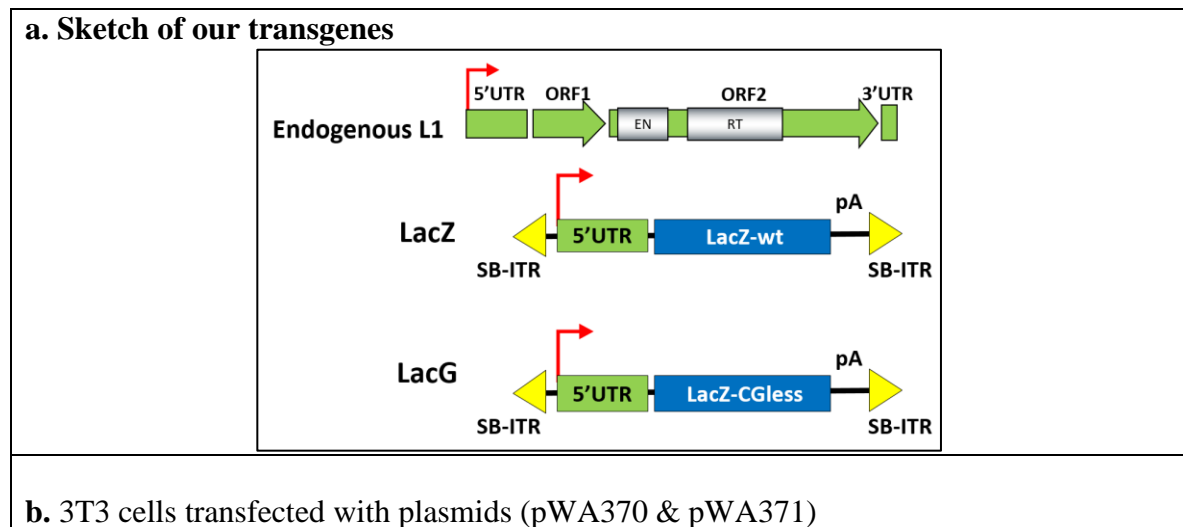
## **2.4 Results**

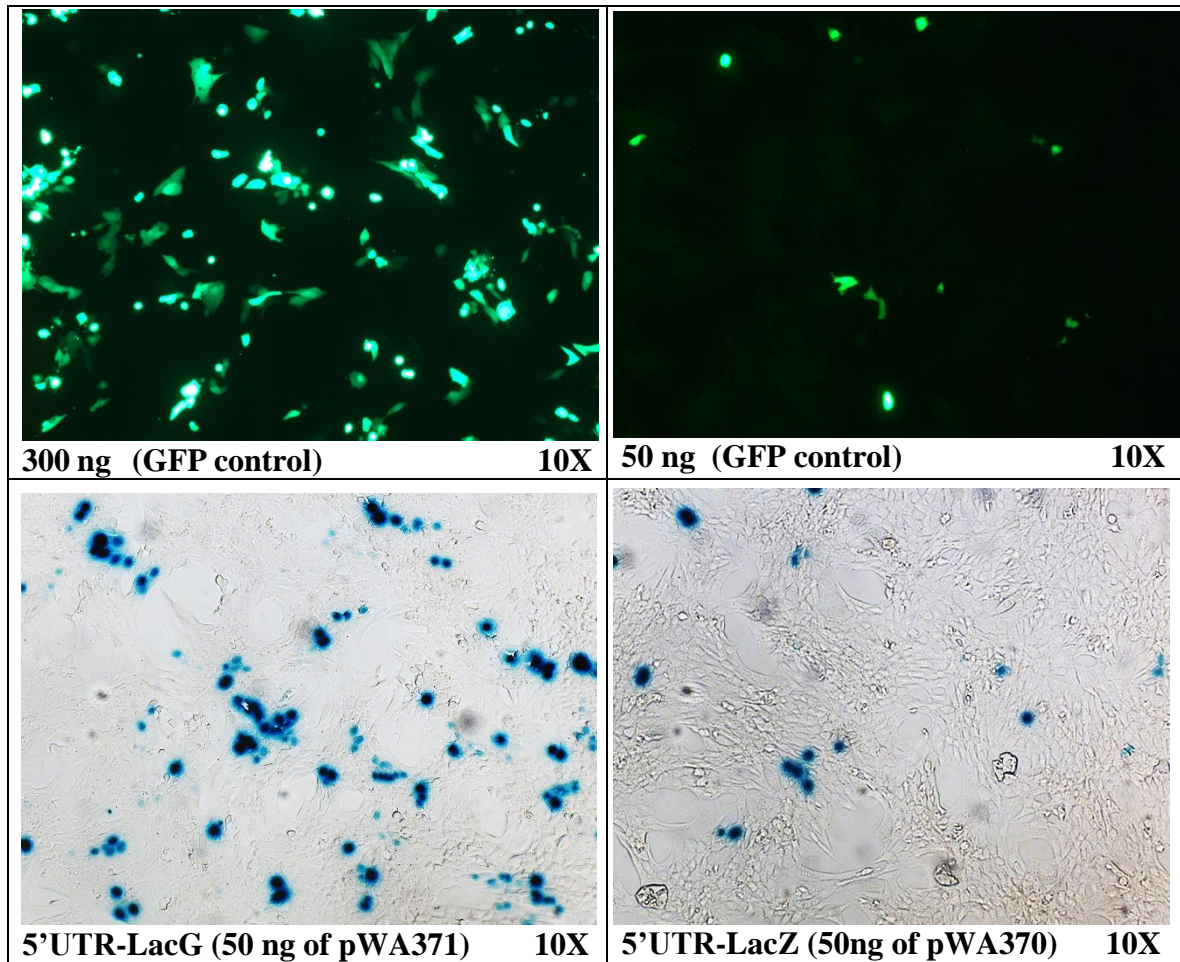
### **2.4.1 Development and preliminary testing of LacZ based reporter transgenes for L1 promoter activity**

To understand the *in vivo* promoter activity, the endogenous 5'-UTR promoter was fused with either of the two reporter genes (**Fig. 2.1a**). The *in vivo* expression from these two

transgenic constructs was analyzed in the transgenic mouse, with wild-type (C57BL/6) genetic background, carrying the same transgene in a unique chromatic location.

Before generating transgenic mouse lines, we analyzed *in vitro* the transient expression patterns of these two constructs in transfected mouse embryonic fibroblast cell line (3T3). Here, we found a significant difference of expression in these two transgenic constructs, in the presence of appropriate control with GFP plasmid (**Fig. 2.1b**). Transfection with a plasmid (pWA370), enclosing 5'UTR-LacZ transgene, generated about 2.5 fold less X-gal positive cells than that of its counterpart plasmid (pWA371), which carries 5'UTR-LacG transgene. *In vivo* differences of their expression patterns in locus-dependent as well as time-dependent manner were also determined in sections 4 & 5.

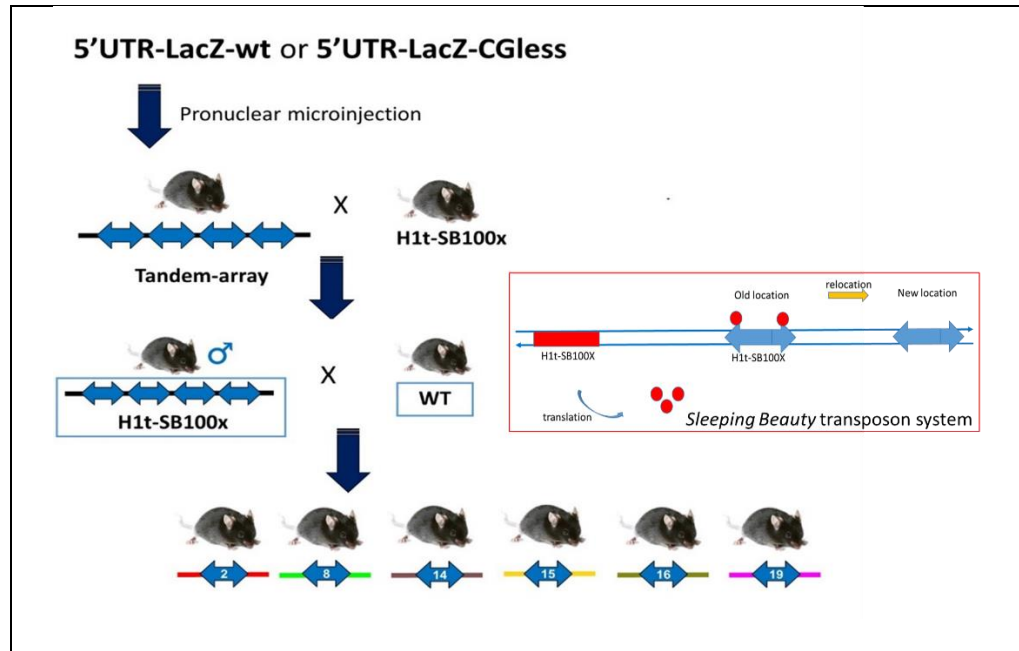




**Figure 2.1.** LacZ and LacG transgenes (a) sketch of the structure of LacZ and LacG transgenes & (b) The transfection results in the 3T3 cells with the plasmids carrying the transgenes.

#### 2.4.2 Generation of transgenic mouse lines each carrying a single-copy 5'UTR-LacG (or 5'UTR-LacZ) transgene at a random genomic locus

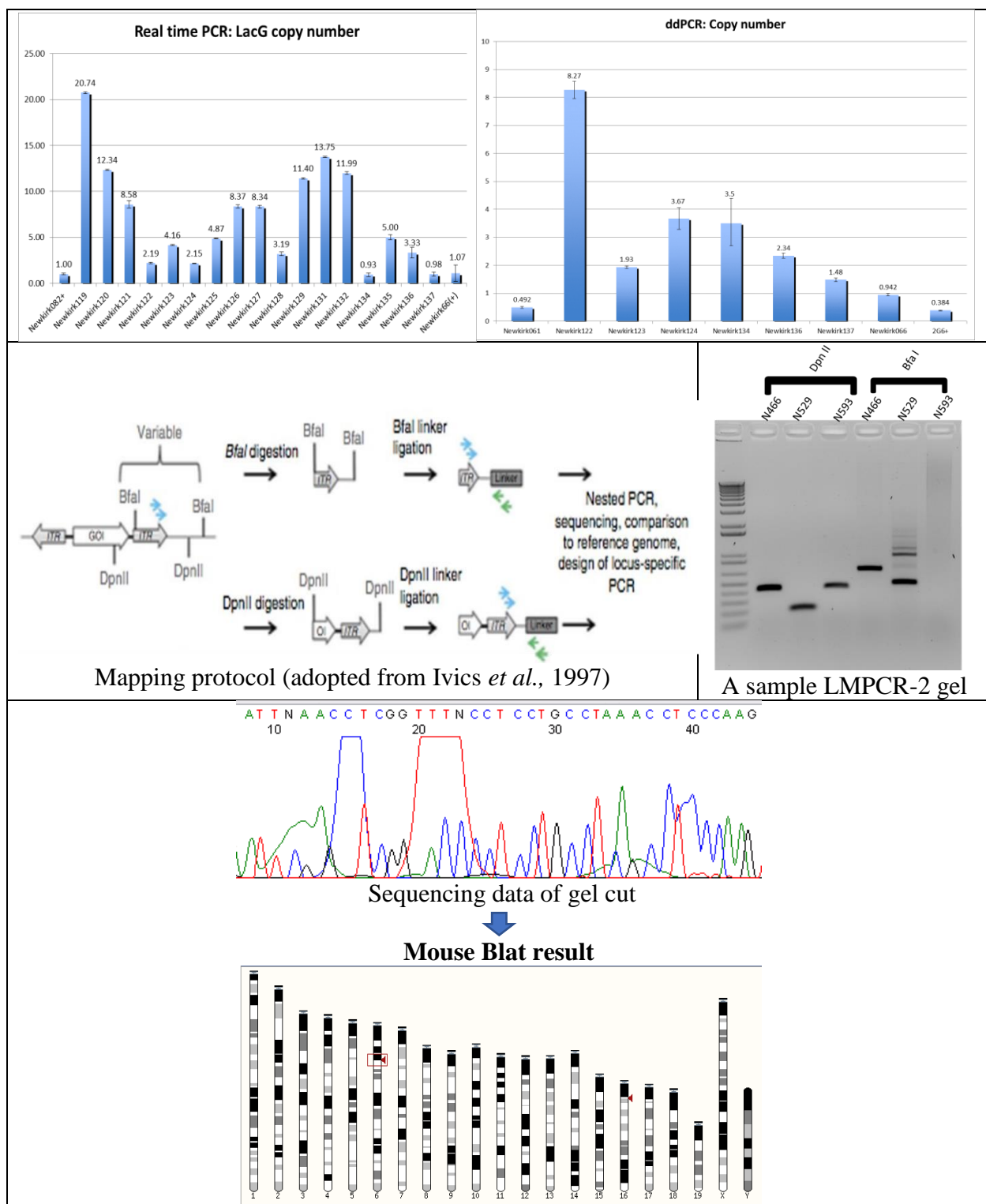
To obtain single-copy germline insertions for our transgene, the donor animal was bred with H1t-SB100X positive transgenic animals (Grandi *et al.*, 2015), expressing a hyperactive *Sleeping Beauty* (SB) transposase specifically in pachytene spermatocytes (Mates *et al.*, 2009). Therefore, the excision of the transgene and their successive mobilization happened in male germ cells, carrying both H1t-SB100X and the L1 transgene (Fig. 2.2).



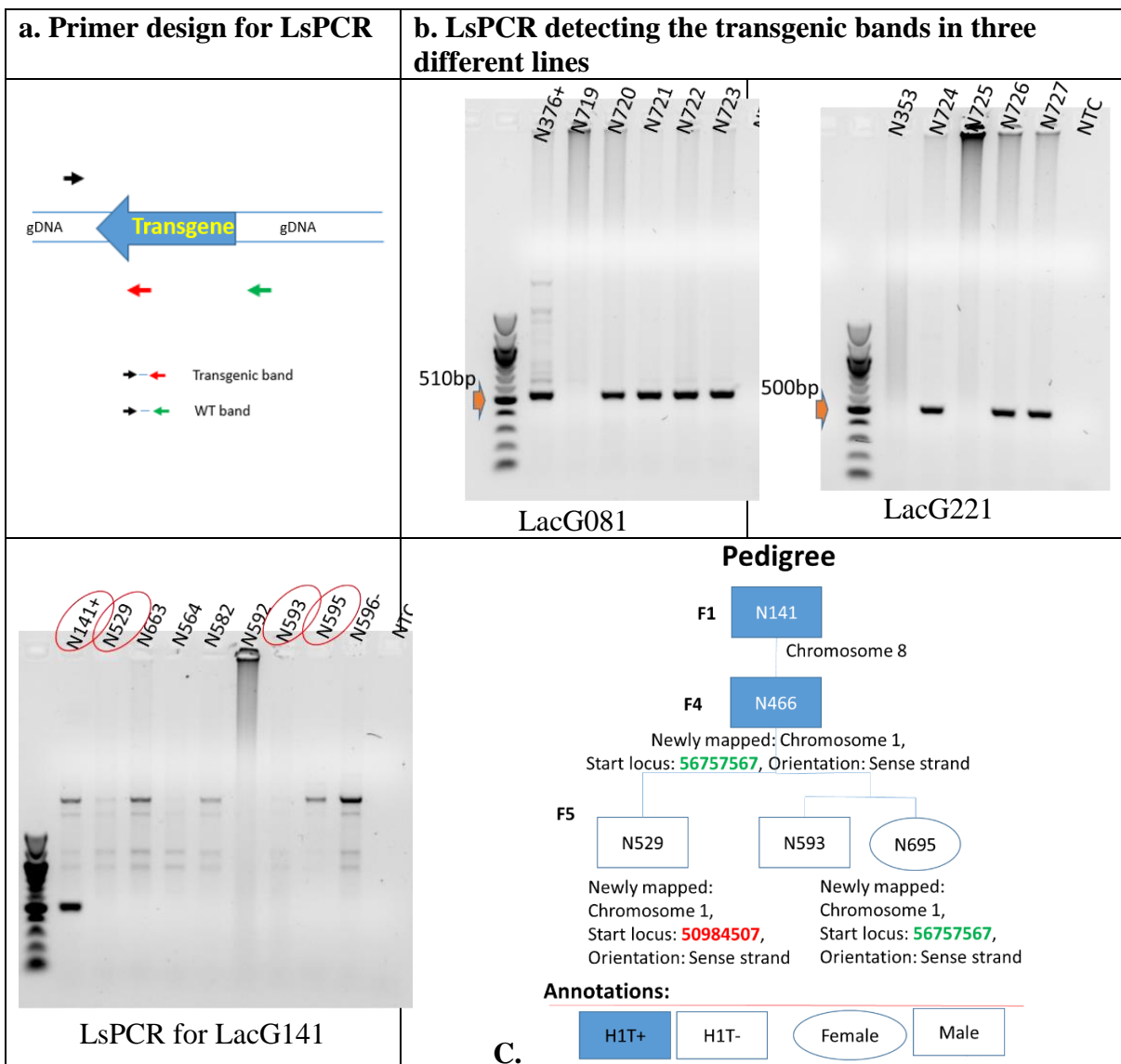
**Figure 2.2** Generation of single-copy mouse lines with the Sleeping Beauty DNA transposon system

The progenies of this male with a wild-type female possess differential copy numbers of transgenes. The transgenic mice with low-copy or single-copy number were identified, using real-time PCR and droplet digital PCR, respectively. Next, the loci of these transgenes were mapped (**Fig. 2.3**), and primers were designed for locus-specific PCR at the junction of the transgene and genomic DNA. On some occasions, when the transgenes moved further due to the presence of H1t-SB100X in the transgene positive animals, the mobilization was detected with the help of the result of locus-specific PCR. In those cases of a new position, the transgene was again mapped. A pedigree was maintained for these single-copy animals (**Fig. 2.4**).





**Figure 2.3** Steps to detect the single-copy mouse from the ones with high copy tandem repeat, using techniques like (a) Real-Time PCR, (b) ddPCR, (c) gene mapping protocol, (d) Ligation-mediated PCR, (e) sequencing of amplicons, & (d) BLAST.



List of loci for LacZ reporter lines			
Mouse i.d.	chromosome	Start locus	+/-
<b>LacZ-WT</b>			
Grandi0160	16	88926935	-
Grandi0161	6	16947936	+
Grandi0163	14	4181303	+
<b>LacG-CGless</b>			
Newkirk061	19	12842495	+
Newkirk066	15	102036502	-
Newkirk071	16	72415026	-
Newkirk082	2	102938808	-
Newkirk221	14	53635623	-
Newkirk141	15	3479407	-
Newkirk529	1	50984507	+
Newkirk466	1	56757567	+

d.

**Figure 2.4** Detection of transgenic loci for the sublines. (a) Design of primers for locus-specific PCR, using the BLAT results, (b) Sample of gels with expected bands from three sublines, (c) a prototype of pedigree maintained for the single-copy mice, and (d) a list specifying the details of the mapping result for the sublines.

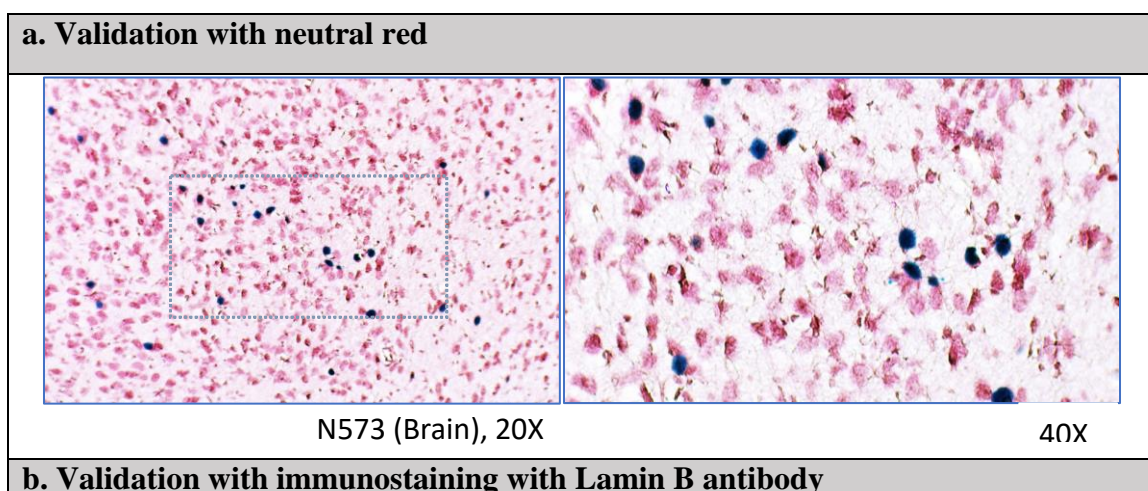
### 2.4.3. X-gal staining to detect 5'UTR-LacG or 5'UTR-LacZ transgene expression in mouse tissues *in situ*

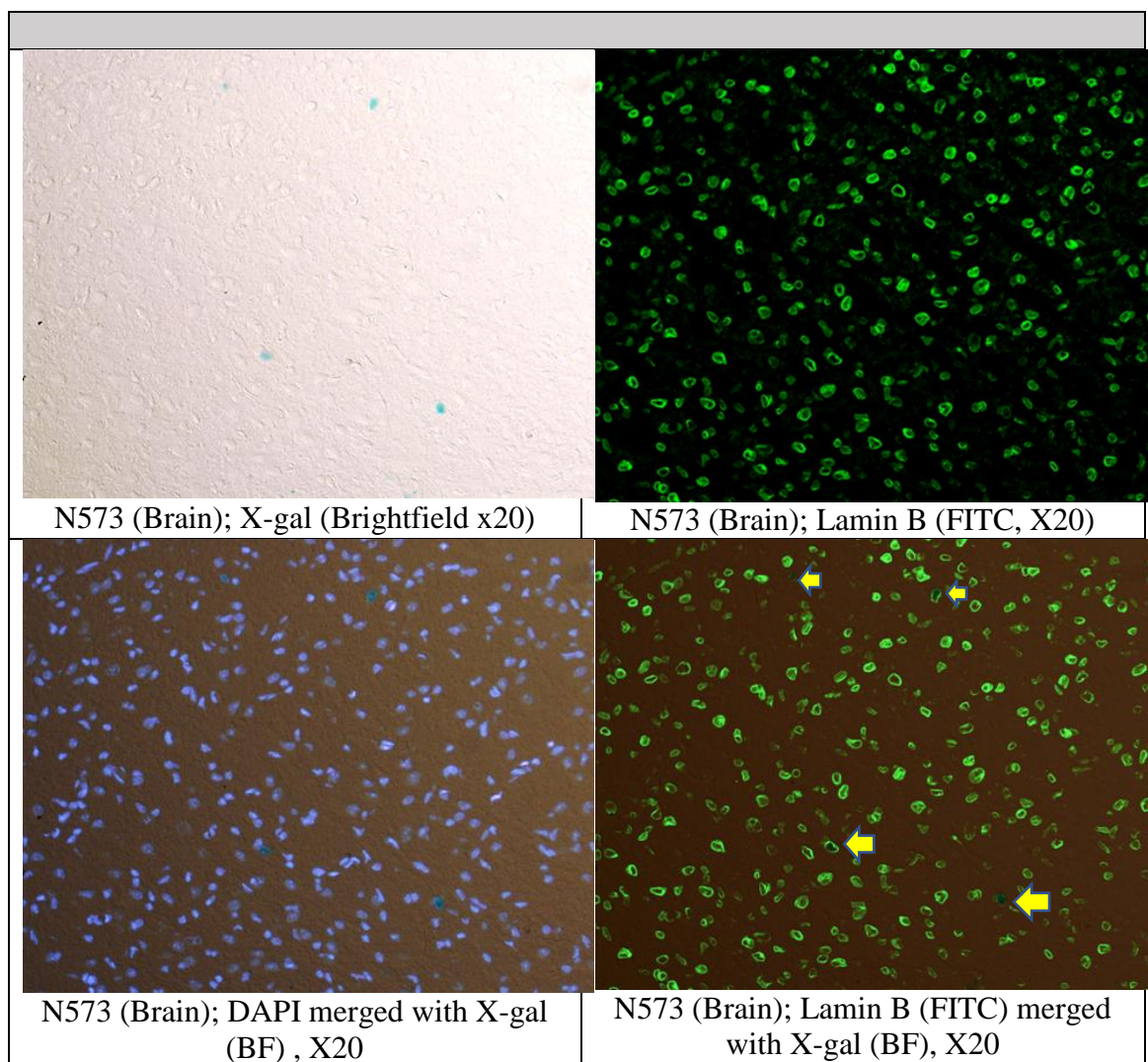
In this part, X-gal staining for the cryosections was validated and optimized. As the cryosections were X-gal stained and followed by counterstain with neutral red. We observed the X-gal signals being colocalized with neutral red (**Fig. 2.5.a**) under light microscopy. Upon immunostaining of the X-gal stained sections, we observed the blue stains surrounded by lamin B signals (**Fig. 2.5.b**). This added to the 2<sup>nd</sup> line of verification of our X-gal stains. Later, we optimized pH of the X-gal solution to an optimum point which is suitable to eliminate the signals from endogenous beta gal if present in any tissue. We use Z/EG transgenic mouse as a control. We also found that at pH 7.7, ZE/G kidney maintained a substantially intense X-gal signal (**Fig. 2.6**).

The optimum level of fixation of tissues in fixative is necessary for an ideal X-gal staining at determined pH. Additionally, a perfect fixation condition would provide us with a better

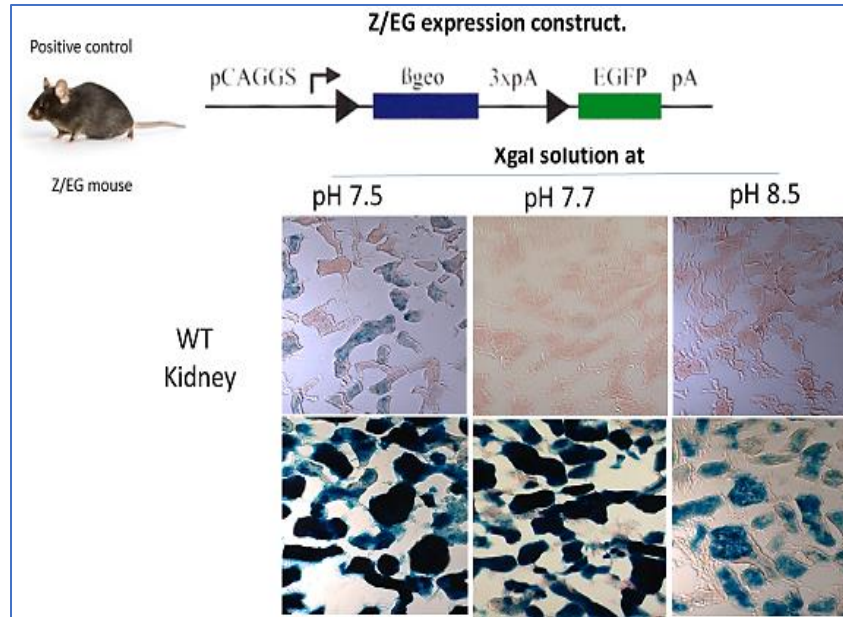
morphology of the tissues in the shortest period, retain the maximum  $\beta$ -gal from a time-dependent degradation, and prevent over-fixation of tissue to guarantee maximum stain. In neonatal time point, we observed that drop-fixation condition with 2% PFA with additives for 4 hours turned out to be the best condition with the shortest period and with using the lowest concentration of fixative possible (**Fig. 2.7**). On the other hand, at an adult time point, the longer time and higher concentration of fixative – i.e. 6 hours into an additive-mixed 4% PFA solution devoid of glutaraldehyde - were essential to stain the tissue samples (**Fig. 2.8**). Notably, in both of these time points, control Z/EG tissues were used.

Finally, desired staining patterns, which matched with their corresponding genotypes, were attained for all five genotypes used in this study (**Fig. 2.9**). Next, the whole slide scanned images of the tissue sections obtained from different lines in various context, we quantified the X-gal signals using the QuPath software. It uses a machine learning approach to bio-analyze the whole slide images. This allows a user to teach QuPath how to distinguish individual cells by separating stains and also determining the intensity peaks for either neutral red, X-gal, or both (X-gal positive cells) within a marked annotation (**Fig. 2.10**). It was possible to make the detections fully automated using scripts.

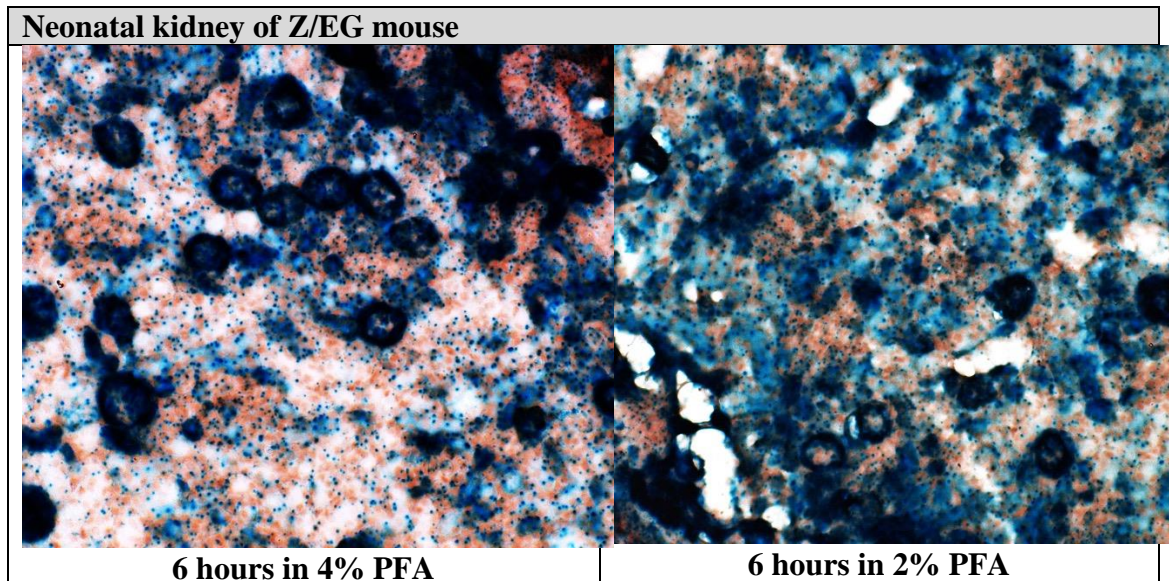


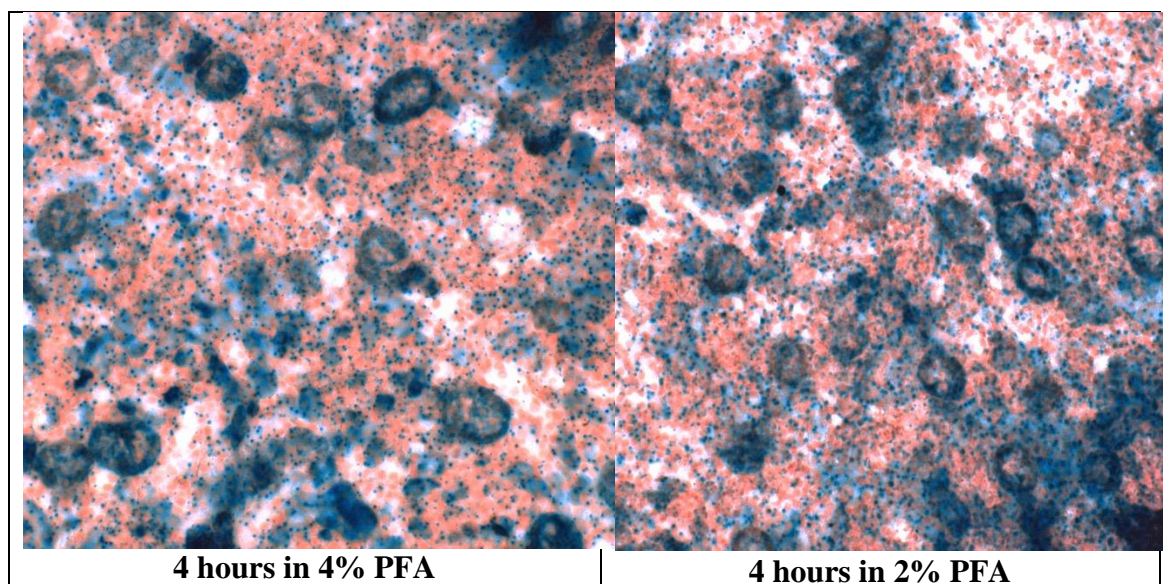


**Figure 2.5:** Validation of X-gal signals. (a) Validation with neutral red co-staining. (b) Validation with X-gal staining followed by immunostaining with Lamin B antibody.

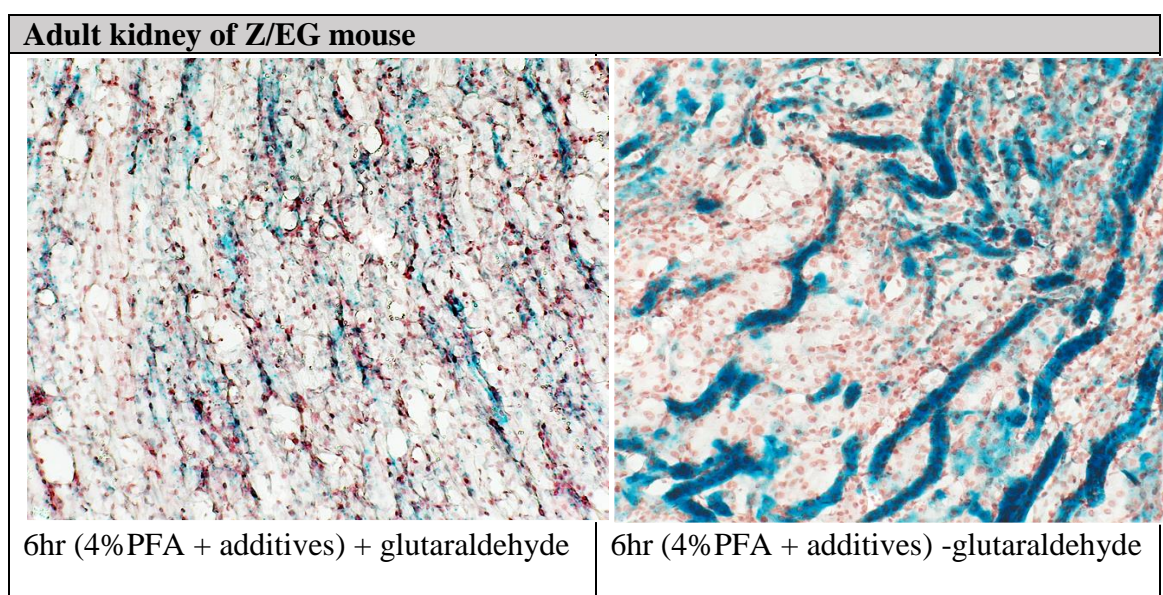


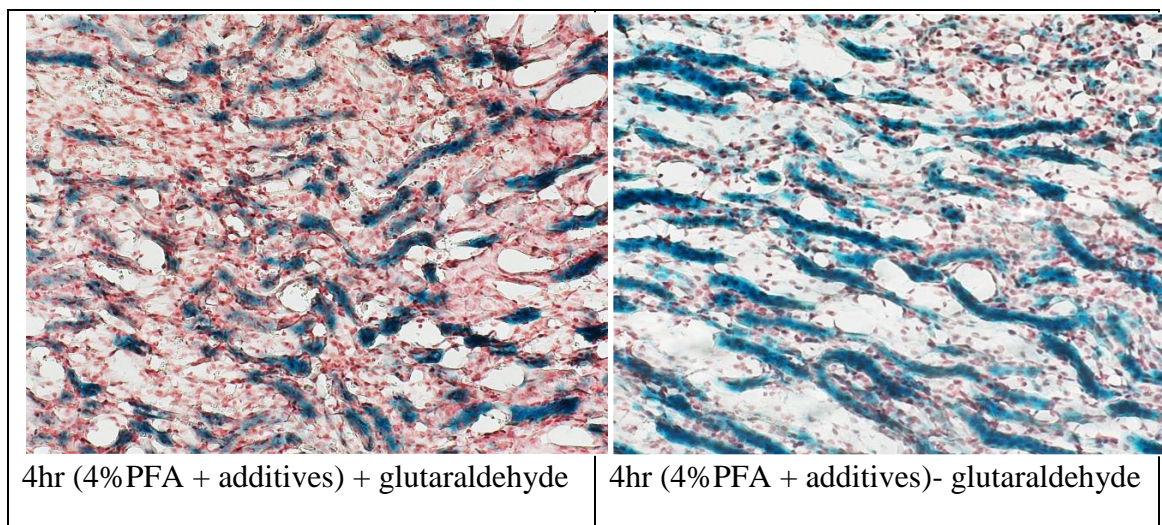
**Figure 2.6** A pH gradient (7.5, 7.7 and 8.5) of X-gal solutions was used to stain adult kidney sections from the non-transgenic mouse as well as Z/EG mouse.



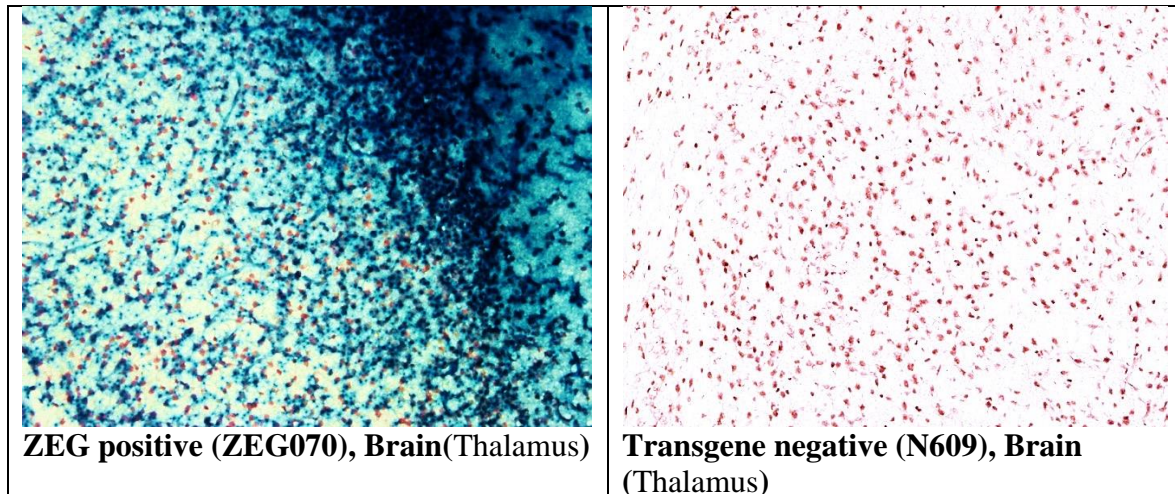


**Figure 2.7:** In search of a suitable tissue fixation condition, different combinations of duration and conc. of fixatives were used in the neonatal kidney of Z/EG mouse.

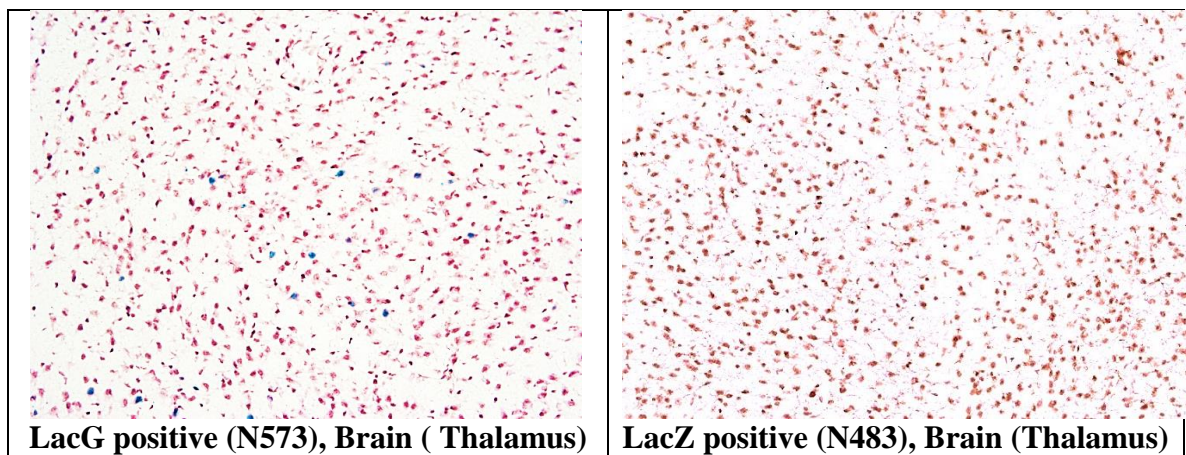




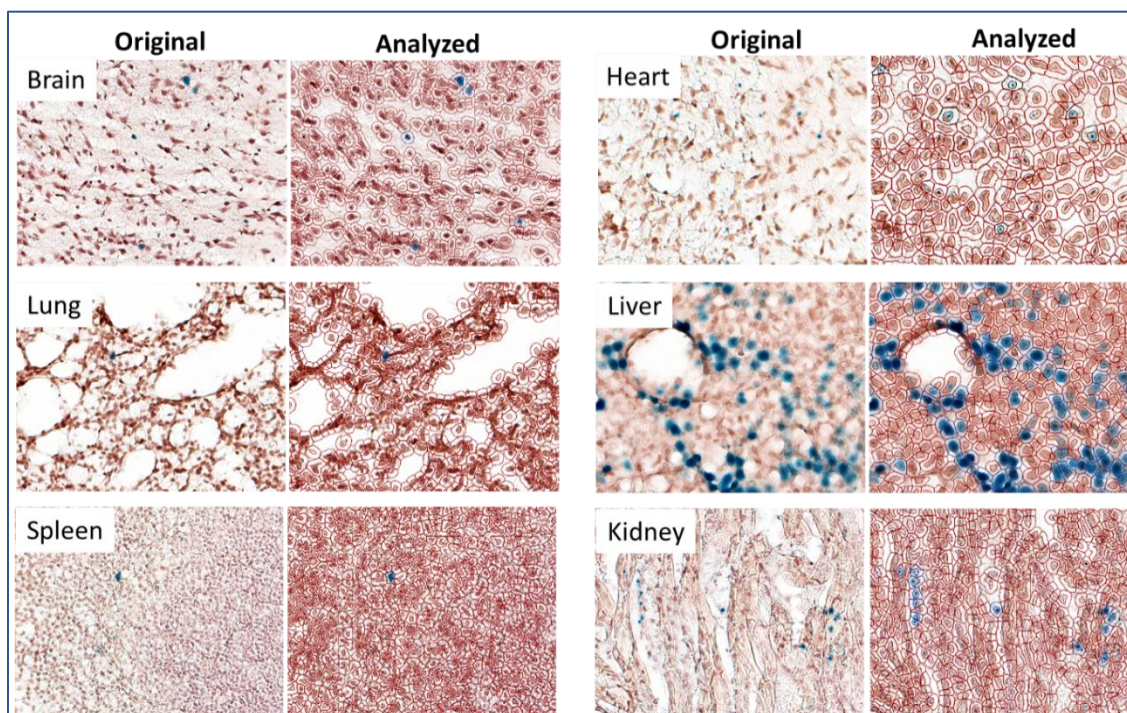
**Figure 2.8:** In search of a suitable tissue fixation condition, different combinations of duration and conc. of fixatives were used in the adult kidney of Z/EG mouse.





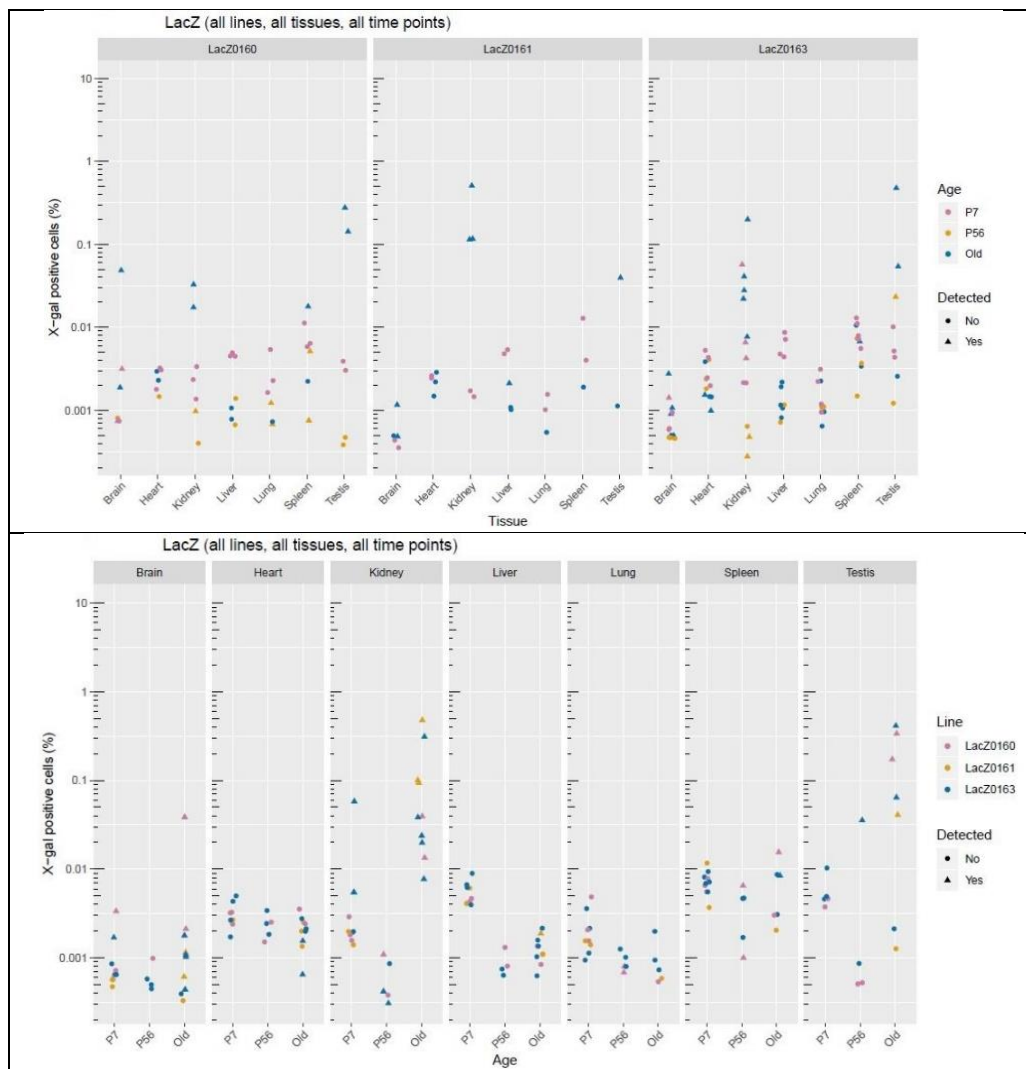


**Figure 2.9:** Desired staining patterns matched with their corresponding four genotypes used in this chapter.



**Figure: 2.10** Detection of X-gal positive cells using QuPath. Here, Original column, corresponding to each organ shows original stained tissues. When these regions were fed in QuPath for quantification, the detections were shown to be annotated accurately with red borders for counterstains (neutral red) and blue borders for X-gal stains.

## 2.4.4 Random lines with 5'UTR-LacZ had low expression at all time points



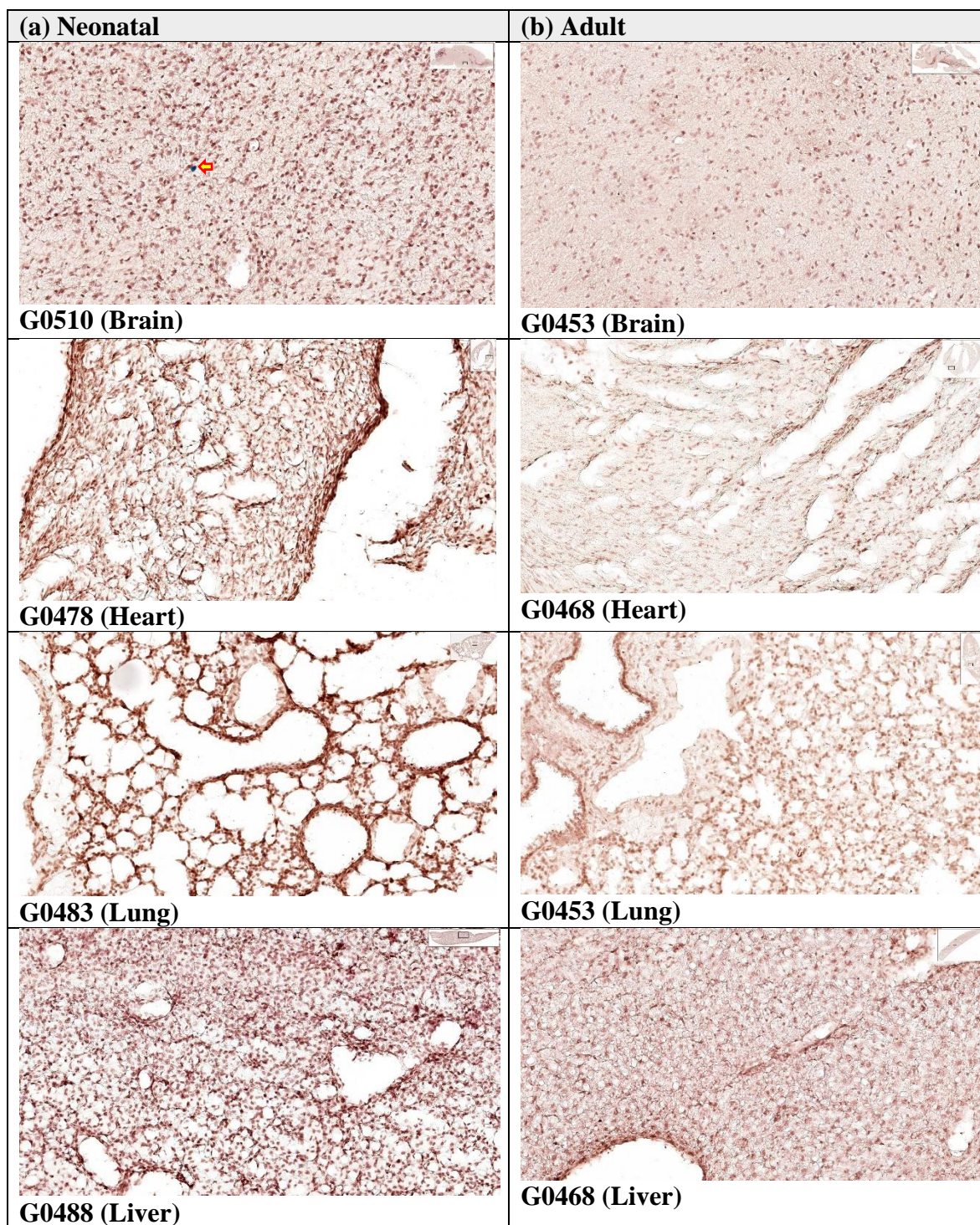
**Figure 2.11** *In vivo* promoter activity of 5'UTR fused with WT-LacZ sequence, being placed in three random genomic loci at three different time points.

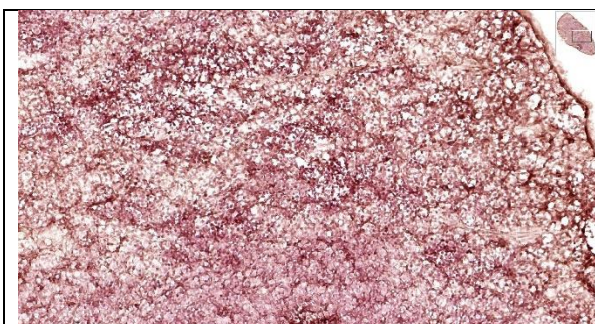
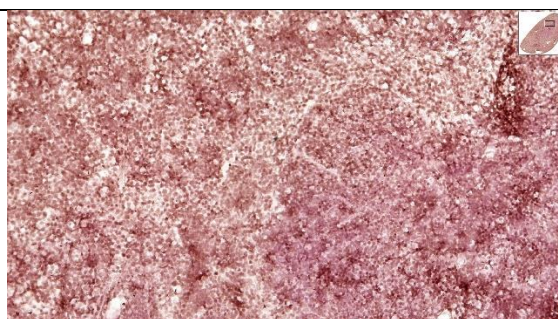
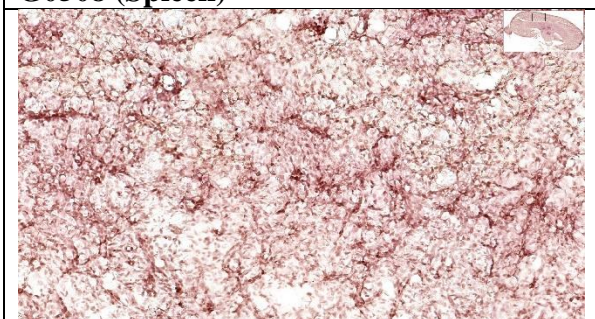
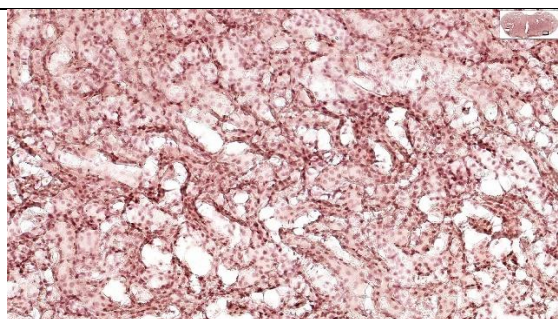
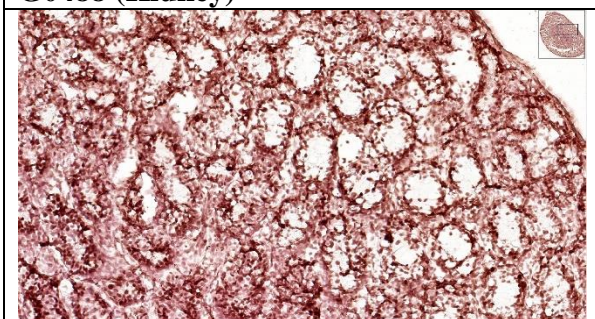
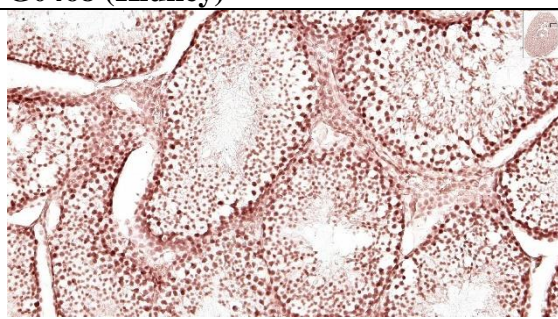
Neonatal time point has detectable signals only in brains (of 2 out of 10 animal for all three lines) and kidneys (of 3 out of 10 animals for all three lines). For kidney, all 3 of 10 animals belonged to one particular line, LacZ0163. For the brain, 2 of the 3 detectable values were from another particular line, LacZ0163. However, in other neonatal organs, there are no detectable signals. The highest level of detectable expression for the neonatal brain was

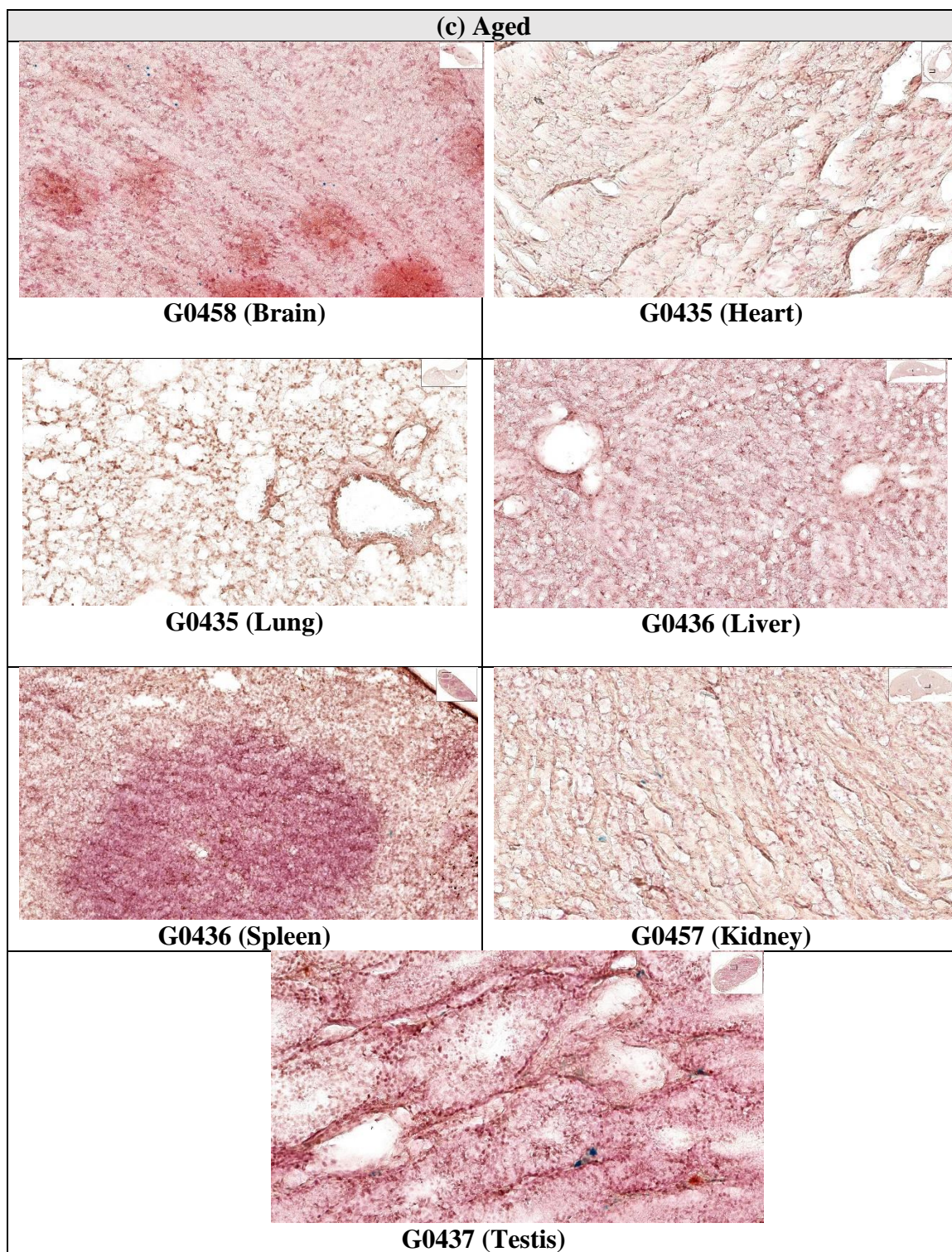
0.003%, and for kidney, it is 0.06%. On the other hand, at an adult time point (aging 2 months), several organs showed to have detectable values, namely lung, kidney, spleen and testis (**Fig. 2.11**).

Finally, most of the old animal (aging 12 – 18 months) brains had a detectable range of values of around (**Fig. 2.11**). Among 10 animals, 2 were undetected with a detectable range of 0.001- 0.04%. Kidney and testes consistently have the highest level of expression at the old-time point. In the case of the kidney, 3 of the 4 highest values are coming from a particular line, LacZ0161. In the case of the testis, 2 of the top 3 highest values are coming from a particular line, LacZ0160. In spleen, lung, liver, and heart, however, showed minimum expression.

Among the three time points, the highest value of the percentage of the X-gal positive cells is still below 1%. Heart, liver, & lung showed the most number of undetectable signals, say 19 out of 21, or ~90 % of animals showed no signals in the heart. Twenty-two out of 23 animals i.e. ~95 % of animals showed no signals in the liver, whereas out 20 of 22 animals, or in other words, ~90 % of animals showed no signals in lungs. Most signals are undetected in neonatal and adult (**Fig. 2.12**); however, the most number of detectable signals were found at aged animals. In all three lines, during the old-time point, kidneys and testes showed a consistent expression pattern. Moreover, the kidney also showed the highest number of expression in old animals. Among three lines, LacG0160 showed comparatively lower expression in kidney than the other two lines, whereas LacZ0161 showed lower testis expression than the other two.

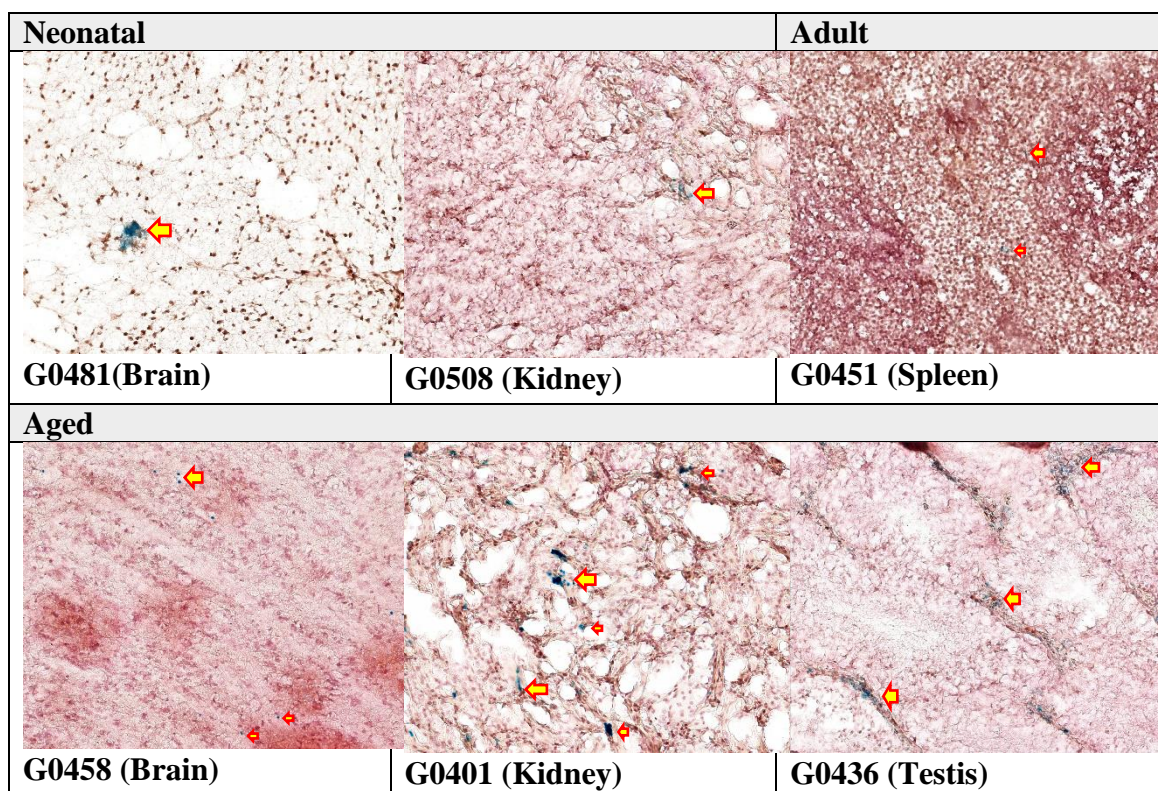


**G0508 (Spleen)****G0452 (Spleen)****G0488 (Kidney)****G0468 (Kidney)****G0510 (Testis)****G0452 (Testis)**



**Figure 2.12.** LacZ tissue sections with low/no staining at 20X

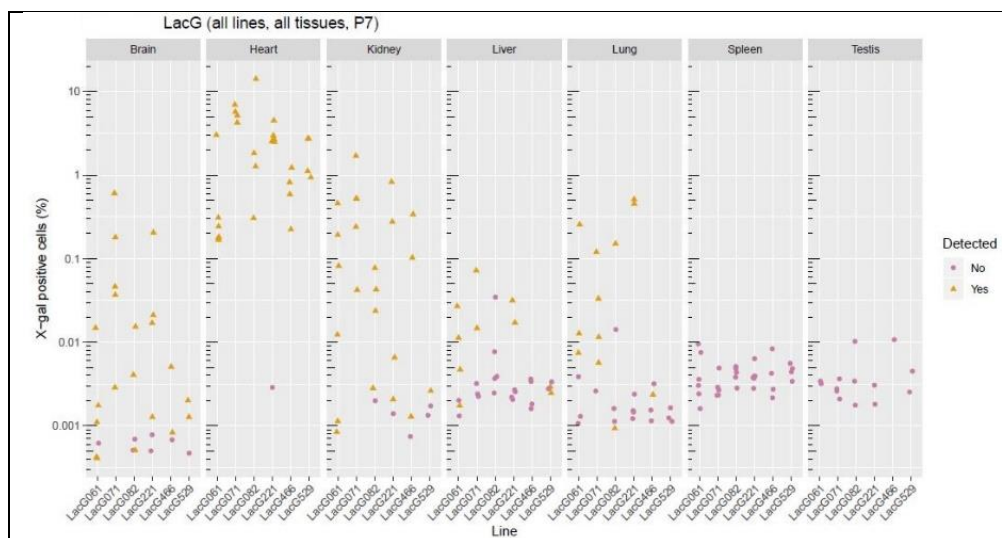
Figure 2.13 showed few of the histological locations on the tissues, where X-gal stains were found to be present relatively more in number in some of the organs across three different time points. Interestingly, some of these animals were litter-mates of the animals showed in the above panel. In testis, X-gal positive cells lied at the border of the seminiferous tubules. In brain and kidney sections belonged to neonatal and old, X-gal positive cells were found to present in the cluster.



**Figure 2.13** Substantially (X-gal) stained LacZ tissue sections at 20X (labelled the positive cells with yellow arrow)



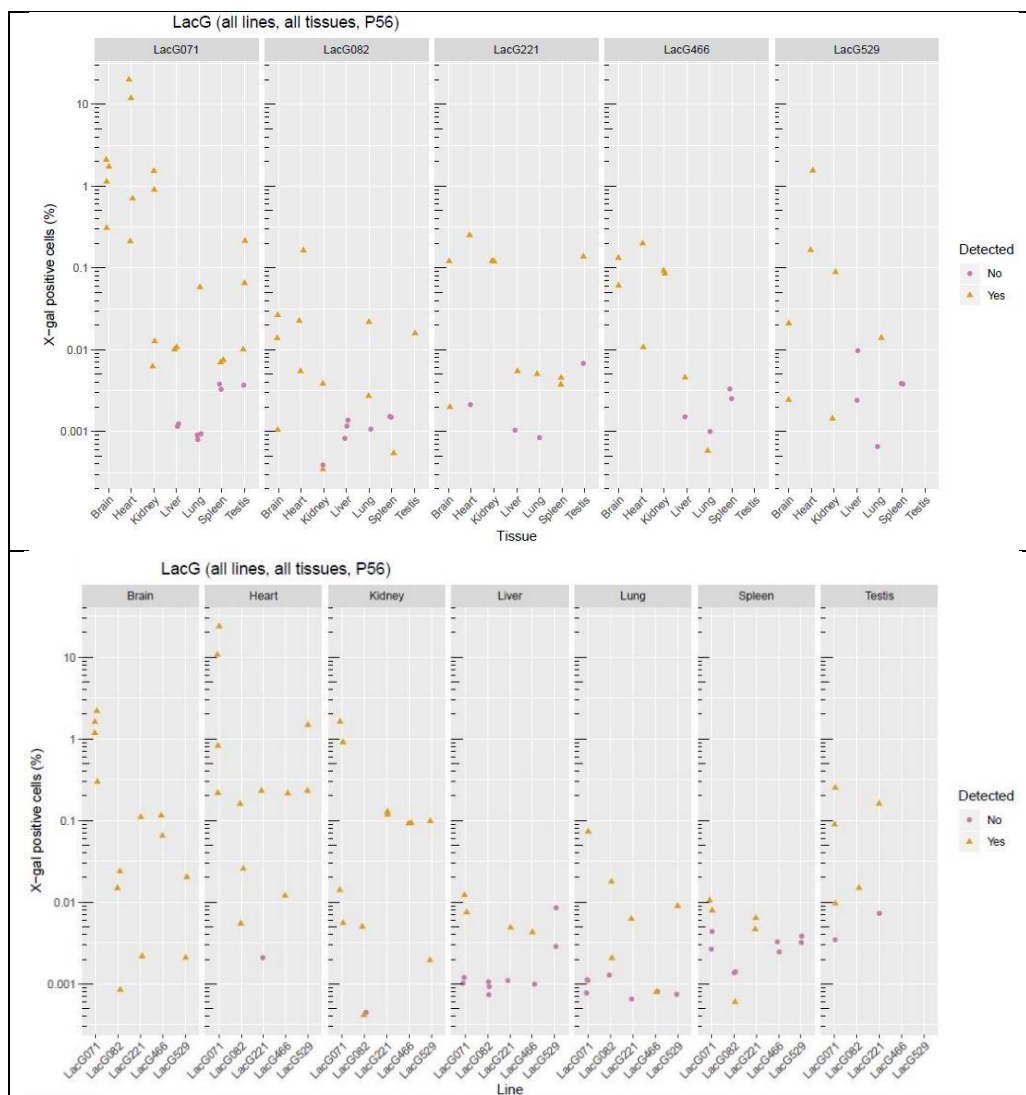




**Figure 2.14** The promoter activities (in percentage of positive cells) in different neonatal organs of all sublines belonged to LacG line. (a) Line-wise (b) organ-wise distributions.

The expression in brain varied across the lines with a least detectable value at 0.003% to highest being at 0.6%. LacG071 is the highest expression in the brain, where the individual values varied highly. Heart values were highest among all the tissues. LacG071 showed the highest activity among the heart values in an average. However, in the heart, many positive signals were not localized in the nuclei. This will be separately reported in the next section below. No detectable values were found in spleen and testis. Expression values fluctuated a lot in most of the organs. LacG071 and LacG061 showed most expressions of high values. Livers & lungs had detectable expressions. In testes, no detectable values were observed at the neonatal stage (**Fig. 2.14**).

### 2.4.5.2 Expression level during adult time point

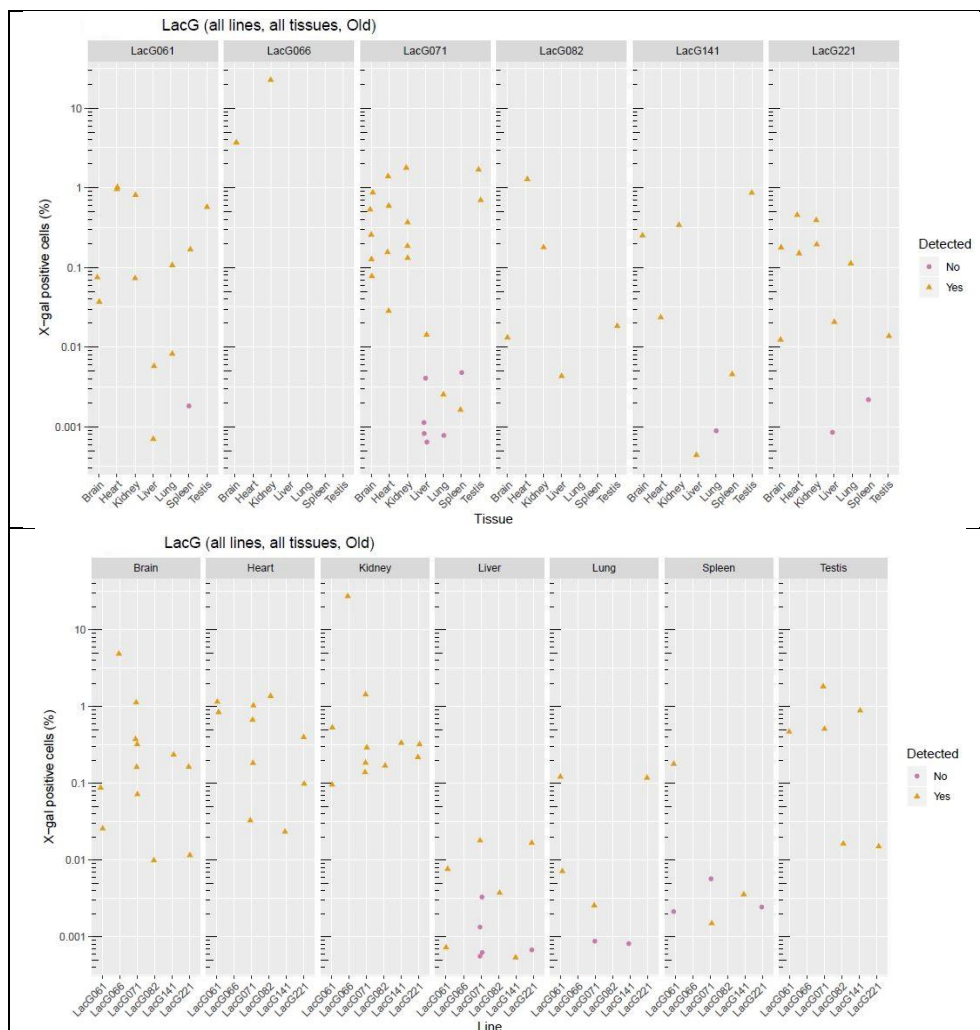


**Figure 2.15** The promoter activities (in percentage of positive cells) in adult (56 days old) organs of all sublines belonged to LacG line. (a) Line-wise (b) organ-wise distributions.

LacG071 has the highest numbers of X-gal positive cells (i.e. ~1-2%) in the brain with less standard deviation (**Fig. 2.15**). Heart again has leading values in all lines, and the levels of these values are comparable to brain and kidney. Liver, lungs and spleen had a low profile, but LacG082 had no signal in the liver. Unlike neonatal time points,

spleens and testes have detectable values in adult time points. Testes have most detectable values in LacG071.

### 2.4.5.3 Expression level during geriatric time point



**Figure 2.16** The promoter activities (in percentage of positive cells) in old animals' organs of all sublimes belonged to LacG line. (a) Line-wise (b) organ-wise distributions.

At this time point (**Fig. 2.16**), some of the new lines included which were not represented earlier in two-time points. LacG066 had the highest level of expression in the brain (4%), disregarding heart with >10% of positive cells. A comparative level of

expression was observed in LacG071 line, however, with at least 10-fold difference among individual mice. Kidney has the highest level of expression in LacG066 among all samples and all lines. Values from kidneys were comparable to the values from the heart in terms of magnitude. Liver and lung showed low profile. Testes values increased quite significantly as compared to the adult time point.

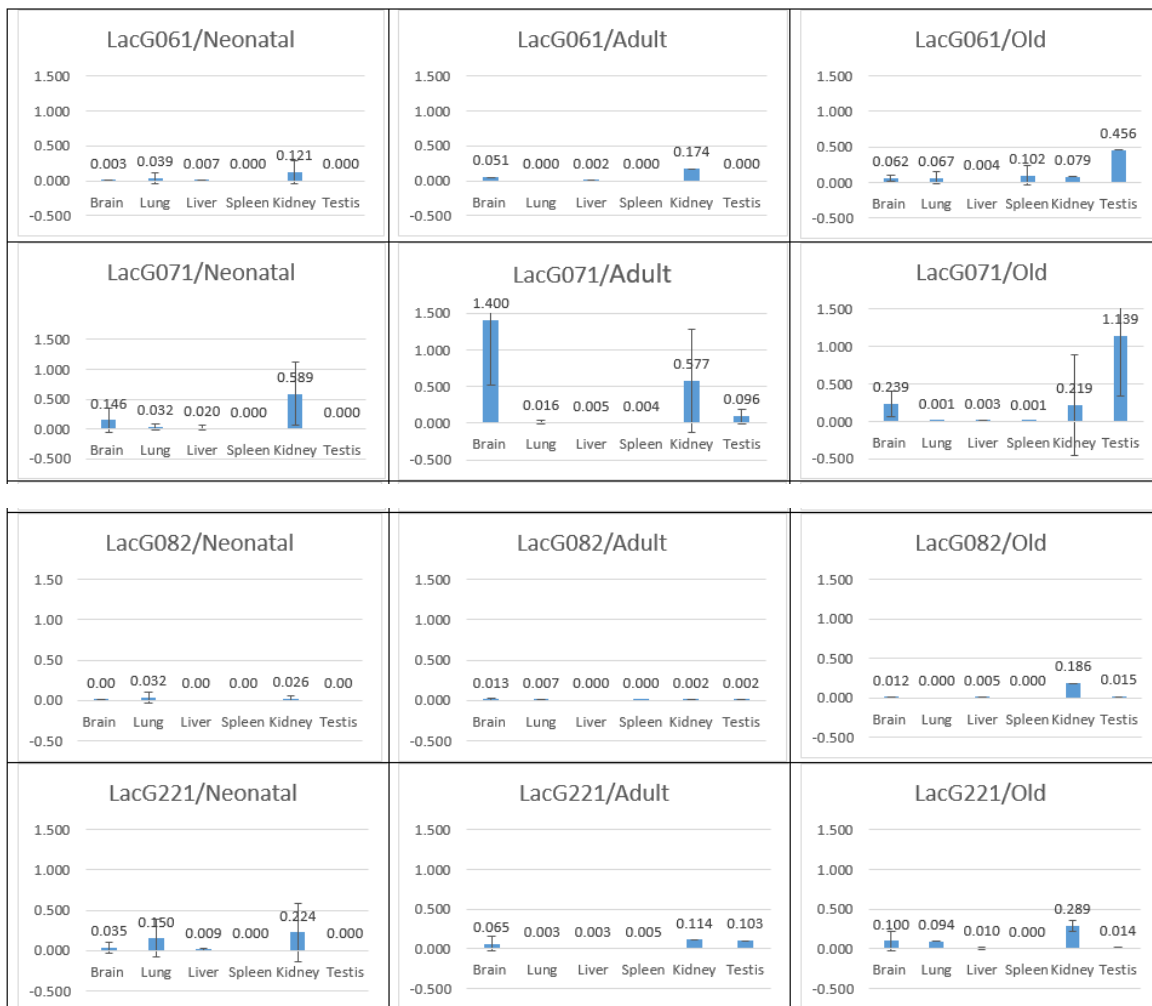
#### 2.4.5.4 Kidney showed as the most common preferential organ with highest number of promoter activity among four lines across different developmental time points

	<b>Neonatal</b>	<b>Adult</b>	<b>Old-aged</b>
<b>LacG061</b>	Kidney (0.121%)	Kidney (0.174%)	Testis (0.456%)
<b>LacG071</b>	Kidney (0.589%)	Brain (0.577%)	Testis (1.139%)
<b>LacG082</b>	Lung (0.032%)	Brain (0.013%)	Kidney (0.186%)
<b>LacG221</b>	Kidney (0.224%)	Kidney (0.114%)	Kidney (0.289%)

**Table 2.6:** The organ had the highest expression in different lines

Above table (**Table 2.6**) summarizes the organs that were found to have the highest mean of the expressions in three-time points across four random lines. Following histograms (**Fig. 2.17**) are showing different mathematical values supporting the table above. In histograms, the y-axis is the percentage (%) of positive cells and x-axis is different organs. Here we see, out of 4 unique organs in this table (brain, lung, kidney and testes) across 4 different lines, the kidney is seemingly the most preferential organ for high promoter activity among all the time points. Spleen seemed to be the lowest expressive organ. Heart values were also excluded from this comparison. Substantial standard deviation was observed. LacG082 line showed the lowest level of expression,

where expression was silent in almost all tissue. LacG071 line showed the highest level of expression in brain, kidney and testes.

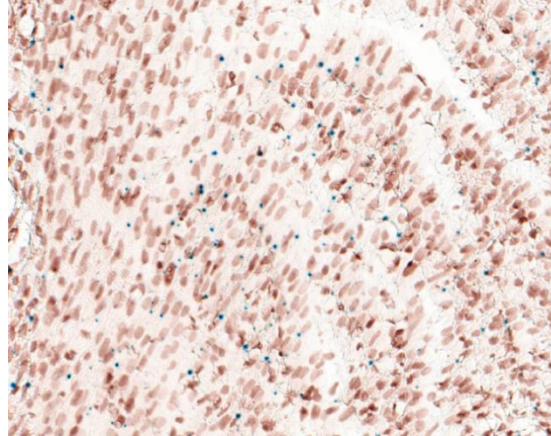
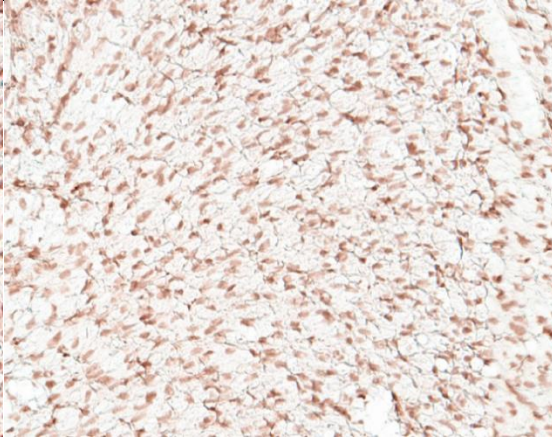
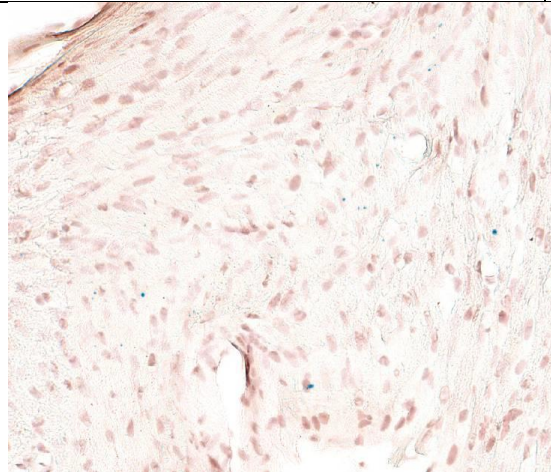

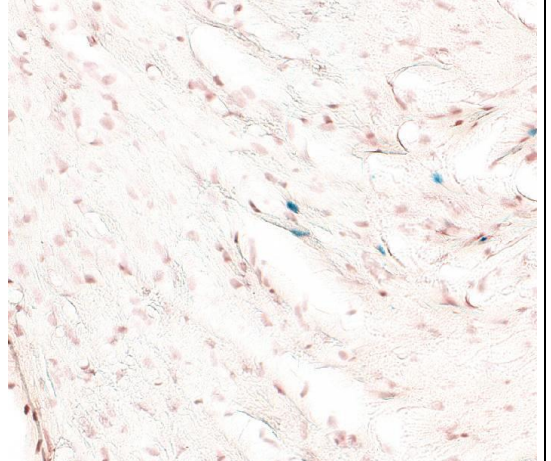
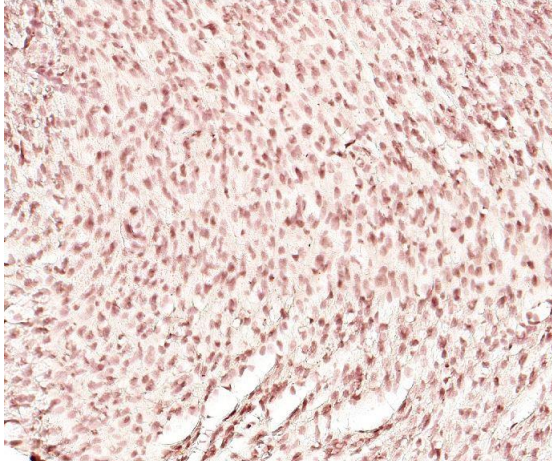


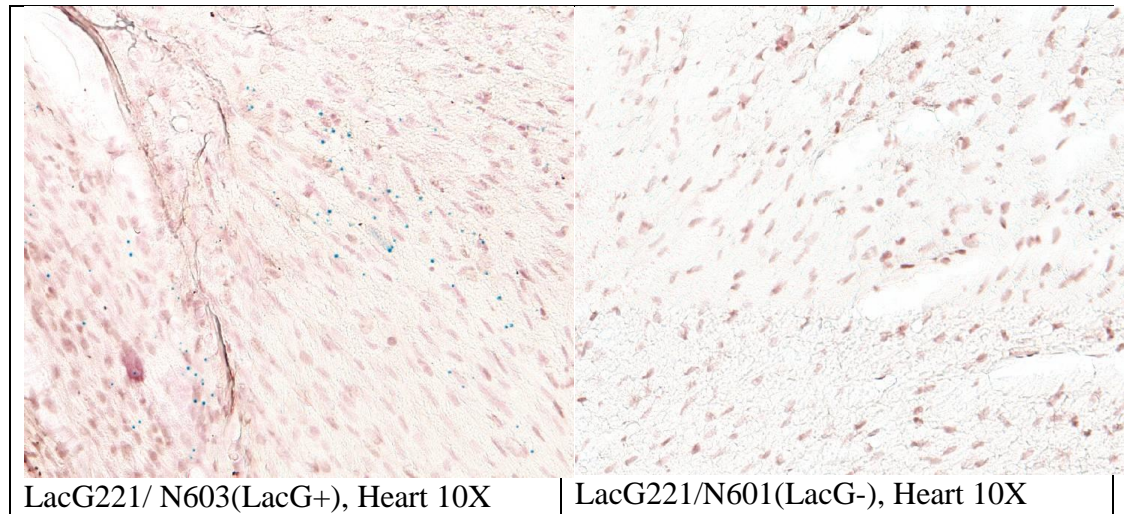
**Figure 2.17** Histograms showing average promoter activity (%) in different main organs across four different lines.

#### 2.4.5.5 Heart showed the highest signals but they are uncommon

In all time points of LacG lines, many granulated, non-nuclear localized X-gal positive signals have been observed, which, in turn, increased the positive signals in the heart.

On the contrary, the non-transgenic animals possessed no signals at all (**Fig. 2.18**).

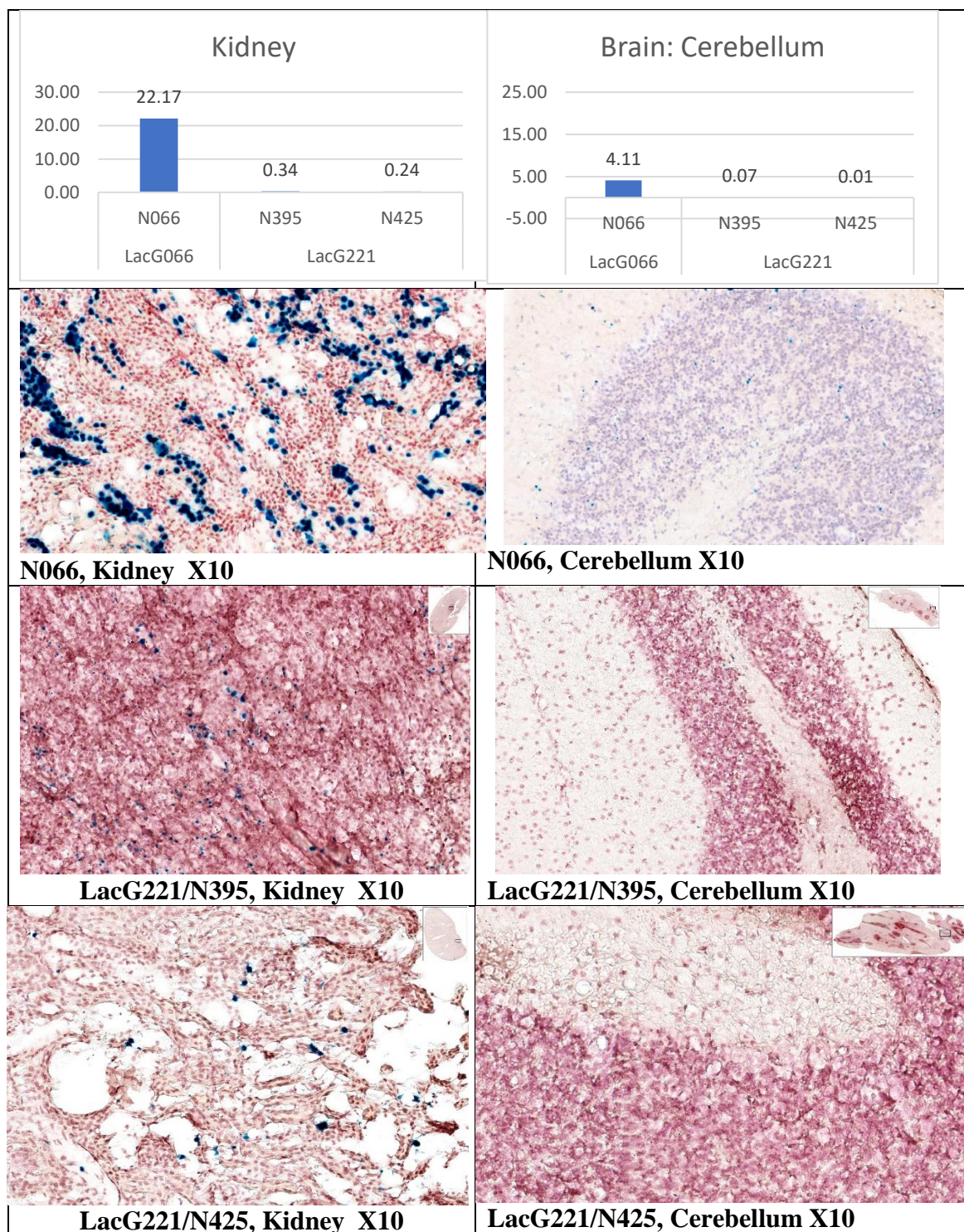
Heart signals in LacG lines	
LacG positive	LacG negative
	
LacG071/ N574 (LacG+), Heart 10X	LacG071/ N807 (LacG-), Heart 10X
	
LacG061/ N617(LacG+), Heart 10X	LacG061/ N710 (LacG-), Heart 10X
	
LacG082/ N583(LacG+), Heart 10X	LacG082/ N626 (LacG-), Heart 10X



**Figure 2.18** Uncommon x-gal signals in heart of LacG positive animals vs LacG negative animals

#### 2.4.5.6 Drastic change of expression was noticed upon moving of transgene to a new locus

During the transition from 2<sup>nd</sup> generation to the 3<sup>rd</sup> generation, transgene moved from a mapped locus point of LacG066 (Chr. 15) to a new location of LacG221 (Chr. 14). More specifically, the only survived animal belonged to LacG066, N066, was a transgenic parent of the founder of LacG221 line. At the old stage, a drastic change of expression was noticed in the brain and kidneys of N066 in compared to the same of the progenies of LacG221. The former showed the highest load of positive signals ever noticed in any animals of LacG lines. On the other hand, a minimum expression was observed in the tissues of animals of comparable age belonged to LacG221 line (that includes N395 and N425 as aged members) (**Fig. 2.19**).



**Figure 2.19:** The contrast between the x-gal signals in kidneys and cerebellums of LacG066 and LacG221 animals.



## 2.4.6. LacG071: The line with high transgenic expression in most of the tissues at all time points

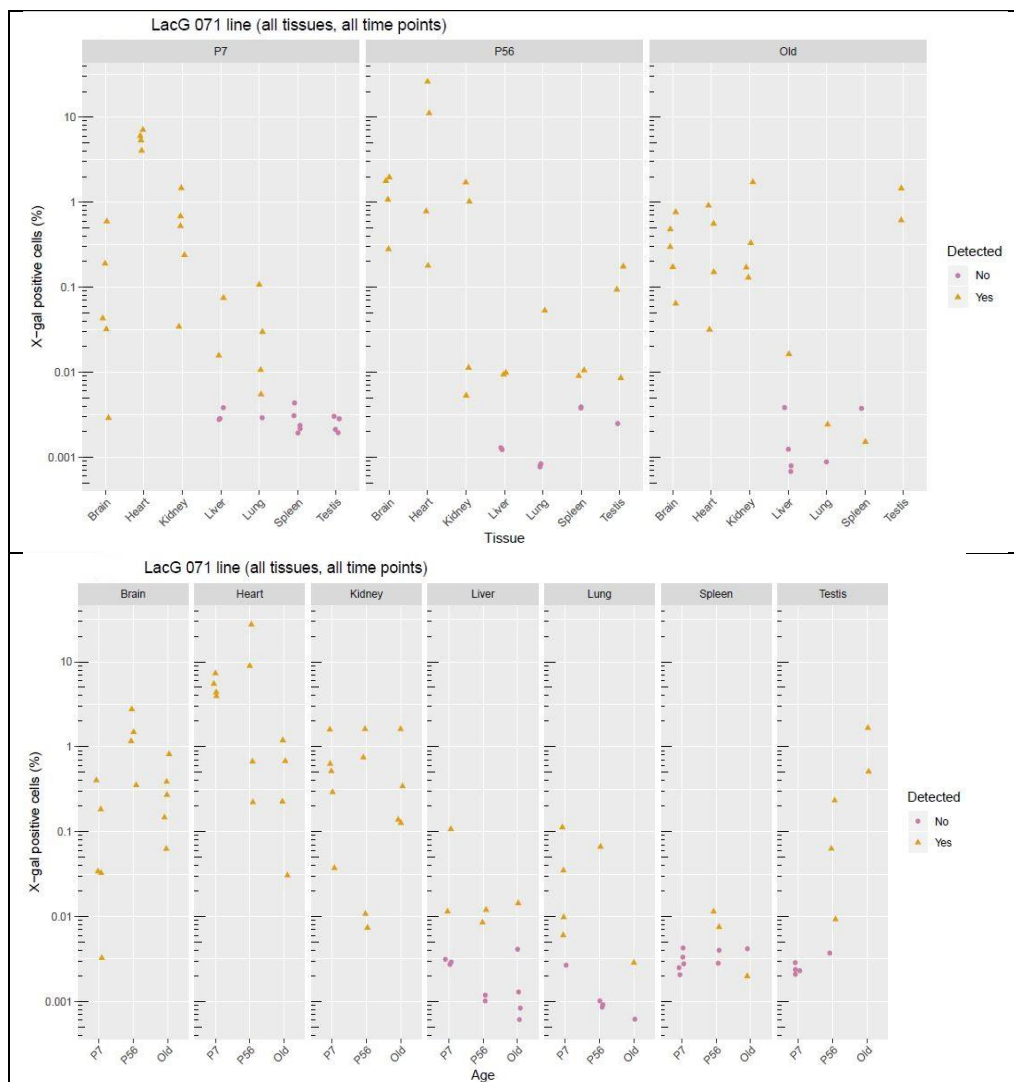
Contents:

**2.3.6.1** The promoter activities in different organs of LacG071 varied across three developmental time points

**2.3.6.2** Brain regions of LacG071 have heterogeneous activity pattern

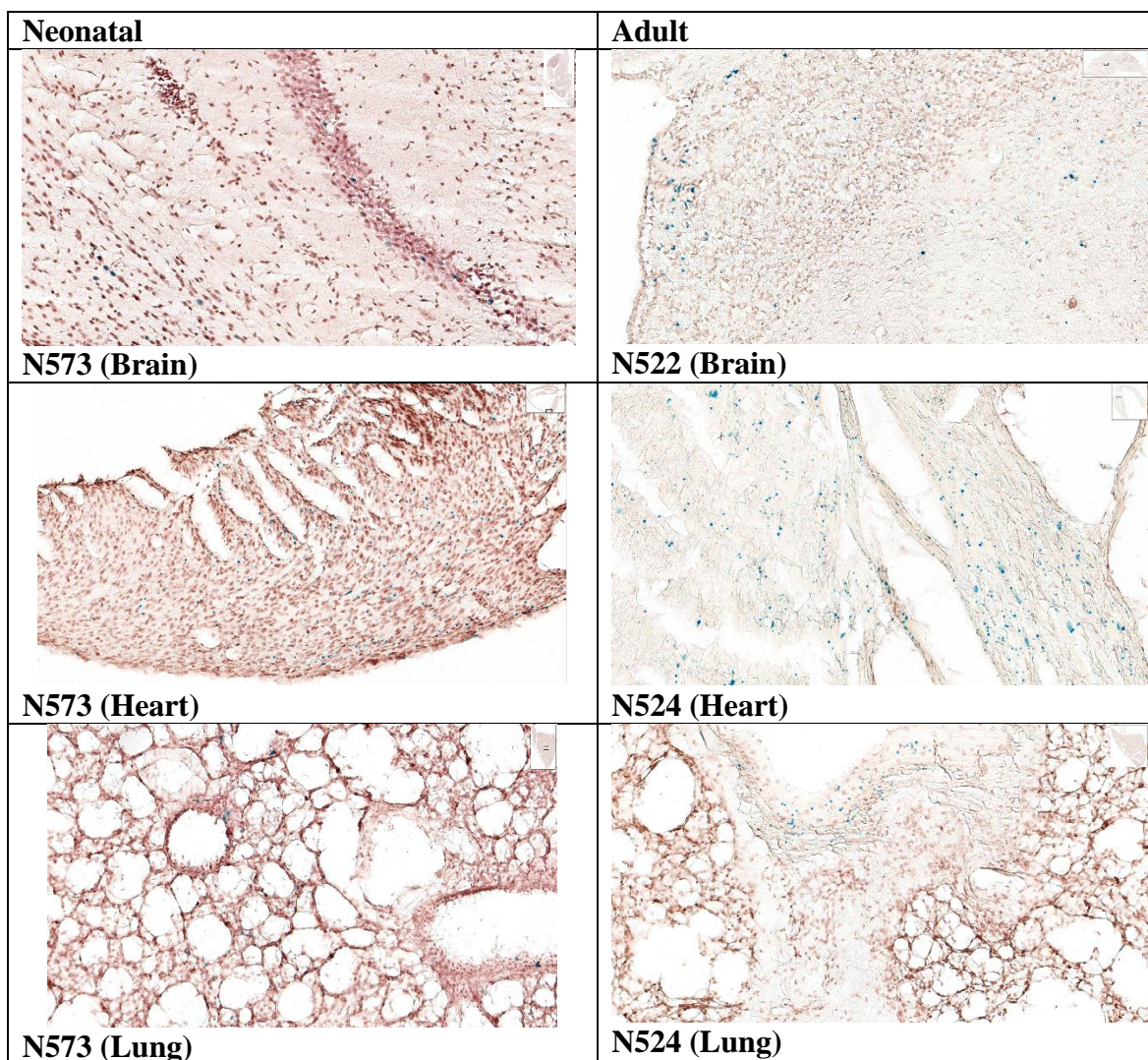
**2.3.6.3** Thalamus is the most common preferential region for promoter activity across different random lines

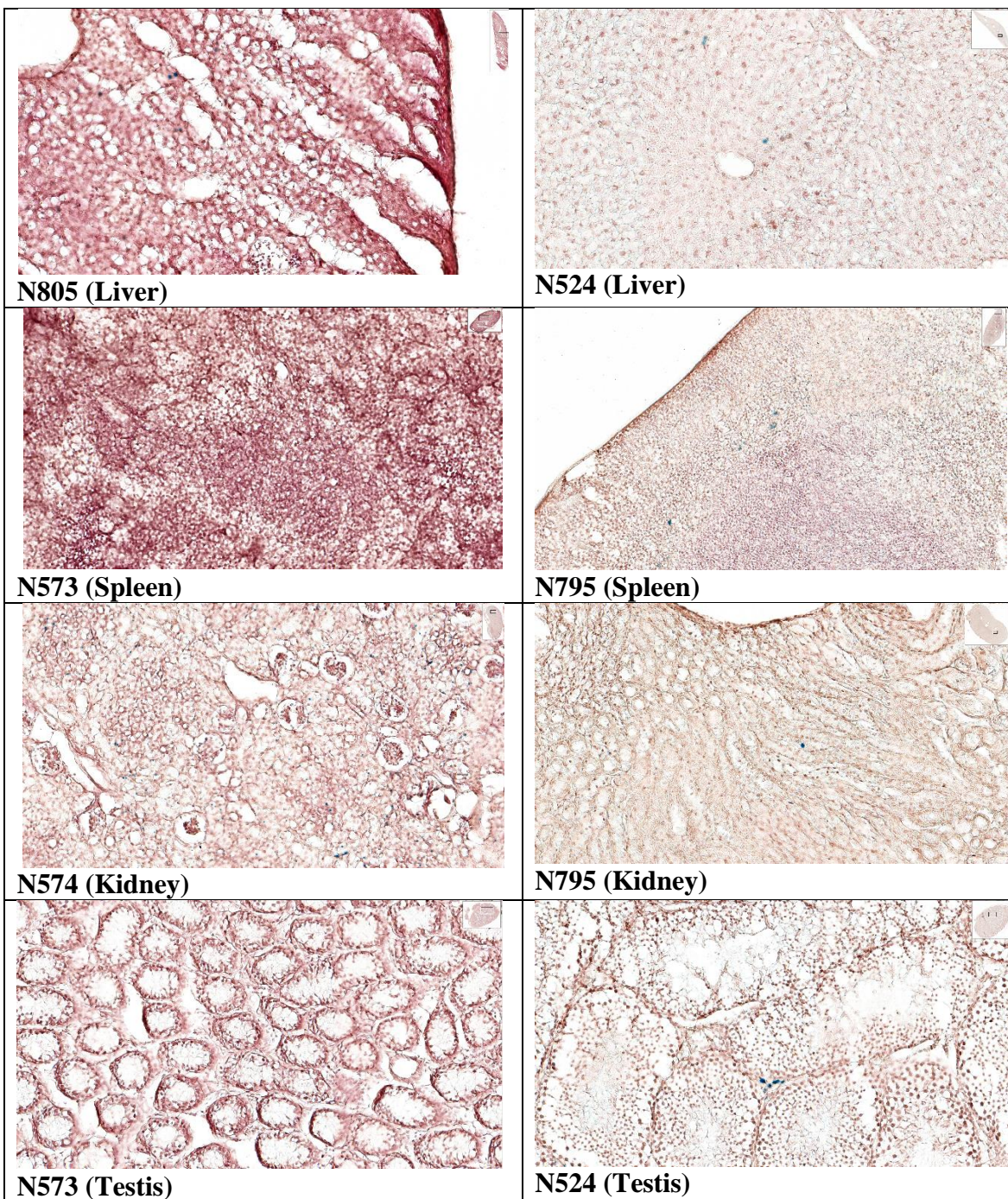
### 2.3.6.1 The promoter activities in different organs of LacG071 varied across three developmental time points

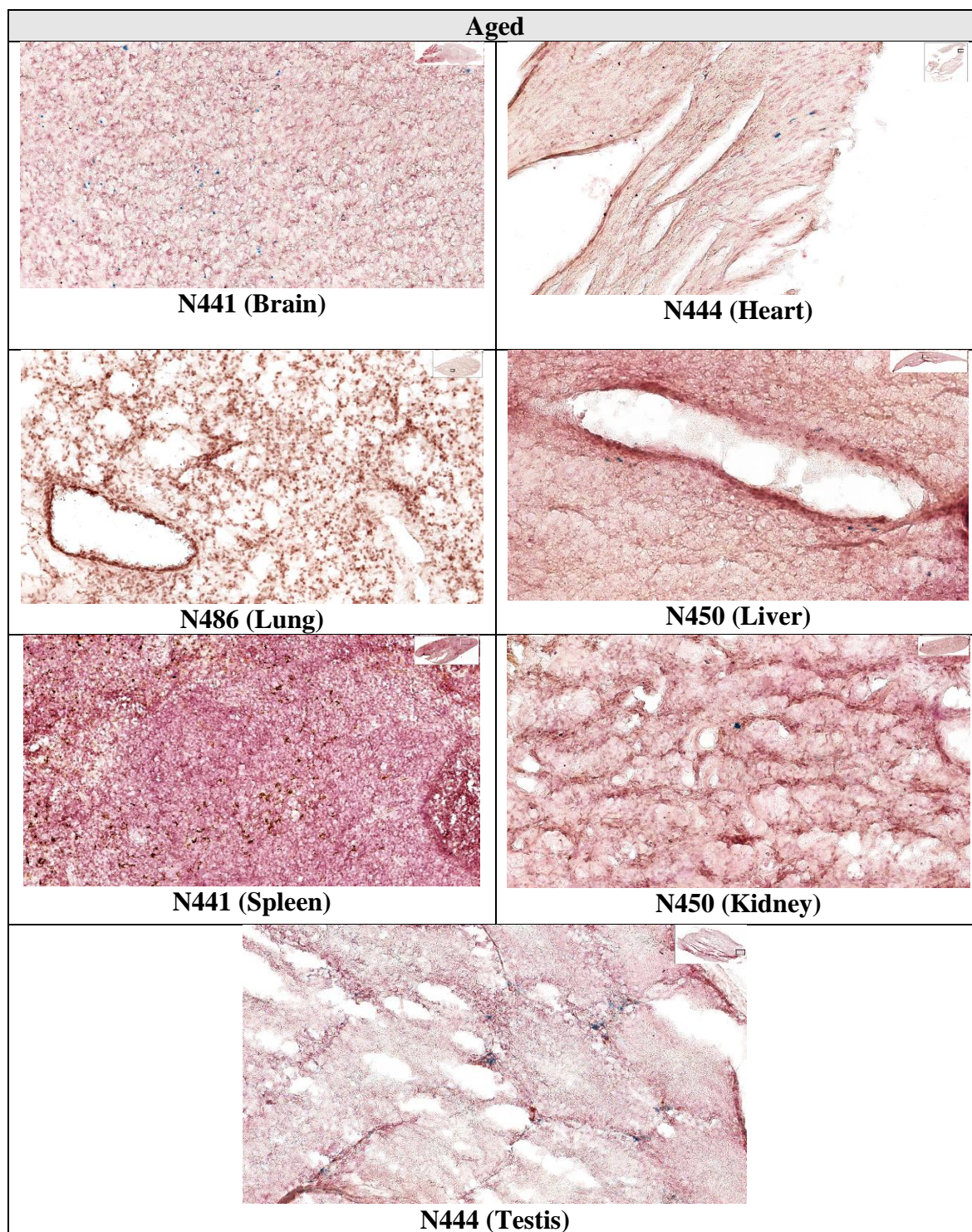


**Figure 2.20** The promoter activities in different organs of LacG071 varied across three developmental time points

Brain, heart and kidney showed high expression (**Fig. 2.20**). With less standard deviation (coefficient of variation (CV) %= 62.82, around 1% of the brain cells showed positive signals in adults, unlike other two time points. While, the liver, lungs and spleen showed a low promoter activity. No detectable value was found in the spleen during neonatal time point; however, in some adult individuals, the average value is 1 positive cell per 10,000 cells. Interestingly, no signal was detected in testes at a neonatal time point, but signal level consistently kept rising in subsequent ages, adults and aged. The above result was supported by the following images of different developmental time points (**Fig. 2.21**).

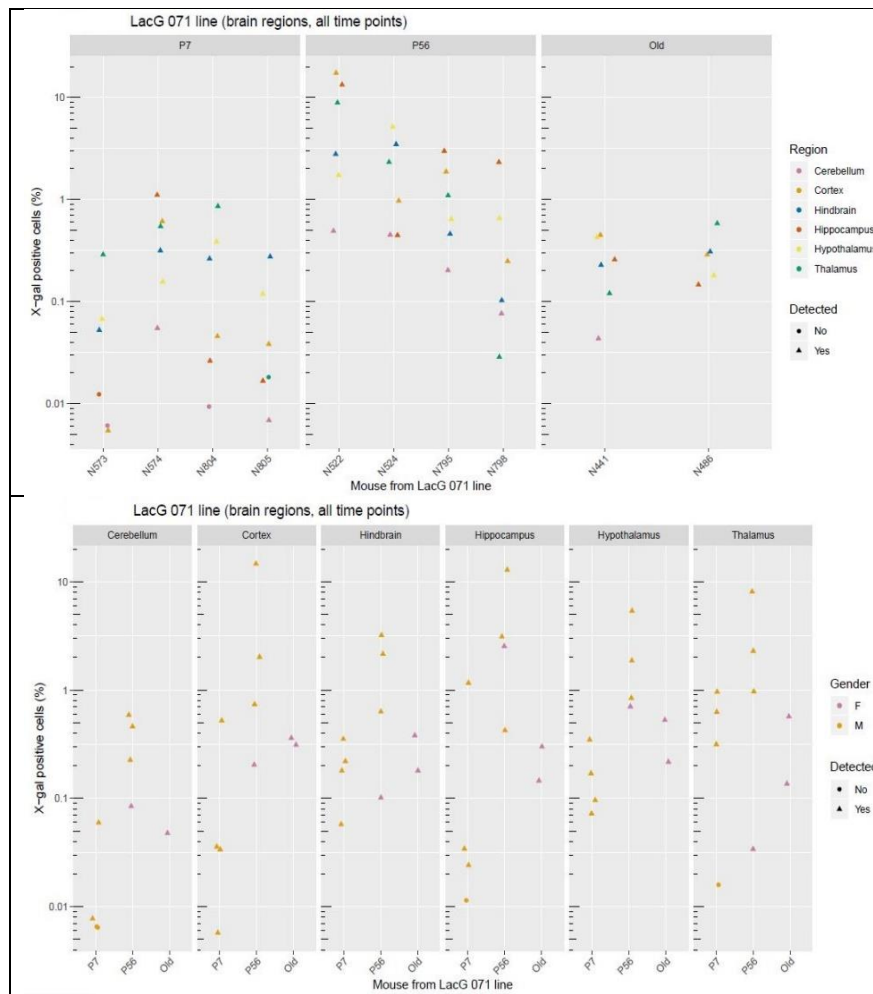






**Figure 2.21** Sample images (at 20X) of all organs of LacG071 at three different time points.

### 2.4.6.2 Brain regions of LacG071 have heterogeneous activity pattern



**Figure 2.22** L1 promoter activity (%) in different brain regions of mice belonging to LacG071 line

#### At Neonatal of LacG071

At the neonatal time point of LacG071, thalamus showed the highest values in at least two animals out of four in total. The average values from all four animals were 0.44%, which is the highest among all other averages from other regions for this time point.

#### At the adult time point of LacG071

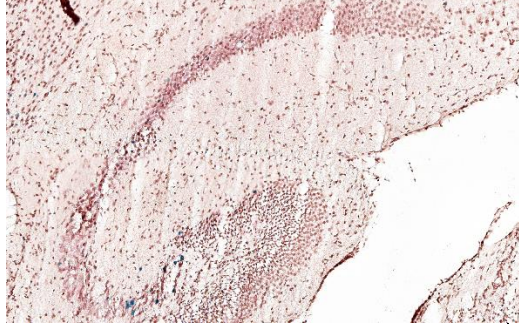
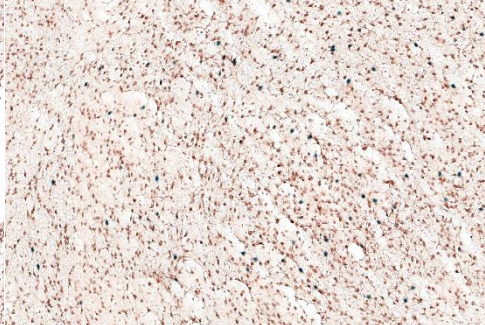


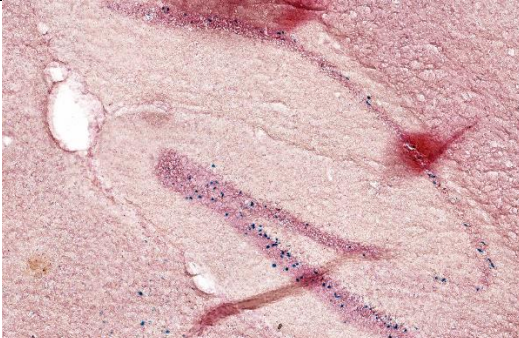
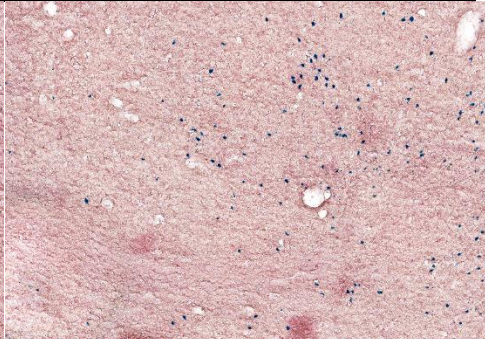
In adult animals of LacG071, out of four animals, three animals showed very high signals in the hippocampus, showing a definite expression in the dentate gyrus (**Fig. 2.23**). One

specific animal, N522, showed a high level of X-gal signals in the cortex (16.56 %) and hippocampus (14.07 %) of all adult stage animals. Except the mentioned, outlying values of N522, first and second most expressive regions were thalamus and Hypothalamus respectively. However, wide variability was observed in thalamus values among the animals (CV% = 118.46).

### **At the old time point of LacG071**

Average values from thalamus fell back sharply during older age from a peak state of adult time point. However, this region still leads in the average transgenic expression among the old animals and ties with the cortex value. A drastic change of falling of hippocampus value is a worth taking attention.

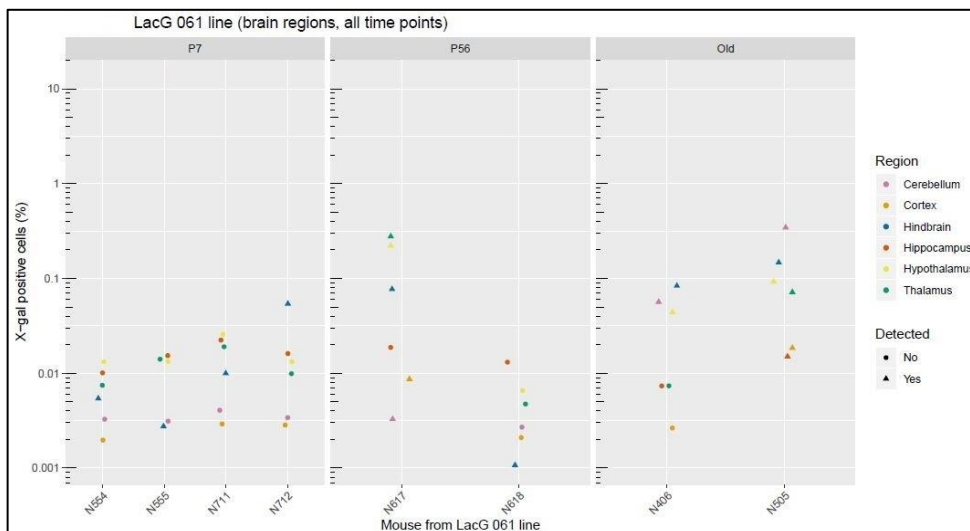
Very interestingly, adult time points showed consistently the highest average levels of positive signals in all brain regions of LacG071. Cerebellum showed minimum expression in all-time points of LacG071 line. Overall, with the change of developmental time point, sharp fall and rise of the expression were observed in hippocampus, cortex, hindbrain, hypothalamus, and thalamus values (also in **Fig. 2.27**). Thalamus was the most common region for the promoter activity in all-time points of the LacG071 line (**Fig. 2.23** and **Fig. 2.27**). A specific expression pattern along the dentate gyrus of the hippocampus in all-time points of the LacG071 line was observed (**Fig. 2.23**).

	Transgenic expression in the dentate gyrus	Transgenic expression in the thalamus
<b>Neonatal</b>	 N574 (Brain, Hippocampus)	 N804 (Brain, Thalamus)
<b>Adult</b>	 N795 (Brain, Hippocampus)	 N524 (Brain, Thalamus)
<b>Aged</b>	 N446 (Brain, Hippocampus)	 N446 (Brain, Thalamus)

**Figure 2.23:** Sample images (at 10X) of all brain regions (hippocampus & thalamus) of LacG071 at three different time points.

### 2.4.6.3 Thalamus is the most common preferential region for promoter activity across different random lines

#### 2.4.6.3.1 Brain regions of LacG061 line

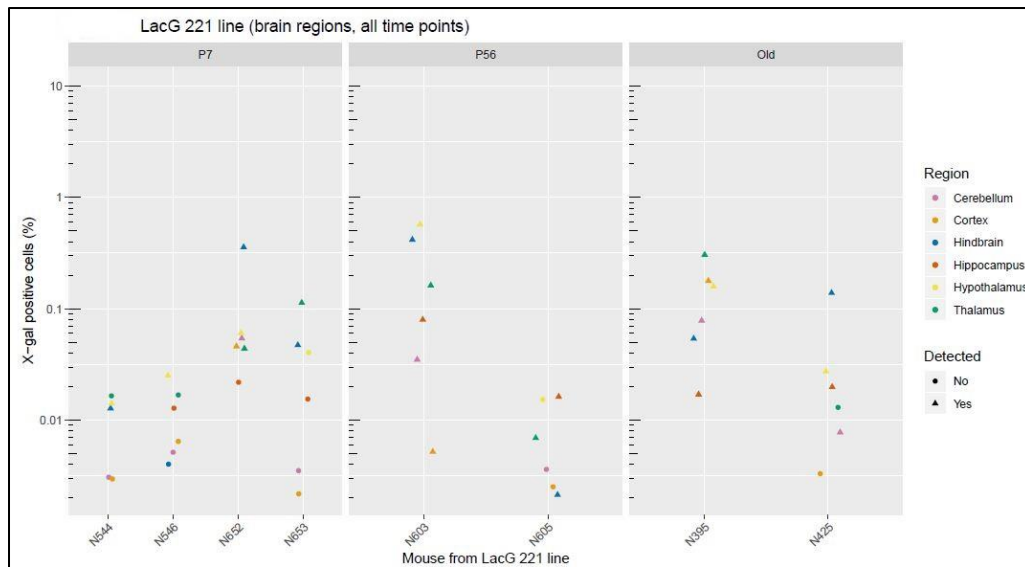


**Figure 2.24:** L1 promoter activity (%) in different brain regions of mice belonged to LacG061 line

Overall, the brain regions of all animals across different time point showed very low expression (**Fig. 2.24**). Hindbrain tops the list. In adults, thalamus and hypothalamus showed increased values in a single animal out of two. Again at the old-time point, hindbrain and cerebellum showed consistently increased values in both animals.



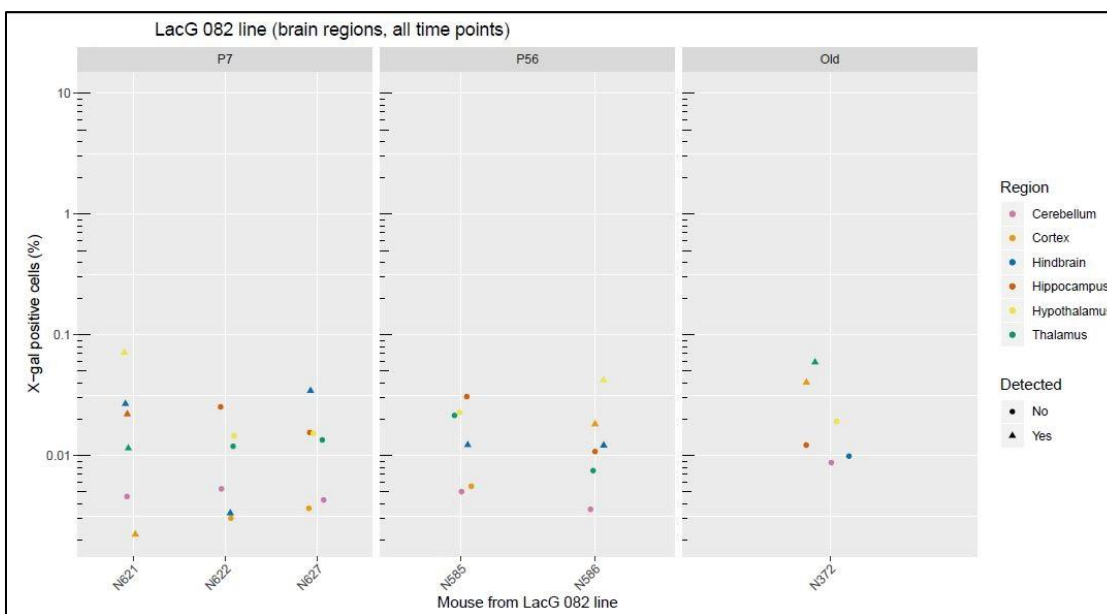
### 2.4.6.3.2 Brain regions of LacG221 line



**Figure 2.25:** L1 promoter activity (%) in different brain regions of mice belonging to LacG221 line

In both neonates and adults, hindbrain showed detectable values in all animals. And among adults, one animal showed high detectable expression in the hypothalamus. Whereas, thalamus topped in one of the old animals as the leading value, whereas in another animal it went undetectable (**Fig. 2.25**).

### 2.4.6.3.3 Brain regions of LacG082 line



**Figure 2.26** L1 promoter activity (%) in different brain regions of mice belonged to LacG082 line

It consistently maintained a low level of expression in all animal regardless of the time point. However, hindbrain leads in Neonatal, the hypothalamus in adults, and thalamus expression was prominent in the old animal (**Fig. 2.26**).

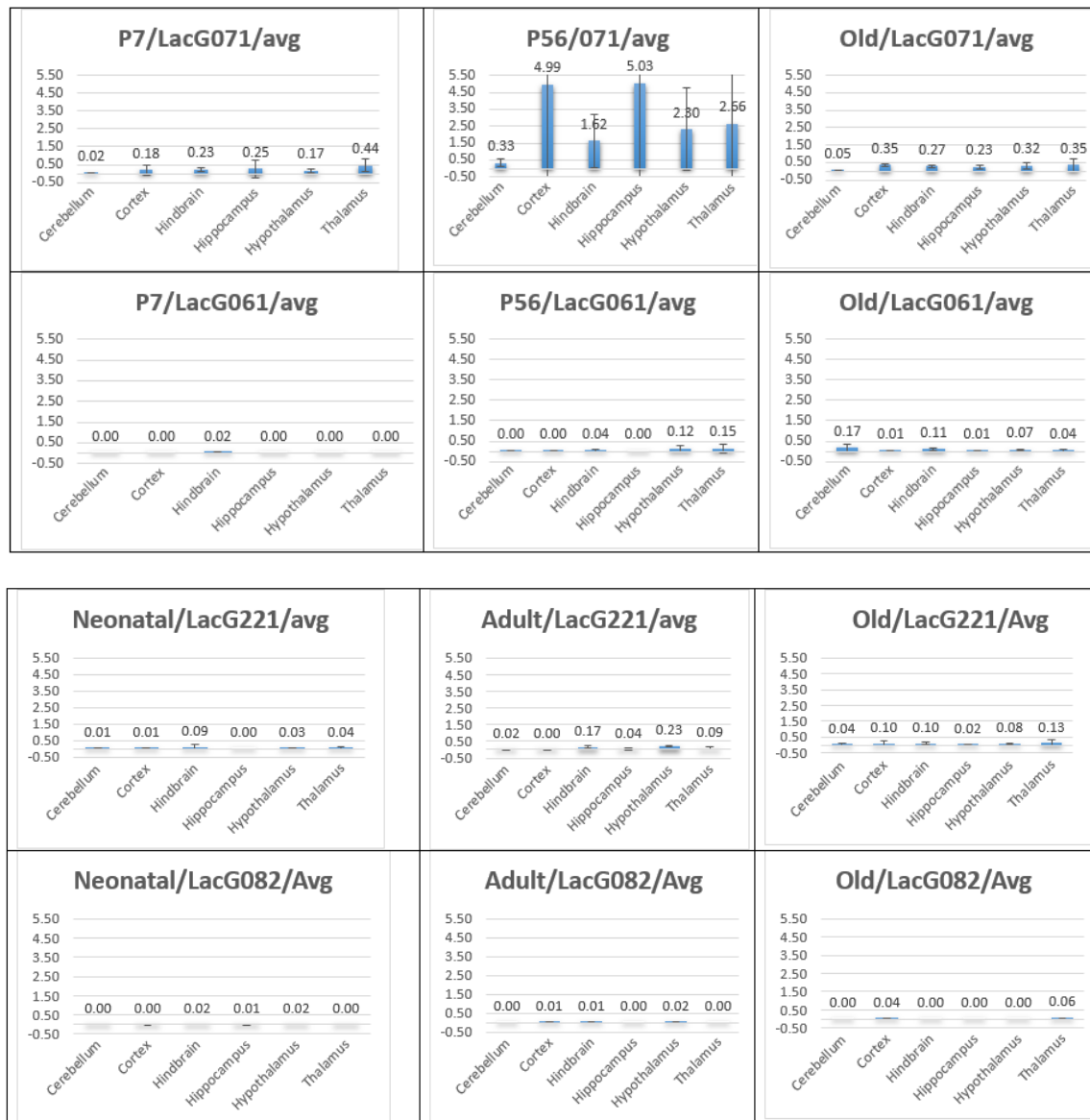
### 2.4.6.3.4 In search of a common region(s) with the most expression:

Among the four lines, LacG071, LacG061, LacG082 & LacG221, it was pertinent to know which brain area had the most expression and also whether any particular region had consistently most transgenic expression. The table (**Table 2.7**) below summarized the histograms (**Fig. 2.27**), showing average promoter activity (%) among three time points across 4 different random lines.

Sublines	Neonatal	Adult	Old
<b>LacG071</b>	Thalamus (0.44%)	Hippocampus (5.03%)	Thalamus & Cortex (0.35%)
<b>LacG061</b>	Hind brain (0.02%)	Thalamus (0.15%)	Cerebellum (0.17%)
<b>LacG221</b>	Hind brain (0.09%)	Hypothalamus (0.23%)	Thalamus (0.13%)
<b>LacG082</b>	Hind brain & Hypothalamus (0.02%)	Hypothalamus (0.02%)	Thalamus (0.06%)

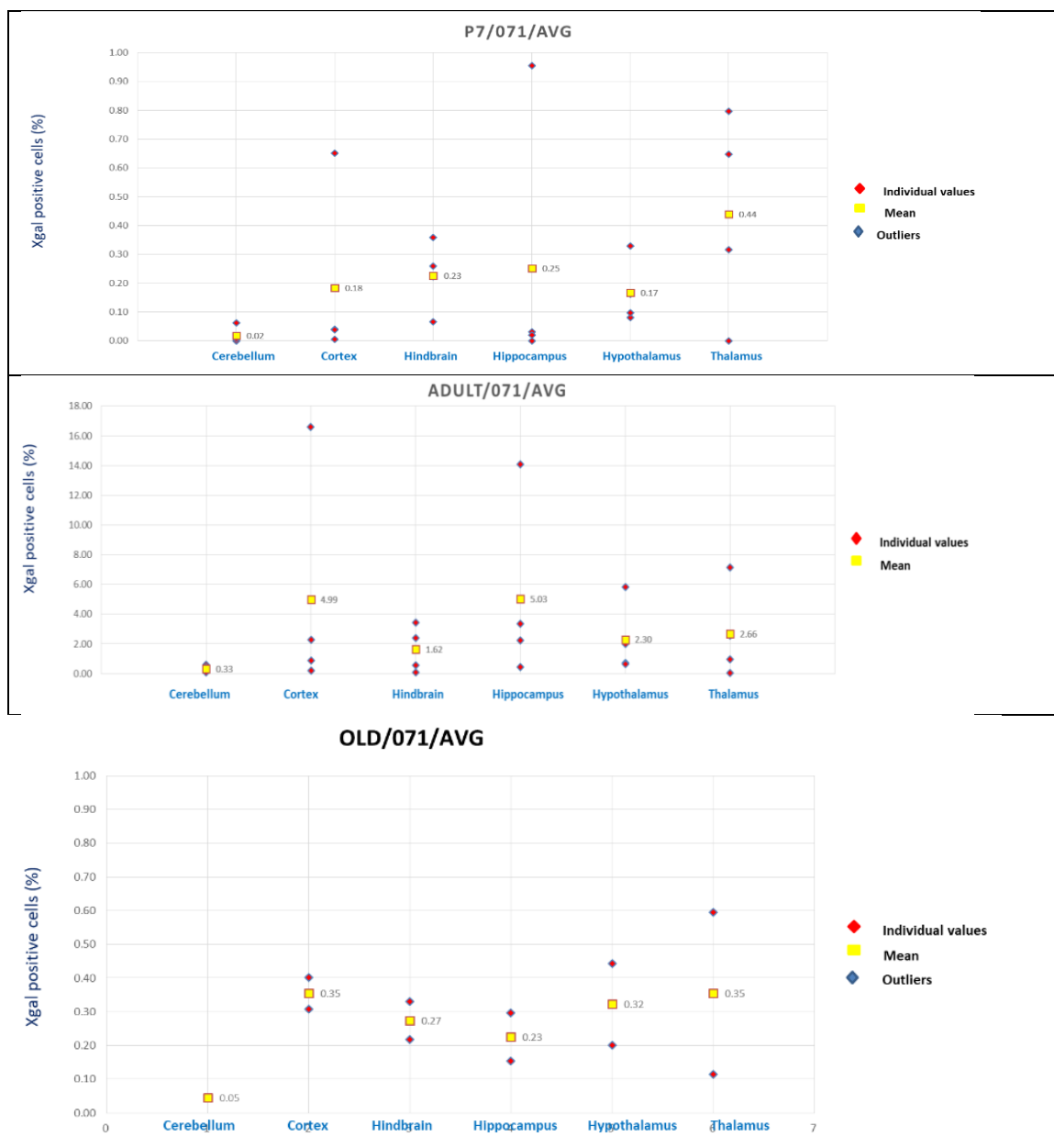
**Table 2.7** The mouse brain regions with the highest promoter activity in four different lines at three developmental time points

Here we see, out of 5 unique regions (hindbrain, hypothalamus, thalamus, cortex, and cerebellum) of 4 different lines, the thalamus is seemingly the most common (as it appeared in 5 cases out of 14), a preferential region of promoter activity of all the time points, whereas hypothalamus and hindbrain take jointly 2<sup>nd</sup> position with 3 appearances, and cerebellum, hippocampus and cortex for 1 time each only. N522 animal particularly showed a high number of positive cells in most of the regions of its brain. Cerebellum showed consistently the lowest level of expression in all-time points, except old time point in LacG061. Hippocampus, though appeared one time but carries the highest value (5.03%) among all the regions been compared. Therefore, it might be the less preferential region of promoter activity. Here, a high standard deviation among the expression values was observed. Figure 2.28 is showing the data distribution of the animals of LacG071 sublines in three time points.



**Figure 2.27:** Histograms showing average promoter activity (%) (in the y-axis) in different brain regions of mice belonged to four different lines at three different time points

**Note: Alternative graphs to observe each data point belonged to LacG071:**

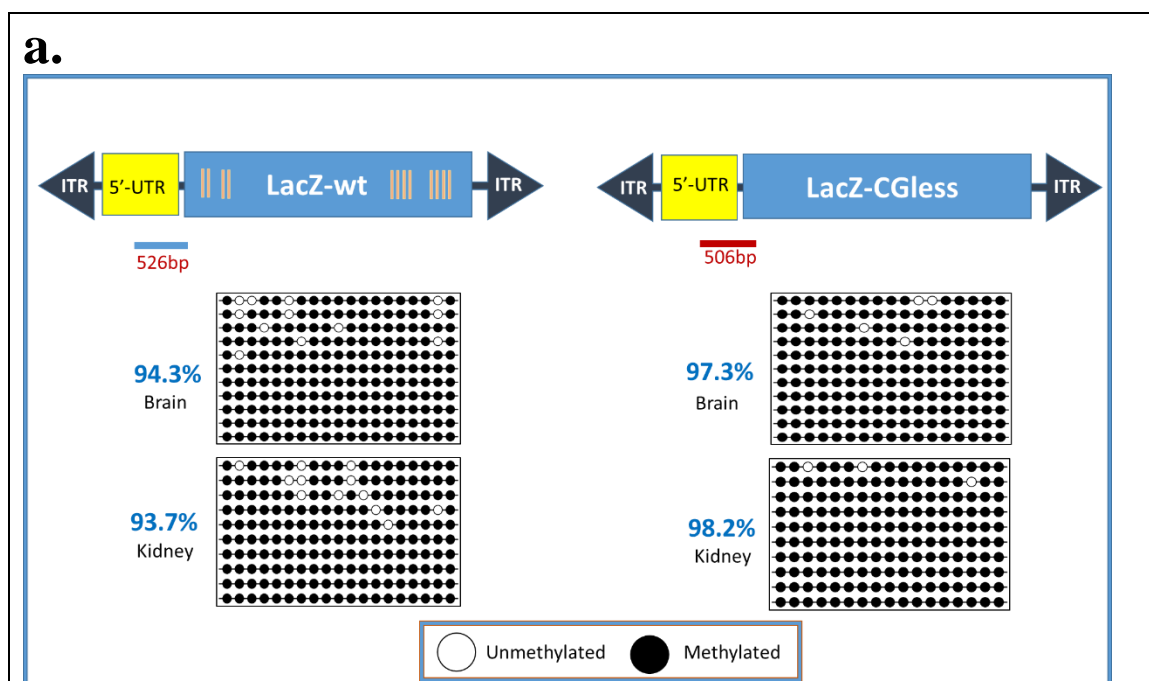


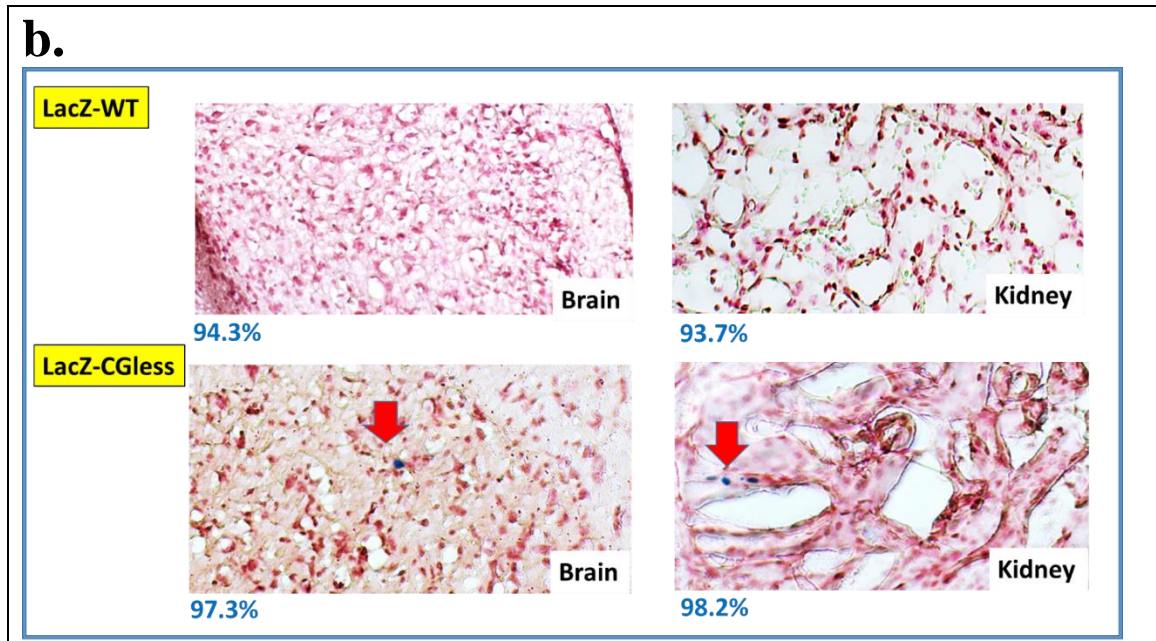
**Figure 2.28** The variation of the values of different brain regions of LacG071 line.

## 2.4.7 Promoter regions of LacZ and LacG transgenes are hypermethylated

Two different tissue samples (brain and kidney) of one animal from each line were extracted for the DNA. The DNA was bisulfite converted, and the methylation of a part of

promoter region DNA was analyzed. The brain and kidney samples of LacZ line showed respectively 94.3% and 93.7% of DNA methylation. While LacG line showed comparatively higher methylation status of 97.3% and 98.2% in brain and kidney, respectively (**Fig. 2.29a**). Figure 2.29b shows that even after having so much methylation at the promoter site, the tissues of the sample animal showed expression in LacG line, not the LacZ line.





**Figure 2.29** The bisulfite sequencing data for DNA methylation analysis. (a) Methylation levels of part of the promoter of both transgenes & (b) transgene expression in the tissues.

## 2.5 Discussion

Vast sequence homology of endogenous LINE-1s barred us from using an appropriate technique to harness the locus-specific information of L1 promoter activity. Here, we utilized a transgenic approach to know the locus-dependent expression pattern of LINE-1 promoter activity.

Unlike A and F subfamilies,  $T_F$  is the youngest subfamily of mouse LINE-1. With an abundant expression in the mouse genome, it carries 4000-500 full-length members. Due to having a large number of active members, it is an expanding sub-family of mouse LINE-1.  $L1_{spa}$  is a name of a full-length LINE-1 insertion caused by this  $T_F$  subfamily, and this  $L1_{spa}$  cause disease in mouse (Kingsmore *et al.*, 1994; Takahara *et al.*, 1996; Naas *et al.*, 1998). Therefore, we wanted to characterize the promoter activity of this  $L1_{spa}$ .

Therefore, we fused the 5'-UTR promoter from this L1spa sequence and attached to LacZ/G reporter to make our transgene) (**Fig. 2.1a**). With the help of pronuclear microinjection procedure, we were able to integrate our transgene into random chromosomal sites. Later, with the help of histochemical detection of the reporter expression, we quantified the promoter activity of the LINE-1 promoter activity. In most of the sublines and organs, we found a silencing effect of the transgene.

There are several general reasons for gene silencing. Due to the influence of the sites of the chromosome, the stably integrated transgene are often poorly expressed (Bestor *et al.*, 2000, Fiering *et al.*, 2000). Several other factors are attributed to the variation of transgene expression, include the differences in copy number and location of the chromosome, and fidelity of the transgene construct itself (Matzke & Matzke, 1998; Kooter *et al.*, 1999) as well as an epigenetic factor, like position-effect variegation (PEV) and DNA methylation. It has already been shown that LINE-1 transcriptions have been influenced by the upstream gene sequences (Lavie *et al.* 2004). Additionally, promoter methylation is one of the strongest candidates for gene silencing. In general, DNA methylation can also be inherited in subsequent generations (Balow, 1995; Schumacher *et al.*, 2000; Kearns *et al.*, 2000). Other candidates for gene silencing could be histone modifications or other epigenetic factors (Kearns *et al.*, 2000; Mehta *et al.*, 2009).

In general, the position effects can be observed in two types *in vitro*, stable and silencing. In the stable category, the transgene expression is obtained pervasively in most of the cells in a similar manner. This expression will be essentially different from the similar transgene or endogenous gene, being embedded in a different site. On the contrary, the silencing



category involves with the heterocellular expression, where except being expressed in few cells, the transgene is silent normally silent (Pikaart *et al.*, 1998; Walters *et al.*, 1996).

Nevertheless, a suitable detection technique to report promoter activity was needed. For the detection techniques, it was vital to opt for a sensitive technique, which should be equally sensitive and quantitative enough for *in situ* signal detection as well as having a robust screening efficiency with less experimental variations. Thereby, we chose LacZ reporter assay or X-gal staining.

Originally, the LacZ gene encodes  $\beta$ -galactosidase in *Escherichia coli*. This enzyme can also hydrolyze different synthetic (chromogenic and fluorogenic) chemical substrates. Thus, X-Gal, a colorless chemical, is cleaved by this enzyme to produce a blue-colored compound, 5-bromo-4-chloro-indole, which is easy to detect in an *in situ* screening procedure. Therefore, LacZ fusion with a promoter of interest could be used to characterize the activity of a promoter in cell lines or can be used to generate a transgenic mouse line. Over 30 years ago, the first transgenic LacZ mice were generated (Kothary *et al.*, 1988; Kothary *et al.*, 1989). Since then this versatile reporter gene has been used to create several mouse models to characterize the expression of genes and also to trace the cells lineage during development. This reporter assay provides an indirect measurement of promoter activity, yet it is a highly straightforward, time-saving, and reproducible, requiring no expensive reagents and equipment.

However, bacterial-origin LacZ gene shows the CpG content of 9.24% (Chevalier-Mariette *et al.*, 2003). The methylation of 5'-CpG-3' dinucleotides within genes mostly creates potential targets for the protein complexes that bind onto methylated DNA sequences and also to histone deacetylases. These bindings could lead to transcriptional repression by

modifying the chromatin landscape. There is evidence as methylation of non-promoter sequences could result in transcriptional silencing of reporter genes (Kass *et al.*, 1997; Hsieh *et al.*, 1994).

Hence, to avoid the silencing of the promoter activity due to the methylation of the gene body containing the reporter cassette, a CpGless version of the CpG-rich LacZ gene (i.e., LacG) was engineered by eliminating the CpG dinucleotides. This class of CpGless-LacG construct was previously designed by Henry *et al.* in 1999 (Henry *et al.*, 1999), and the *in vivo* activity in late-stage of developing mouse was not realized in wide-range of mouse organs until 2003 by the same group (Chevalier-Mariette *et al.*, 2003). There, widespread repression of CpG-rich *LacZ* transgene was observed even at single copy level in all somatic tissues, whereas substantial expression was acquired with the CpG-less *LacZ* transgene in contrary (Chevalier-Mariette *et al.*, 2003).

Since we did not have any data how the transgene should function *in vivo*, therefore initially, we tried to understand the difference of X-gal staining of these two transgenic constructs with the help of *in vitro* transfected cells, embryonic fibroblast cell (3T3 cell line). We found that 5'UTR-LacZ transgene showed less expression than the 5'UTR-LacG (**Fig. 2.1b**). However, we created two mouse lines to understand the difference between their *in vivo* regulations.

Importantly, when the transgenes (5-UTR-LacZ and 5'UTR-LacG) were constructed, we also included an important design in which both transgenes were flanked by two inverted terminal repeats (ITRs) from the Sleeping Beauty (SB) DNA transposon. The utility of this design will be described in the text below. These plasmids were injected into the pronuclei

of fertilized oocytes by Cyagen Biosciences. Subsequently, founder mice carrying multi-copies of either transgene in a tandem array within a single locus were generated (**Fig. 2.2**).

As explained below, the embedding of SB-ITRs in the transgene would allow us to derive an unlimited number of transgenic mouse lines carrying a single copy of the transgene. The rationale for using single-copy instead of multi-copy transgenic lines to profile L1 promoter activity is two-fold. First, endogenous L1s, by definition, are interspersed and present in the mouse genome as discrete, individual copies, not in the form of tandem arrays. Second, it is known that tandem arrayed sequences subject to additional mechanisms of transcriptional regulation. This phenomenon has been previously documented for transgenes and referred to as “repeat-induced gene silencing (RIGS)”. Silencing of transgenes in animals with a high copy number is a regular feature of pronuclear microinjection method of animal transgenesis. Expression of large tandem arrays of repeated sequences may suppress the efficient transcription of transgenes (Simon & Knowles, 1993; Dorer & Henikoff, 1994). Indeed, our lab has observed RIGS for a tandem repeated L1 transgene carrying a heterologous CAG promoter (Rosser & An, 2010). Therefore, it was essential to reduce the number of copy number of transgenes. At the same time, we wanted to mobilize the single-copy transgenic cassette to random loci in order to create unique locus-specific mouse lines. These two purposes were resolved with the help of the Sleeping Beauty (SB) DNA transposon system.

Sleeping Beauty (SB) transposon system is a novel genetic tool that was developed around two decades ago (Ivics *et al.*, 1997). It consists of two components: first, an excisable DNA sequence flanked by two ITRs, which are the essential sequence for the recognition and

mobilization by the SB transposase, the second component. Thus with the help of the latter, the former component is eventually excised and reinserted into other locations of the genome (Ivics *et al.*, 1997).

To obtain single-copy germline insertions for our transgene, the donor animal was bred with H1t-SB100X positive transgenic animals (Grandi *et al.*, 2015), expressing a hyperactive SB transposase specifically in pachytene spermatocytes (Mates *et al.*, 2009). Therefore, the excision of the transgene and their successive mobilization happened in male germ cells, carrying both H1t-SB100X and the L1 transgene).

The progenies of this male with a wild-type female possess differential copy numbers of transgenes. The transgenic mice with low-copy or single-copy number were identified, using real-time PCR and droplet digital PCR, respectively. Next, the loci of these transgenes were mapped (**Fig. 2.3**), and primers were designed for locus-specific PCR. On some occasions, when the transgenes moved further due to the presence of H1t-SB100X in the transgene positive animals, the mobilization was detected with the help of the result of locus-specific PCR. In those cases of a new position, the transgene was again mapped (**Fig. 2.4**).

X-gal staining is a fast and convenient histochemical technique to detect the expression of the LacZ reporter gene. The bacteria (*Escherichia coli*) derived LacZ gene encodes the  $\beta$ -galactosidase ( $\beta$ -gal) enzyme.  $\beta$ -gal can hydrolyze different synthetic substrates. For example, X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside), a synthetic colorless  $\beta$ -gal substrate. It can be enzymatically cleaved by  $\beta$ -gal into galactose and 5-

bromo-4-chloro-3-hydroxyindole. The latter compound gets further oxidized into 5, 5'-dibromo-4, 4'-dichloro-indigo, which is blue. In sum, cells expressing the LacZ reporter can be visualized with X-gal staining.

In this study, the wild-type LacZ and its CpG-less derivative LacG were placed under the control of endogenous promoters for LINE-1, i.e. 5'UTR. The expression from them was marked by a dark blue stain, which was detected and quantified at the single-cell level, providing a robust visual readout of LINE-1 promoter activity in the main organs of developing mouse.

It is already known that many mammalian tissues synthesize endogenous  $\beta$ -gal (GLB1 gene product). This is a physiologically significant lysosomal enzyme that helps in the enzymatic degradation of glycolipids in some of the tissues. Kidney, intestine, and epididymis are some of these tissues, which are a particularly rich source of endogenous  $\beta$ -gals in mammals, although their presence may vary species to species (Conchie *et al.*, 1958; Pearson *et al.*, 1963).

Lysosomes have an acidic environment within it, and therefore, consistent with its localization lysosomal  $\beta$ -gal exhibits maximal activity within the range of pH 4.0 and 4.5 but significantly lower activity at pH 6.0 or higher (Zhang *et al.*, 1994). Importantly, proliferating cells have undetectable  $\beta$ -gal activity *in situ* with an X-gal staining buffer at pH 6.0 (Kurz *et al.*, 2000). Yet, in senescent cells, the GLB1 gene product was reported to be the origin of senescence-associated  $\beta$ -gal activity (SA  $\beta$ -gal) (Lee *et al.*, 2006). Therefore, if X-gal staining is found at or higher pH 6.0, it should be either originated from senescent cells with increased lysosomal  $\beta$ -gal activity or transgenic b-gal. Therefore, it is a technical challenge for X-gal staining to distinguish between these two types of  $\beta$ -gals.

Different modifications to the X-gal technique were previously adopted by various groups to increase the specificity of exogenous LacZ detection in respective experimental setups. These included exposing LacZ tissues to X-gal at a point below 37°C (Weber- Benarous et al, 1993; Sanes et al, 1986), or utilizing alternative chromogenic substrates as well as the fluorescent substrates of X-gal compound (Weis et al, 1991; Liu et al, 1996, Aguzzi and Theuring, 1994; Pearson *et al.*, 1963; Brunet *et al.*, 1998; Kishigami *et al.*, 2006, Zhang *et al.*, 1991). However, none of these methodologies is widely used. Here, we utilized two different strategies to enhance the specificity of X-gal detection.

First, to enhance the specificity, the reporter gene, carrying a nuclear localization signal that results in producing nuclear-localized  $\beta$ -gal rather than cytosolic forms, can also be utilized to distinguish between endogenous and exogenous signals (Bonnerot *et al.*, 1987). Therefore, in our transgenic construct, we embedded a sequence of nuclear localization signal as part of the coding sequence. When the cryosections of the organs were treated with X-gal staining solution, we observed the X-gal signals being colocalized with neutral red (**Fig. 2.5.a**), a planar phenazine dye, which found to interact with DNA (Wang *et al.*, 2003; Huang *et al.*, 2001) in the nucleus. In addition to that, we validated our X-gal signals with immunostaining. Here, we used fluorescently labelled antibodies against lamin B, a protein that is anchored to the inner nuclear membrane. The antibody fluorescently marks the circumference of nuclei. Indeed, the X-gal stains, blue under the visible light source, were found to be surrounded by lamin B signals (**Fig. 2.5.b**). Therefore, these results confirmed that the observed X-gal signals were solely derived from the transgene.

Second, we used an optimized pH of the X-gal solution to rule the chances of detection of the X-gal signals out of endogenous  $\beta$ -gal, which, as mentioned earlier, were generally

active in acidic pH and inactive in high pH. To do so, we had to select an exact pH point from a relatively high pH gradient (7.5, 7.7 and 8.5). We observed that (**Fig. 2.6**) at pH 7.7, non-specific signals are eliminated in non-transgenic, adult mouse kidney. This condition also maintained substantially intense X-gal signals in the positive control Z/EG mouse. Z/EG (LacZ/EGFP) mouse line is a double-reporter mouse line. The reporters are driven by a strong CMV early enhancer/chicken beta-actin (CAG) promoter. A LacZ gene, one of the reporter gene, is flanked by two loxP sequences and also followed by enhanced green fluorescent protein (EGFP) sequence (**Fig. 2.6**). Cre-mediated recombination removes the LacZ gene, and thus activates expression of the second reporter. Nonetheless, the CAG promoter maintains the expression of LacZ throughout its all developmental stages by default (Novak *et al.*, 2000). In this control mouse line, the LacZ is not embedding a nuclear localization signal; as a result, the X-gal signals mostly come from the cytoplasm, unlike the reporter mouse lines generated for this study. Moreover, this signal mostly found to diffuse to form a smear over the surrounding cells.

Later on, we optimized drop fixation condition also depending upon the mouse developmental time points to get a substantial signal intensity and a better morphology (**Fig. 2.7 & Fig. 2.8**). Multiple tissue sections from the both of major transgene lines were screened for quantification of the X-gal signals.

In the LacZ line, a very small percentage of neonatal animals showed positive signals in no more organs than the brains (20%) and kidneys (30%) only (**Fig. 2.11**). The highest detectable signal obtained in the brain was this clearly shows that at this time point, the expression of LacZ transgene was silent. Therefore, not much information regarding other organs can be harnessed with the help of this transgene construct for this time point.

However, interestingly, all animals with the detectable expression values in the kidney belonged to one particular locus. Similarly, 2 out of 3 detectable brain values are also confined to another unique locus. Although the sample numbers are not large enough, still it might be suggesting that with taking age as a factor, expression from this transgene is dependent on the locus point where transgene is placed in the genome. It is also true that the detectable values are very rare in both brains and kidneys. The highest detectable value in the neonatal kidney was recorded as 6 positive cells in 10,000 cells. Upon transitioning to adult time point, we observed more number of organs showed detectable values outside the brain and kidney. At the geriatric time point (12 to 18 months old mice), X-gal stained cells were mostly found as shown in the scattered plot. This is a situation opposite to the neonatal time point. Here, rather, 20% of the brains showed detectable signals, although the detectable signals were of low range i.e. 0.001 - 0.04%. At this time point, besides the brain, kidney and testes showed a locus-dependent expression pattern with relatively higher detection values compared to other organs, like spleen, lung, liver and heart. The kidneys of the old animals showed more number of expression particularly. This might be an age-dependent, tissue-specific change of DNA methylation (Spiers *et al.*, 2016; Stubbs *et al.*, 2017). Moreover, in this case, especially, this might be due to the presence of accidental detection of senescence-associate beta-gal. If not in all cases, but in some cases, age-associated glomerulosclerosis might have influenced the methylation pattern (Davies *et al.*, 1989; Hackbarth & Harrison, 1982). In the kidney, the positive cells had no consistent pattern, rather scattered throughout the cortex and medulla. In some cases in brain and kidney, positive cells were found to exist in the cluster, with having no positive cells in the vicinity. These clusters of X-gal expressing cells might have suggested clonal expansion



originating from a committed progenitor, of course not let alone the other possibilities. Yet, the X-gal positive cells lied at the border of the seminiferous tubules in testes. They are believed to be somatic Leydig cells or Sertoli cells, and need further confirmation. However, this line is not suitable for further screening. First, it had a very low range of detection (**Fig. 2.12**). For instance, the highest level of signal detected at all time point is less than 1 per 100 cells. Second, around 90% of animals in all time points showed no detectable values in heart, liver and lungs. Third, only one time point i.e. the aged mice showed more number of detectable signals in most of the organs **Fig. 2.13**. However, this transgenic line helped us to understand the locus dependency of expression in certain cases. Certainly, there is an observable difference of expression between neonatal and old-time points, at least in kidney and testes. This contrast of expression between these two developmental time points helped us to understand the notion of age-dependent expression. The detection values are expressed in percentage; therefore, we do believe that the low detections in overall is not due to difference of the sections' surface areas in between two different time points. Finally, the cases where a detectable littermate of non-detectable animals hinted us regarding the variation due to individual expression pattern.

However, mechanistically we tried to understand the reason behind the difference of expression of the LacZ and LacG construct whether it is because of the methylation of the promoter region. We did bisulfite sequencing to analyze the level of methylation present in a part of the promoter regions of the gDNA of the mice of both lines. We found that the promoter regions were highly methylated in both cases. More specifically, LacG had more methylated than the LacZ line. Despite this difference, we observed some positive cells in

LacG line, unlike LacZ cells (**Fig. 2.29**). This tells us that there might be some other reason involved in the silencing of the transgene, than the methylation alone.

Therefore we had to rely on other LacG-CGless lines, with several mapped sublimes, for the expression study. Each of these sublimes has transgene placed to a unique locus point in the genome of the mouse. Similarly to the study with the LacZ-WT, we screened the tissues from mice of three different developmental time points. Unlike the LacZ line at a neonatal time point, where the highest detectable value in the brain was 0.0038%, the lowest detectable value was 0.003% for this line. Surprisingly, the highest obtained percentage of X-gal signal in the neonatal brain for this line was 0.6% (**Fig. 2.14**). Although it varied among the individual animals, the LacG071 line showed the highest expression among all the sublimes present in this LacG-CGless line. In the case of heart, although we obtained the highest amount of expression, these signals were mostly non-nuclear localized. This is very unusual as because the beta-gal expressed from our transgene should be localized in the nuclear instead. However, we believe these signals are from exogenous beta-gal, not from the counterpart, because we did not observe any expression in non-transgenic animals. In line to support a belief, we can put forward two pieces of evidence of the detection of L1 activity in a human sample. First, in a study in post-mortem human heart, the presence of Orf1 protein was detected highest among the non-brain samples (Sur *et al.*, 2017). In another study, in an attempt of detection of full-length L1-mRNA on human heart muscle with northern blot assay, the relative expression was found to be 200%, which was again one of the highest detection in somatic samples (Belancio *et al.*, 2010). Therefore, it is evidence that mammals have LINE-1 expressed in the heart. Here although,

the morphological entity of being expressed outside the nuclei of the cardiac cells was not understood well.

Unlike the LacZ-WT line, liver and lungs showed detectable values, accounting for more number of organs showing up expression at the neonatal stage. The difference of the transgenes must be a reason why this transgene is detectable in terms of detectable positive signals. Testes and spleen, like LacZ-WT line, however, did not show any detectable values. This perhaps indicates further age-dependent silencing of the promoter in these two organs.

Compared to the neonatal time point, we observed certain changes in the adult and aged time points (for all sublines). Firstly, in comparison to the highest of neonatal brain value (0.6%), the adult had a peak of ~2% and old animals had a peak of ~4% (**Fig. 2.15** and **2.16**). These three values were from three different sublines. Moreover, limited sample number restrict us to claim a significant correlation of this upward trend with the age. High expression levels compared to LacZ-WT line was observed. Nevertheless, this result is consistent with another study on the human sample. In that study, a similar age-dependent L1 activity was observed by Sur et al, where they compared the ORP1p expression among the 15, 55, 80-year-old post-mortem human brain samples (Sur *et al.*, 2017). They observed almost no expression in the sample from 15-year-old; the intermediate value was obtained from the 55-year old brain, and the highest value was obtained in the sample from 80 years old (Sur *et al.*, 2017). Secondly, average values dropped in case of adult and geriatric lungs and livers from the high neonatal counterpart. However, the average values in kidneys of adult and aged time points did not change in comparison to that of neonatal kidneys, and values of spleen rose from undetectable to detectable during neonatal to adulthood, but no

change was observed in old animals. In the case of testes, however, a significant increase of values observed in transitioning from adulthood to old-time point, whereas there were no detectable values in case of neonatal time points. It might be due to gonadotropic hormones-regulated hypomethylation in testis, associated with the progression of the age of the rodents (Reddy *et al.*, 1990). Apparently, similar to adult, the tissues can be categorized into relatively “high” (brain, heart, kidney) and “low” (spleen, lung, liver) in old animals. Surely, a different organ-specific regulation, aided with the age factor, is observed in all lines.

Instead of concluding the expression pattern for every line concerning age, we tried to find out a specific organ, which is influenced mostly by different locus positioning of the same transgene. To understand that, we tried to compare the highest average expression values from the different organs of four (commonly used) sublines of animals of three different time points. Notably, heart, even though having the highest expression values, were ruled out from the test. This is because the expression pattern of heart as discussed previously is not well understood. Therefore, we had to confine our test with only six organs in total. Therefore, among these six organs, four organs (brain, lung, kidney and testes) had the highest level of mean expression values. Surprisingly, kidney most frequently showed high expression values - at least once in all three time points of the animals belonged to all four lines. The notion of the frequent occurrence of the kidney might be comparable to that of high expressive kidneys at the old stage of LacZ-WT strain. Similar reasons should be echoed here. However, too little data is available on mouse kidney with a mouse-derived promoter for supporting our finding. However, with the help of quantitative PCR (qPCR), Kano et al measured the de novo retrotransposition in several adult tissues of LIRP mouse,

a transgenic mouse carrying L1RP element (a highly active human L1) followed by an enhanced green fluorescent protein (EGFP) cassette (Ostertag et al. 2000). In adult tissues of brain, lung, liver, kidney, and tail, they found a persistent occurrence of kidney in all the tissues (Kano *et al.*, 2009). It is true that since this is a data from a human retrotransposon, a little can be extrapolated about the regulation of a mouse promoter activity. Therefore, it can be claimed here regarding our data that kidney could be a preferential place of high endogenous 5'-UTR promoter activity in a mouse. However, further assays might be helpful to establish this claim.

It is also should be noted as this point about three outlying values which were eventually opted out to determine the mean value. More specifically, these all three outliers were from old animals, and out of three, 2 of them were in kidneys of two different mice of two different sublines (LacG071 and LacG061). Rest one value comes from the brain of different individual animal of LacG071 subline. These outliers might have caused by the cell-specific epigenetic dysregulation or some confounding senesce associated beta-gal escaped from pH control due to profound presence. Regardless of high inter-individual variation among the values, this test also empowered us to see the distribution of the organ-specific quantification values. This indicated us LacG071 and LacG082 sublines as the most and least expressive lines, respectively. Most of the sublines screened were found to be less expressive, except LacG071. This together shows that there is a presence of locus-dependent expression.

By far, it was evident that the presence of the transgene in different locus is associate with differential expression of the promoter activity. Another line of evidence comes in favor of locus-dependent expression when we observed a drastic change was observed when

transgene moved from chromosome 15 (subline LacG066) to chromosome 14 (subline LacG221) within one generation apart (**Fig. 2.19**). LacG066 showed the highest amount of expression ever in its brain and kidney among all lines used for the collection of tissue sections. Unfortunately, the line of LacG066 could not be continued, but the tissues of the old, founder animal were collected and compared to the corresponding tissues of several old transgenic mice of subline LacG221. In compared to LacG066, tissues from LacG221 line had sporadic incidences of expression for all organs. Although we have reported previously the incidences, where inter-individual differences of expression were present within an organ of an animal of particular subline and age, the ~20 fold difference in kidney and ~4 fold difference in the cerebellum was a strong indicator of the crosstalk between the promoters and the surrounding chromosomal environment. This certainly informs how L1 promoters function concerning location on chromosomes.

Besides locus dependency, organ dependent promoter activity might be an obvious phenomenon. As we have seen that LacG071 is the subline with high promoter activity in terms of positive X-gal staining cells, we would like to take a closer look regarding the expression pattern in the organs of it in different developmental time points. In short, we found an organ-dependent expression differential pattern, where different organs expressed during different time points. These organs can be categorized into three different classes, namely ‘no’ expressing (spleen), ‘low’ expressing (liver, lungs), and ‘high’ expressing (brain, heart, kidney, and testes) (**Fig. 2.20**). Similar classification can be obtained from other LacG sublines at all points of development (**Fig. 2.14 to 2.16**). One of the best examples of organ-dependent expression was observed in spleen and testes. At the neonatal time point, the X-gal values in these organs were in the undetectable range. However, in

the two later time points, we observed them in the detectable range (**Fig. 2.14 to 2.16**). We also found that the kidney, compared to other organs, showed to as a preferential organ for high promoter activity (**Table 2.6**). Overall, it indicated that there exists an organ dependent promoter activity.

We also found an age-linked promoter activity. The first evidence was the incident when we observe the values of neonatal spleens and testes moved from undetectable to detectable range in later stages of development (**Fig. 2.14 to 2.16**). The second evidence comes when we see that the brain showed a peak activity during adult time point among other points (**Fig. 2.22**). There might be different known or unknown scientific causes to address this observation. Here, we would like to explain this concerning the abundance of sex hormones. There is strong evidence that sex hormones powerfully modulate the neurogenesis in both males and females (reviewed in Mahmoud *et al.*, 2016). For instances, long-term exposure to testosterone increased hippocampal neurogenesis in adult male rodents (Galea *et al.*, 2013). Again, adult hippocampal neurogenesis was essentially found to be regulated with estrogens (Pawluski *et al.*, 2009). In a rat study, adults showed the highest level of testosterone than the young and old rats (Ghanadian *et al.*, 1975). This suggested young rodents do not have enough sex hormone present. Evidence are there that both sexes of rodents have drops in sex hormones are age progresses (Ghanadian *et al.*, 1975; Morley and Perry, 1999). It is already known also that the LINE-1 activity is a prominent phenomenon during neurogenesis and neuronal differentiation (Coufal *et al.* 2009; Muotri *et al.* 2005). Taking together, we can support our data, saying that LINE-1 promoter activity should be highest during the period of highest neurogenesis. Sex hormones are either not developed or its level dropped during neonatal and old-stage,

respectively. Therefore, these two time points most likely showed low promoter activities. On the contrary, whereas sex hormone level is at its peak during the adult stage showed high promoter activity. The values in kidneys showed the highest variations. Both neonatal spleen and testes did not show any promoter activity until adulthood. Unlike spleen, in the values of testes showed an increasing trend from the adult stage to the old stage might be based on the increased hypomethylation as stated earlier.

The brain is a classic hub of retrotransposition, where LINE-1 is present in mosaic form. In most of the LacG sublines, except LacG071, we observed very low promoter activities across different brain regions. In LacG071 animal model, we also saw an approximate frequency of one X-gal positive neuron per 10000 of neurons. Besides, we observed an age-dependent change of brain region with high expression; in some cases, a sharp rise and fall of activity observed across different time points. For example, an elevated expression was found in the thalamus during neonatal age, but it shifted to the hippocampus for adult time point. In some animals of three time points, an expression pattern in the dentate gyrus was visible (**Fig. 2.23**). Overall, it suggested to us that there is an observable brain-region specific activity for our promoter, which is also age-linked.

In adulthood, the hippocampus has been reported to be associated with the neurogenesis (Eriksson *et al.*, 1998). The alterations to the hippocampal neurogenesis have been observed in post-mortem tissues of patients with severe depression (Boldrini *et al.*, 2012) and also Alzheimer's disease (Crews *et al.*, 2010, Jin *et al.*, 2004). And a higher copy number of LINE-1 was found in the hippocampal dentate gyrus (DG) (Coufal *et al.*, 2009; Baillie *et al.*, 2011). Therefore, this finding in our transgenic model is relevant to the findings in the works of literature.



A mouse, an exception, with very high values (>10% positive cells) showed in several parts of its brain. This might be possible either by a persistent global hypo-methylation in all organs or due to the effect of some unknown health issue. Alternatively, this example might also indicate the effect of the environment on the LINE-1 promoter activity. In another example, during the old-time points, brain expression of hippocampus and thalamus fell from their peak state during adulthood. This is quite obvious as the hippocampus is associated with keeping memory, and thalamus is associated with sensory integration of information and also with alertness. Sur *et al.*, immune detected LINE-1 activity in the thalamus of a geriatric human (Sur *et al.*, 2017). However, the existence of only two samples might not tell all. Overall, during the adult time point, most of the brain regions are elevated, and cerebellum remained with low profile among all the brain regions. Muotri *et al.*, observed mobility in this region for the first time. It surprised them as this area is a non-neurogenic. A mobilization happening here would potentially be suggested that LINE-1 transposition occurred in post-mitotic cells, or the cell born elsewhere but ended up migrating to this location (Muotri *et al.*, 2009). Another important observation is that at the individual animal level for the adult animals, all brain regions tracked together, at a higher range or lower range (**Fig. 2.22**). That might suggest that level of expression is a characteristic at the individual level. In other words, if any mouse had high expression in a region of brain, high likelihood that other brain regions of the same will have high promoter activity too.

Next, we tried to find out whether any specific brain region has commonly been influenced by the promoter activity by the different locus positioning of the same transgene. To understand that, we compared the highest average expression values from the different

brain regions of four (commonly used) sublines of animals of three different time points. Among these six brain regions, four regions (hindbrain, hypothalamus, thalamus and cerebellum) appeared with the highest level of mean expression values. Surprisingly, thalamus most frequently showed high expression values - at least once in all three time points of the animals belonged to all four lines. It is also should be noted at this point about two outlying values, from a particular adult animal belonged to LacZ071, were opt out to determine the mean values from cortex and hippocampus. These outliers might have caused by the cell-specific epigenetic dysregulation. Regardless of high inter-individual variation among the values, this test also empowered us to see the distribution of the organ-specific quantification values. This indicated us LacG071 and LacG082 sublines as the most and least expressive brain regions, respectively. This highlighted us with a clear locus-specific expression. And the cerebellum is the lowest expressive brain region, except in one incident of the brain of one old LacG061. This again indicates for a brain region-specific promoter activity. This might be a cell-specific incident and can be addressed with the help of single-cell analysis.

In this study, we would like to observe how much reproducibility the transgene is expressed when it is placed in the same genetic locus of the different mouse. On the other hand, we also wanted to know how transgene is expressed being placed into distinct regions in different integrant mouse lines. In the first observation, we saw that upon introducing the transgene in the same locus, the mice were capable of producing a more or less similar level of signals in littermates as well as in the progenies (for example, LacG072 line). We also observed a profound gene silencing, which gave rise to the mosaic gene expression pattern.

The eukaryotic genome has different structural variance due to insertions, duplications, deletions, and rearrangements. This structural variance can affect the gene expression of the transgene (Lydiard-Martin *et al.*, 2014). Chromosomal localization of transgene at the proximity to centromere or any other heterochromatic region may contribute to the gene silencing. Chromosome locations which promote transgenic expression are considered as the transcriptionally active sites of euchromatin. The heterochromatin sites are inaccessible to the transcription factors and are often correlated with cytosine hypermethylation and also with histone hypoacetylation (Ng and Bird, 1999). As the heterochromatin integrate itself to the proximity of this chromatin architecture, it shows the variability of the transgene expression (Dobie *et al.* 1996; Iglesias *et al.* 1997). Moreover, the neighboring regulatory elements at the site of insertion can silence or modulate the expression of the transgene in mice (Hatada *et al.*, 1999; Al-Shawi *et al.*, 1990).

It is common that different transgenic mouse same transgene exhibit different pattern of expression. Position effect variegation is rather a common phenomenon that is responsible for sectoring of gene expression patterns in both amounts and tissue specificity. It is a common term concerning transgene silencing in a mouse model (Dobie *et al.*, 1997). It is a position-dependent silencing of the transgene in a fraction of cells that lead to forming tissues, where these cells continue to inherit this trend to the daughter cells. Therefore, the tissue shows a variable and mosaic pattern of transgene expression. This stochastic silencing of the transgene is mostly caused by different mechanisms, such as chromosome localization of the transgene, transgene copy number, DNA methylation, aggregation of multiprotein complexes at the promoters with repeated sequences, etc. However, the effect

of variegation is difficult to detect in the tissues of the immune system and brain. This is because in these tissue the daughter cells happen to migrate following cell division. (Mentioned in Dobie *et al.*, 1997).

In two mice studies, the evidence of position-effect variegation was observed. First, in two of the 3 sublines of a transgenic mouse line, of variable positioning of the transgene (with beta-1actoglobulin (BLG) transgene), showed up to ten-fold inter-individual differences in transgenic expression within the individual mice (Dobie *et al.*, 1996). In another example, in situ staining of the sections of thymus from transgenic mice carrying a human CD2 transgene showed some clustering of transgene expressing cells (Festenstein *et al.*, 1996). However, these observations point to that individual cells could able to suppress the transgenic expression and also can propagate the inactive state through cell division, which might lead to mosaic or variegated expression patterns.

Another possibility with the silencing is the position-effect model, which is similar to PEV but involves the spreading of the heterochromatin region to the gene or integration of the locus into a heterochromatin complex in the form of loop (Wakimoto 1998). However, there could be many unexplained reasons for this variation in transgenic mice (Jetton *et al.*, 1994; Hennighausen *et al.*, 1995).

### **2.5.1 Overall conclusion**

This study helps us to understand many fundamental locus-specific and organ-specific regulation of promoter activity *in vivo*. However, the exact mechanism was not known very well. It is believed that the major factor of the differential expression is because of the positioning of the transgene in the chromosome (Shaw-White *et al.*, 1993). Promoter

occlusion could be one of the reasons for the variability of expression due to the position effect.

The striking thing that was observed here is the organ-specific regulation i.e. in some lines activity was low and in some lines, activity was shown to be higher, so much so we could categorize the organs based on the level of promoter activity. As we have observed gene silencing in most of the loci, we can extrapolate that the phenomena of position effect are a ubiquitous characteristic for the given transgene in the same genetic background.

Age-dependent change of promoter activity was observed. For example, in spleen and testes, the activity was critically low in neonates but was detectable in later stages. In LacG071, we showed how promoter activity in the brain (**Fig. 2.22**) showed an individualized pattern as we observed in some animals. An animal showed high expression in one part of the brain, continued showing higher expression in other parts as well. Also, in rare cases, some animal showed a very high level of expression persistently in most of the organs.

However, the exact factor controlling the transgenic expression was not understood. Therefore, further studies with transgenic mouse having a targeted placement of the transgene at a chose locus was necessary to observe a stable expression if any (Jasin *et al.*, 1996; Wallace *et al.*, 2000). The technical caveat of this study was included, low small sample number, presence of high variability of the signal quantification. This expression pattern has relied on one type of LINE-1 promoter.

## 2.6 References

1. Aguzzi, A & Theuring, F. (1994) Improved in situ beta-galactosidase staining for histological analysis of transgenic mice. *Histochemistry*; 102: 477-481.

2. Al-Shawi R, Kinnaird J, Burke J, Bishop JO. (1990) Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Mol Cell Biol*; 10: 1192–8.
3. Baillie JK *et al.* (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*; 479(7374):534–537
4. Balow D. (1990) Gametic imprinting in mammals. *Science*; 270:1610–1613.
5. Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P. (2010) Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.*; 38: 3909–3922.
6. Bestor TH. (2000) Gene silencing as a threat to the success of gene therapy. *J Clin Investig.*; 105: 409–411.
7. Boldrini M, Butt TH, Santiago AN, Tamir H, Dwork AJ, Rosoklija GB, Arango V., Hen R, Mann JJ. (2014) Benzodiazepines and the potential trophic effect of antidepressants on dentate gyrus cells in mood disorders. *Int. J. Neuropsychopharmacology*;17, 1923–1933.
8. Bonnerot C, Rocancourt D, Briand P, Brimber G, and Nicolas J. (1987) A Beta-galactosidase hybrid protein targeted to nuclei as a marker for developmental studies. *Proc. Natl. Acad. Sci. USA*; 84: 6795-6799.
9. Brunet LJ, McMahon JA, McMahon AP, Harland RM. (1998) Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science*; 280: 1455-1457.
10. Chevalier-Mariette C, Henry I, Montfort L, Capgras S, Forlani S, Muschler J, and Nicolas J. (2003) CpG content affects gene silencing in mice: evidence from novel transgenes. *Genome Biol.*; 4(9): R53.
11. Conchie J, Findlay J, and Levvy GA. (1958). Mammalian glycosidases: Distribution in the body. *J. Biol. Chem.*; 71: 318-325.
12. Coufal NG *et al* (2009) L1 retrotransposition in human neural progenitor cells. *Nature* 460(7259):1127–1131.
13. Crews L, Adame A, Patrick C, Delaney A, Pham E, Rockenstein E, Hansen L, Masliah E. (2010) Increased BMP6 levels in the brains of Alzheimer's disease patients and APP transgenic mice are accompanied by impaired neurogenesis. *J. Neurosci.* 30, 12252–1226.
14. Davies I, Fotheringham AP, Faragher BE. (1989) Age-associated changes in the kidney of the laboratory mouse. *Age Ageing.*; 18(2):127-33.
15. Deininger P, Morales ME, White TB, Baddoo M, Hedges DJ, Servant G, Srivastav S, Smither ME, Concha M, DeHaro DL, Flemington EK, Belancio VP. (2017) A Comprehensive Approach to Expression of L1 Loci. *Nucleic Acids Research*; 45 (5): e31.
16. Dobie K, Mehtali M, McClenaghan M, Lathe R. (1997) Variegated gene expression in mice. *Trends Genet.*; 13(4):127-30.
17. Dobie KW, Lee M, Fantes JA, Graham E, Clark AJ, Springbett A, Lathe R, McClenaghan M. (1996) Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. *Proc. Natl. Acad. Sci.*; 93:6659–6664.
18. Dorer DR & Henikoff S. (1994) Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell*; 77: 993—1002.
19. Eriksson PS *et al.* (1998) Neurogenesis in the adult human hippocampus. *Nat Med*; 4(11):1313–1317
20. Festenstein R, Tolaini M, Corbella P, Mamalaki C, Parrington J, Fox M, Miliou A, Jones M, Kioussis D. (1996) Locus control region function and heterochromatin-induced position effect variegation. *Science*; 271(5252):1123-5.
21. Fiering S, Whitelaw E, Martin D I. (2000) To be or not to be active: the stochastic nature of enhancer action. *Bioassays*; 22:381–387.

22. Galea LAM, Wainwright SR, Roes MM, Duarte-Guterman P, Chow C, Hamson DK. (2013) Sex, Hormones and Neurogenesis in the Hippocampus: Hormonal Modulation of Neurogenesis and Potential Functional Implications. *J Neuroendocrinol*; 25: 1039–1061.
23. Ghanadian R, Lewis JG, Chisholm GD. (1975) Serum testosterone and dihydrotestosterone changes with age in rat. *Steroids*; 25(6):753–62.
24. Grandi FC, Rosser JM, Newkirk SJ, Yin J, Jiang X, Xing Z, et al. (2015) Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. *Genome Res.*; 25(8):1135–46.
25. Hackbarth H, Harrison DE. (1982) Changes with age in renal function and morphology in C57BL/6, CBA/HT6, and B6CBAF1 mice. *J Gerontol.*; 37(5):540–7.
26. Hatada S, Kuziel W, Smithies O, Maeda N. (1999) The influence of chromosomal location on the expression of two transgenes in mice. *J Biol Chem*; 274: 948–55
  
27. Hennighausen L, Wall RJ, Tillmann U, Li M, Furth PA. (1995) Conditional gene expression in secretory tissues and skin of transgenic mice using the MMTV-LTR and the tetracycline responsive system. *J Cell Biochem.*; 59(4):463–72.
28. Henry I, Forlani S, Vaillant S, Muschler J, Choulika A, Nicolas JF. (1999) LagoZ and LagZ, 2 genes depleted of CpG dinucleotides, derived from the LacZ gene for the study of epigenetic control. *C R Acad Sci III.*; 322:1061–1070.
29. Hsieh CL. (1994) Dependence of transcriptional repression on CpG methylation density. *Mol Cell Biol.*; 14:5487–5494.
30. Huang CZ, Li YF, Feng P. (2001) A spectrophotometric study on the interaction of neutral red with double-stranded DNA in large excess. *Talanta*; 55: 321–328.
31. Iglesias VA, Moscone EA, Papp I, Neuhuber F, Michalowski S, Phelan T, Spiker S, Matzke M, Matzke AJ. (1997) Molecular and cytogenetic analyses of stably and unstably expressed transgene loci in tobacco. *Plant Cell*; 9:1251–1264.
32. Ivics Z, Izsvák Z, Medrano G, Chapman KM, Hamra FK. (2011) Sleeping Beauty transposon mutagenesis in rat spermatogonial stem cells. *Nat Protoc.*; 6(10):1521–35.
33. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell*; 91(4): 501–510.
34. Jasin M, Moynahan ME, Richardson C. (1996) Targeted transgenesis. *Proc. Natl. Acad. Sci.*; 93:8804–8808.
35. Jetton T.L, et al. (1994) Analysis of upstream glucokinase promoter activity in transgenic mice and identification of glucokinase in rare neuroendocrine cells in the brain and gut. *J. Biol. Chem.*; 269: 3641–3654.
36. Jin K, Peel AL, Mao XO, Xie L, Cottrell BA, Henshall DC, Greenberg DA. (2004) Increased hippocampal neurogenesis in Alzheimer's disease. *PNAS*; 101: 343–347.
37. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH Jr (2009) L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev*; 23(11):1303–1312.
38. Kass SU, Landsberger N, Wolffe AP. (1997) DNA methylation directs a time-dependent repression of transcription initiation. *Curr Biol.*; 7:157–165.
39. Kearns M, Preis J, McDonald M, Morris C, Whitelaw E. (2000) Complex patterns of inheritance of an imprinted murine transgene suggest incomplete germline erasure. *Nucleic Acids Res.*; 28:3301–3309.

40. Kingsmore SF, Giros B, Suh D, Bieniarz M, Caron MG, Seldin MF. (1994) Glycine receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. *Nat Genet.*; 7:136–141.
41. Kishigami S, Komatsu Y, Takeda H, Nomura-Kitabayashi A, Yamauchi Y, Abe K, Yamamura K, Mishina Y. (2006). Optimized beta-galactosidase staining method for simultaneous detection of endogenous gene expression in early mouse embryos. *Genesis*; 44: 57-65.
42. Kooter JM, Matzke MA, Meyer P. (1999) Listening to the silent genes: Transgene silencing, gene regulation and pathogen control. *Trends Plant Sci.*; 4:340–347.
43. Kothary R, Clapoff S, Brown A *et al.* (1988) A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature*; 335(6189):435–437.
44. Kothary R, Clapoff S, Darling S *et al* (1989) Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development*; 105(4):707–714.
45. Kumaki Y, Oda M, Okano M. (2008) QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.*; 36: 170–175.
46. Kurz DJ, Decary S, Hong Y, Erusalimski JD. (2000) Senescence-associated Beta-galactosidase reflects an increase in lysosomal mass during replicative ageing of human endothelial cells. *J. Cell Sci.*; 113: 3613–3622.
47. Lavie L, Maldener E, Brouha B, Meese EU, Mayer J. (2004) The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res.*; 14(11):2253-60.
48. Lee BY, Han JA, Im JS, Morrone A, Johung K, Goodwin EC, *et al.* (2006) Senescence-associated beta-galactosidase is lysosomal beta-galactosidase. *Aging Cell*; 5(2):187–95.
49. Li LC, Dahiya R. (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics.*; 18: 1427–1431.
50. Liu Y, Mounkes LC, Liggitt HD, Brown CS, Solodin I, Heath TD, and Debs RJ. (1996). Factors influencing the efficiency of cationic liposome-mediated intravenous gene delivery. *Nature Biotechnol.*; 15: 167-172
51. Mahmoud R, Wainwright SR, Galea LA (2016) Sex hormones and adult hippocampal neurogenesis: Regulation, implications, and potential mechanisms. *Front Neuroendocrinol.*; 41:129-52.
52. Mates L, Chuah MK, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, Grzela DP, Schmitt A, Becker K, Matrai J. *et al.* (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nature Genetics*; 41(6): 753-761.
53. Matzke AJ, Matzke MA. (1998) Position effects and epigenetic silencing of plant transgenes. *Curr. Opin. Plant Biol.*; 1:142–148.
54. Mehta AK, Majumdar SS, Alam P, Gulati N, Brahmachari V. (2009) Epigenetic regulation of cytomegalovirus major immediate-early promoter activity in transgenic mice. *Gene*; 428:20–24.
55. Morley J, Perry H. (1999) Androgen deficiency in aging med. *Med. Clin.N orth Am.*;83, 1279–1289
56. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*; 435:903–910.
57. Muotri AR, Zhao C, Marchetto MC, Gage FH. (2009) Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus*; 19(10):1002-7.



58. Naas TP, DeBerardinis RJ, Moran JV, Ostertag EM, Kingsmore SF, Seldin MF, Hayashizaki Y, Martin SL, and Kazazian HH. (1998). An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J.*; 17, 590–597.
59. Newkirk SJ, Lee S, Grandi FC, Gaysinskaya V, Rosser JM, Vanden Berg N, Hogarth CA, Marchetto MCN, Muotri AR, Griswold MD, Ye P, Bortvin A, Gage FH, Boeke JD, An W (2017) Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proc. Natl. Acad. Sci. USA.*; 114 (28):E5635-E5644.
60. Ng HH, Bird AP. (1999) DNA Methylation and chromatin modification. *Curr. Opin. Genet. Dev.*; 9:158–163.
61. Novak A, Guo C, Yang W, Nagy A, Lobe CG. (2000) Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision. *Genesis*; 28 (3-4):147-55.
62. Ostertag EM, Prak ET, DeBerardinis RJ, Moran JV, Kazazian HH Jr. (2000) Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.*; 28(6):1418-23.
63. Pawluski JL, Brummelte S, Barha CK, Crozier TM, Galea LAM. (2009) Effects of steroid hormones on neurogenesis in the hippocampus of the adult female rodent during the estrous cycle, pregnancy, lactation and aging. *Front Neuroendocrinol*; 30, 343–357.
64. Pearson B, Wolf PL, and Vazquez J. (1963). A comparative study of a series of new indolyl compounds to localize Beta-galactosidase in tissues. *Lab. Invest.*; 12: 1249-1259.
65. Philippe C, Vargas-Landin DB, Doucet AJ, van Essen D, Vera-Otarola J, Kuciak M, Corbin A, Nigumann P, Cristofari G (2016) Activation of Individual L1Retrotransposon Instances Is Restricted to Cell-Type Dependent Permissive Loci. *ELife*; 5. pii: e13926. doi: 10.7554/eLife.13926.
66. Pikaart MJ, Recillas-Targa F, Felsenfeld G. (1998) Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev.*; 12:2852–2862.
67. Potter CJ, Tasic B, Russler EV, Liang L, Luo L. (2010) The Q System: A Repressible Binary System for Transgene Expression, Lineage Tracing, and Mosaic Analysis. *Cell*; 141( 3): 536-548.
68. Prak ET, Dodson AW, Farkash EA, Kazazian HH, Jr (2003) Tracking an embryonic L1 retrotransposition event. *Proc Natl Acad Sci*; 100:1832–1837.
69. Reddy PM, Reddy PR. (1990) Differential regulation of DNA methylation in rat testis and its regulation by gonadotropic hormones. *J Steroid Biochem.* ;35(2):173-8.
70. Rosser JM & An W. (2010) Repeat-induced gene silencing of L1 transgenes is correlated with differential promoter methylation. *Gene*; 456: 15– 23.
71. Ruehl-Fehlert C, Kittel B, Morawietz G, Deslex P, Keenan C, Mahrt CR, Nolte T, Robinson M, Stuart BP, Deschl U; RITA Group; NACAD Group. (2003) Revised guides for organ sampling and trimming in rats and mice--part 1. *Exp Toxicol Pathol.*; 55(2-3):91-106.
72. Sanes JR, Rubenstein JLR, and Nicolas J. (1986). Use of a recombinant retrovirus to study post implantation cell lineage in mouse embryos. *EMBOJ*; 12: 3133-3142.
73. Schumacher A, Koetsier PA, Hertz J, Doerfler W. (2000) Epigenetic and genotype-specific effects on the stability of de Novo imposed methylation patterns in transgenic mice. *Biol Chem.*; 275:37915–37921.
74. Shaw-White JR, Denko N, Albers L, Doetschman TC, Stringer JR. (1993) Expression of the lacZ gene targeted to the HPRT locus in embryonic stem cells and their derivatives. *Transgenic Res.*; 2(1):1-13.

75. Simon D. & Knowles BB. (1993) Newly acquired peri-telomeric heterochromatin in a transgenic mouse line. *Cytogenetics and Cell Genetics*; 62: 211-213.
76. Spiers H, Hannon E, Wells S, Williams B, Fernandes C, Mill J. (2016) Age-associated changes in DNA methylation across multiple tissues in an inbred mouse model. *Mech Ageing Dev.*; 154:20-3.
77. Stubbs TM, Bonder MJ, Stark AK, Krueger F, Team BIAC (2017) Multi-tissue DNA methylation age predictor in mouse. *Genome Biology*; 18(1): 68,
78. Sur D, Kustwar RK, Budania S, Mahadevan A, Hancks DC, Yadav V, Shankar SK, Mandal PK (2017) Detection of the LINE-1 retrotransposon RNA-binding protein ORF1p in different anatomical regions of the human brain. *Mob DNA*; 8:17. doi: 10.1186/s13100-017-0101-4. eCollection 2017.
79. Takahara T, Ohsumi T, Kuromitsu J, Shibata K, Sasaki N, Okazaki Y, Shibata H, Sato S, Yoshiki A, Kusakabe M, et al. (1996) Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an L1 sequence into the skipped exon. *Hum Mol Genet.*; 5:989–993.
80. Lydiard-Martin T, Bragdon M, Eckenrode KB, Wunderlich Z, DePace AH. Locus architecture affects mRNA expression levels in *Drosophila* embryos; *bioRxiv* <https://doi.org/10.1101/005173>
81. Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas CP, Zamora J, Raine K, Menzies A, Roman-Garcia P *et al.* (2014) Mobile DNA in Cancer. Extensive Transduction of Nonrepetitive DNA Mediated by L1 Retrotransposition in Cancer Genomes. *Science*; 345 (6196): 1251343.
82. Wakimoto BT. (1998) Beyond the nucleosome: Epigenetic aspects of position-effect variegation in *Drosophila*. *Cell*; 93:321–324.
83. Wallace H, Ansell R, Clark J, McWhir J. (2000) Pre-selection of integration sites imparts repeatable transgene expression. *Nucleic Acids Res.*; 28:1455–1464.
84. Walters MC, Magis W, Fiering S, Eidemiller J, Scalzo D, Groudine M, Martin DIK. (1996) Transcriptional enhancers act in cis to suppress position-effect variegation. *Genes Dev.*; 10:185–195.
85. Wang Z, Zhang Z, Liu D, Dong S. (2003) A temperature-dependent interaction of neutral red with calf thymus DNA. *Spectrochim. Acta A.*; 59: 949–956.
86. Weber-Benarous A, Decaux JF, Bennoun M, Allemand I, Briand P, and Kahn A. (1993). Retroviral infection of primary hepatocytes from normal mice and mice transgenic for SV40 large T antigen. *Exp. Cell Res.*; 205, 91-100.
87. Weis J, Fine SM, David C, Savarirayan S, and Sanes JR. (1991) Integration site-dependent expression of a transgene reveals specialized features of cells associated with neuromuscular junctions. *J. Cell Biol.*; 113: 1385-1397
88. Zhang S, McCarter JD, Okamura-Oho Y, Yaghi F, Hinek A, Withers SG, Callahan JW. (1994) Kinetic mechanism and characterization of human Beta-galactosidase precursor secreted by permanently transfected Chinese hamster ovary cells. *Biochem. J.*; 304: 281–288.
89. Zheng K, *et al.* (2010) Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc. Natl. Acad. Sci. USA*; 107:11841–11846.

## Chapter 3

### Specific locus- and orientation-dependent LINE-1 promoter activity

#### 3.1 Abstract

Randomly integrated transgenes are prone to gene silencing, while targeted delivery of transgene to a chosen locus in the mouse genome has few advantages. Therefore, the transgene was targeted to Rosa26 locus, which is well known for giving higher expression in different animal models. To characterize the orientation-dependent expression of the endogenous LINE-1 promoter, we also orientated in two different directions and created two gene-targeted mouse lines: sense and anti-sense. We observed a thousand-fold higher expression in the anti-sense orientated form, unlike the sense oriented promoter. Expression from the sense oriented form is believed to be affected by the transcriptional interference from upstream Rosa promoter that might have led to read-through transcription that interfered with transcription initiation from the transgene promoter. No age-dependent difference in expression in the anti-sense line was observed, except in the cerebellum. In short, this orientation-dependent expression also indicated how an inversely placed LINE-1 insertion would be expressed *in vivo* in mammals.

#### 3.2 Introduction

Gene silencing can be experienced by the introduced transgene in a random location of the genome of the host animal (Clark *et al.*, 1994). Most of the mechanisms are unknown for this silencing, but two of the prominent reasons can be the status of the chromatin landscape

of the site of the host genome and the copy number of the introduced transgene (Clark *et al.*, 1994; Dorer, 1997; Garrick *et al.*, 1998). The consequences of this silencing can be either in mosaic expression, which might indicate a progressive silencing, or the complete shutdown of the transgene's overall expression (Martin & Whitelaw, 1996).

In this occasion, to avoid this silencing effect, a preferred chromosomal site can be chosen in the mouse genome (Misra & Duncan, 2002; Bronson *et al.*, 1996). This approach can also give some of the added advantages. First, a favorable chromosomal site for a consistent expression may help to avoid the possibility of undesirable insertional mutagenesis. Second, the transgene can be introduced in a single copy to exclude the problems associated with large tandem repeats.

One of such chromosomal locus for ubiquitous expression can be found in mouse chromosome 6 (Zambrowicz *et al.*, 1997). This locus is known as *Gt(ROSA)26S* or *Rosa26 locus*. Since the *Rosa26* locus is mostly active in most of the cells or organs, any genetic sequence inserted into this locus must not be shut down for its expression by chromatin's unfavorable configurations. Therefore, this locus position is often used to express endogenous sequences, and often for a reporter sequence attached with an endogenous promoter (Soriano, 1999; Mao *et al.*, 2001). In addition, the promoter present on the *Rosa26* locus can also be applied to drive a widespread expression of reporter genes in transgenic mice and rats (Awatramani *et al.*, 2001). In these animal models, a ubiquitous, stable expression out of a single-copy transgene, being at this locus, can be utilized for lineage-tracing experiments in different stages of development (Soriano, 1999).

In the earlier chapter, the expression from the transgene construct was checked by inserting LacG transgene in several random chromosomal sites. The neighboring regulatory

elements at the site of integration of this transgene might have played a vital role in modulating the expression (Hatada *et al.*, 1999; Al-Shawi *et al.*, 1990). In this study, to eliminate the influence of this factor on the expression, the same transgene construct was inserted into the Rosa26 locus by targeted recombination, in order to test if the Rosa26 locus would affect transgene expression (**Fig. 3.1**). This will help us to understand whether the placement of transgene at a specific chromosomal address expressed in the same manner *in vivo* or more organs.

Notably, two ideas are important to understand at this stage as well. Firstly, it is also known that the position effects are influenced by the orientation of a transgene with respect to flanking chromatin (Feng *et al.*, 2001). Secondly, in a human genome, in addition to sense-oriented transcripts, antisense transcripts are produced by the anti-sense promoter present in the human 5'-UTR promoter, and nearly one-third of the human LINE-1s possess active anti-sense promoters (Speek, 2001). It might be very obvious that some of these antisense promoters are translated. These antisense transcripts also have regulatory roles; one of them is the base pair formation with the sense-oriented transcript in order to form a dsRNA substrate for the Dicer protein for degradation (Levin *et al.*, 2011). On the other hand, it is already known that anti-sense oriented insertion is prone to produce truncated transcripts with premature polyadenylation (Han *et al.*, 2004). It is also reported that antisense promoters can act as an alternate promoter of the neighboring sequencing, deriving the formation of chimeric mRNA, which, again, interfere with the regulation of the adjacent genes. In humans, these anti-sense promoters have locus-dependent activities (Matlik *et al.*, 2006). In contrast, LINE-1 insertion in a sense orientation is rather rare in the protein-coding gene in the human reference genome (Ewing and Kazazian, 2011). However, they

are more detrimental to a gene because RNA polymerase II strives hard to process the sense oriented LINE-1 insertions (Chen *et al.*, 2006; Han and Boeke, 2004; Han *et al.*, 2004). Therefore, we see that LINE-1 insertions in both orientations are dissimilar based on the differences of their actions and regulations. Therefore, it is vital to characterize how mouse 5'UTR LINE-1-promoter activities are expressed in sense and anti-sense being in the same locus to better understand the endogenous regulation of LINE-1 promoters.

Hence, taking the above-mentioned points into account, we systematically varied the orientation by letting the transgene integrate in sense and inverse orientations into the Rosa26 locus, and thus generated two independent mouse lines, namely sense and antisense lines. They will allow us to assess the effect of the construct's orientation for the integration site.

### **3.3 Materials and Methods**

#### **Ethics statement**

The study was carried out under the strict accordance with the guidelines provided by the National Institutes of Health in the Guide for the Care and Use of Laboratory Animals. The protocols were thoroughly approved by the Institutional Animal Care and Use Committees (IACUC).

#### **Mice**

WT and transgenic mouse lines were maintained in the C57BL/6J (B6) background. Mice were housed in well-ventilated cages. In maximum, up to five adult mice per cage was allowed. They fed on quality-controlled standard pellet chow and pure water *ad libitum* in a regular 12-hour light/dark cycle at room temperature. Rosa26 sense and Rosa26 antisense

lines were produced with gene targeting (homologous recombination) method by Cyagen Biosciences Inc., US.

### **Cryosectioning and X-gal staining of fixed tissue**

Mostly the same as Chapter 2. But the X-gal staining duration was re-optimized for Rosa26 antisense line.

### **Tissue harvesting and preparation**

Same as Chapter 2

### **Statistics**

Statistical analysis and mathematical calculations were performed using either Microsoft excel. Sample means were compared with the help of two-tailed unpaired T-test used and expressed in terms of the *p*-value. Analysis of replication consistency was tested with Coefficient of variation (CV%) was used to analyze replicate consistency. Any CV values with <100% were regarded as with a lesser variation.

### **Genotyping PCR**

Genotyping of the mice was performed with gDNA mainly from their tail biopsies. A PCR reaction using ExTaq or ExTaq HS polymerase was run for the detection of the presence or absence of expected transgene with the set of specific primers (**Table 3.1**).

Line	Primer pairs	Expected band	PCR protocol	
Mov10L1 <sup>+/-</sup>	WA549 & WA550	461 bp (Transgenic band)	Genotyping PCR	
	WA567 & WA568	398 bp (WT band)		
Rosa26 sense	Region 1 WA1550, WA1555	Mutant= 373bp WT= na	Rosa26 genotyping PCR protocol:	
	Region 2 WA1551, WA1552, WA1553	WT=617bp MT=335bp		
Rosa26 antisense	Region 1 WA1550, WA1554	Mutant= 476bp WT= n/a	C	min/s
	Region 2 WA1555, WA1552 WA1553	WT=617bp Mutant=348bp	<b>94</b>	3min
			<b>94</b>	30
			<b>62</b>	35
			<b>72</b>	35
			<b>72</b>	5min

**Table 3.1** Genotyping protocols for two transgenic mouse lines.

### Microscopy and image analysis

Same as Chapter 2

### QuPath

Same as Chapter 2

### QuPath Data Plot

Same as Chapter 2

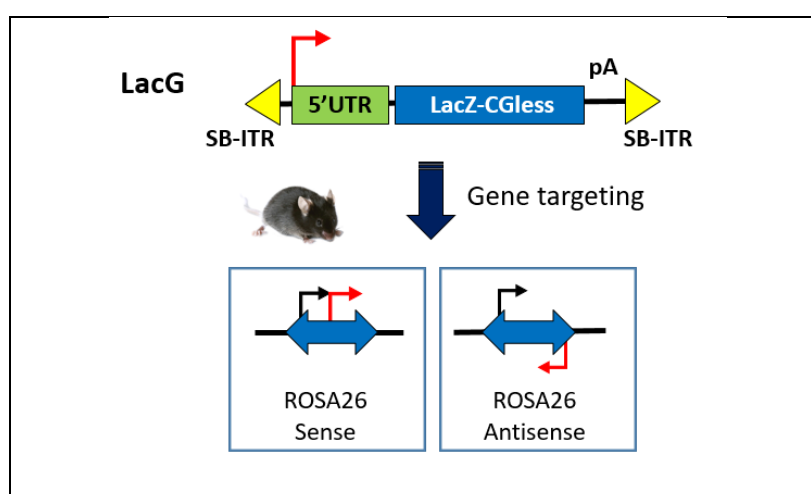
## 3.4 Results

### 3.4.1 Unlike Rosa26 sense line, Rosa26 antisense tissues needed further optimization due to high abundance of the signals

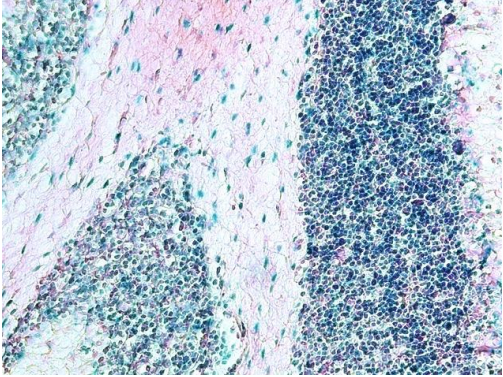
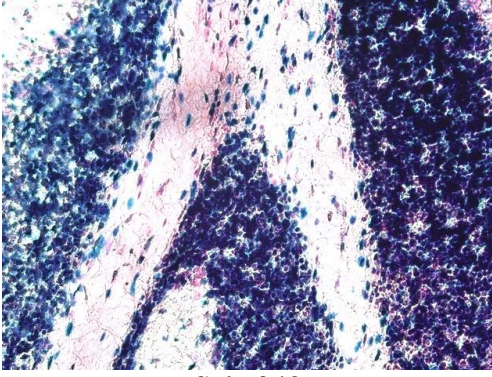
At the outset, the same X-gal staining protocol as mentioned for the genotypes in Chapter 2 was adopted. However, excessive staining was obtained for the anti-sense line. Therefore, to restrict the excessive staining, which might interfere with the quantification

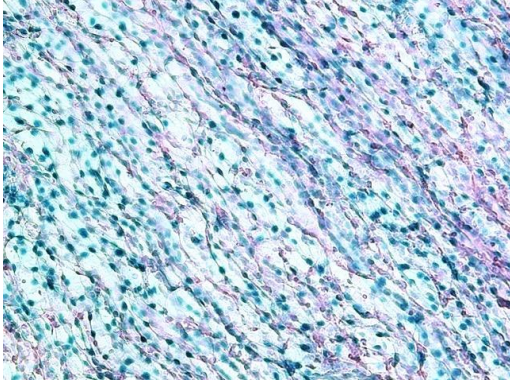
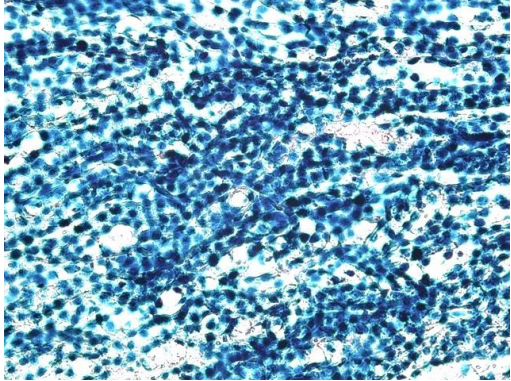
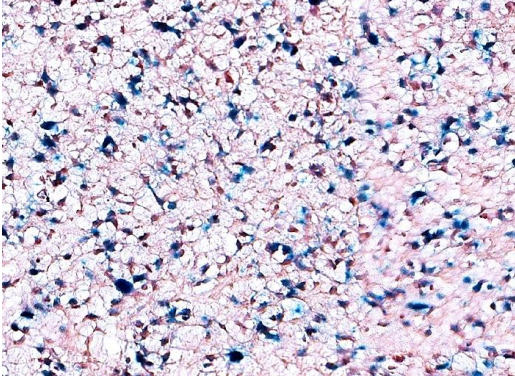
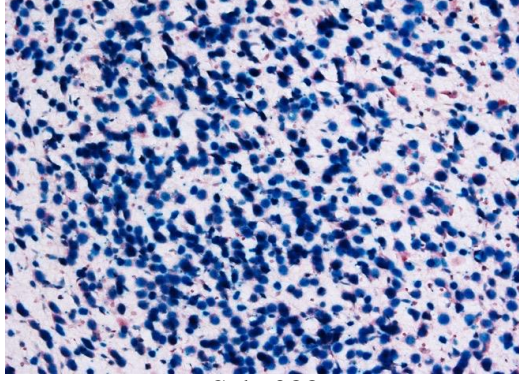


of the signals, we had to reoptimize the staining time for Rosa26 antisense animals only, depending on the age of the animals. For adult tissues, time was reduced to 45 minutes for brain sections (**Fig. 3.2.1a**) and 120 minutes for other tissues (**Fig. 3.2.1b**) using a standard concentration of X-gal (25mg/ml). For neonatal tissues, the optimal staining of all tissues was obtained with a lower concentration of the X-gal compound (6.25mg/ml) for 4 hours (**Fig. 3.2.2**). Figure 3.3 shows the typical staining patterns of these two lines in the hypothalamus of neonatal mice.

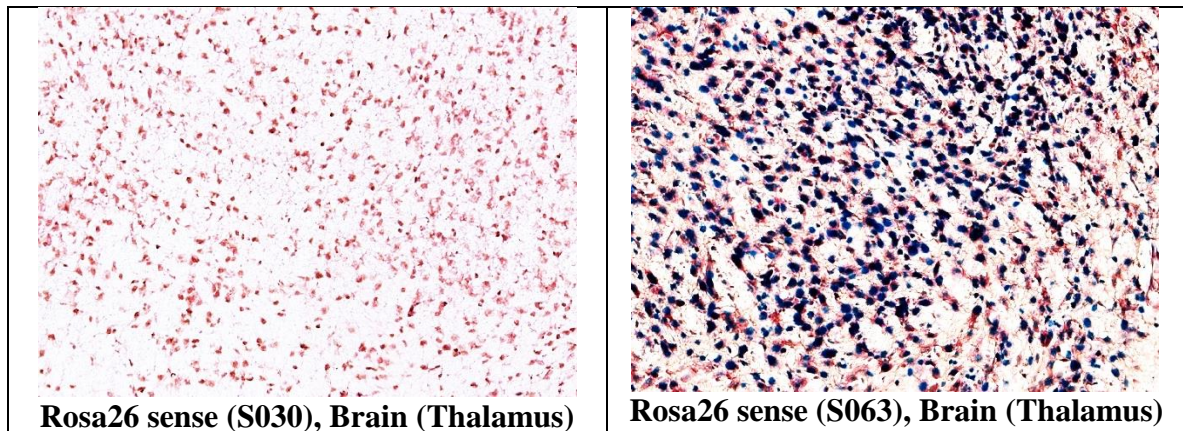


**Figure 3.1.** Two different orientations of the promoter targeted into the Rosa26 locus.

<b>1. Adult tissues of Rosa26 anti-sense line</b>	
<b>(a) Brain (Cerebellum)</b>	
45 mins incubation with X-gal 25mg/ml	3 hours incubation with X-gal 25mg/ml
	
Saha042	Saha042

<b>(b) Kidney</b>	
120 mins incubation with X-gal 25mg/ml  Saha042	3 hours incubation with X-gal 25mg/ml  Saha042
<b>2. Neonatal tissues of Rosa26 anti-sense line</b>	
<b>Brain (Thalamus)</b>	<b>Brain (Thalamus)</b>
4 hours incubation with X-gal 25mg/ml  Saha088	Overnight incubation with X-gal 25mg/ml  Saha088

**Figure 3.2:** Optimization of staining condition for Rosa26 antisense line. **1.** Adult time point in brain and kidney **2.** Neonatal time point in the brain.

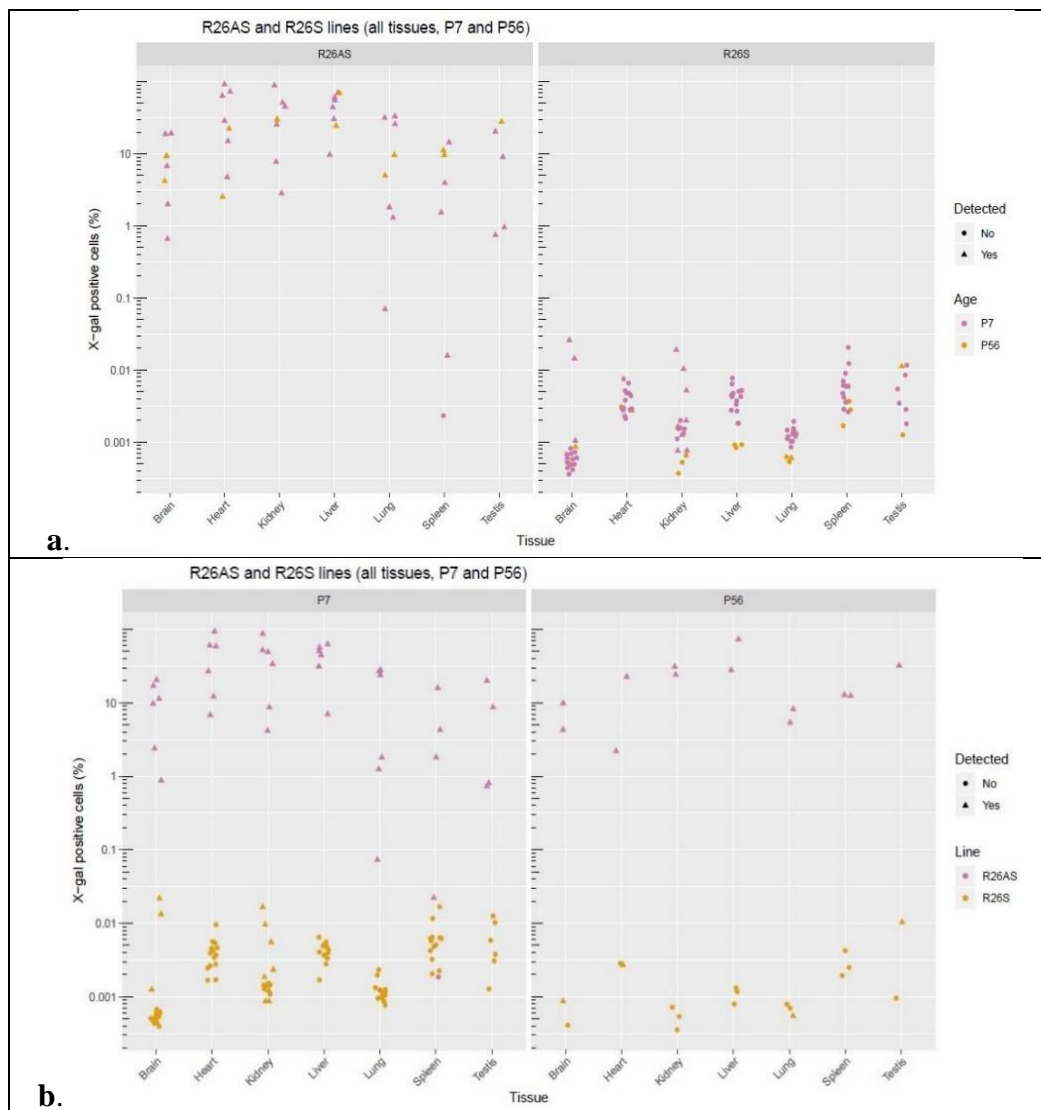


**Figure 3.3:** Desired staining patterns matched with their corresponding the two genotypes used in this study.

### **3.4.2 Promoter orientation altered the level of expression: Rosa26 antisense had 1000 fold more expression than the Rosa26 sense line**

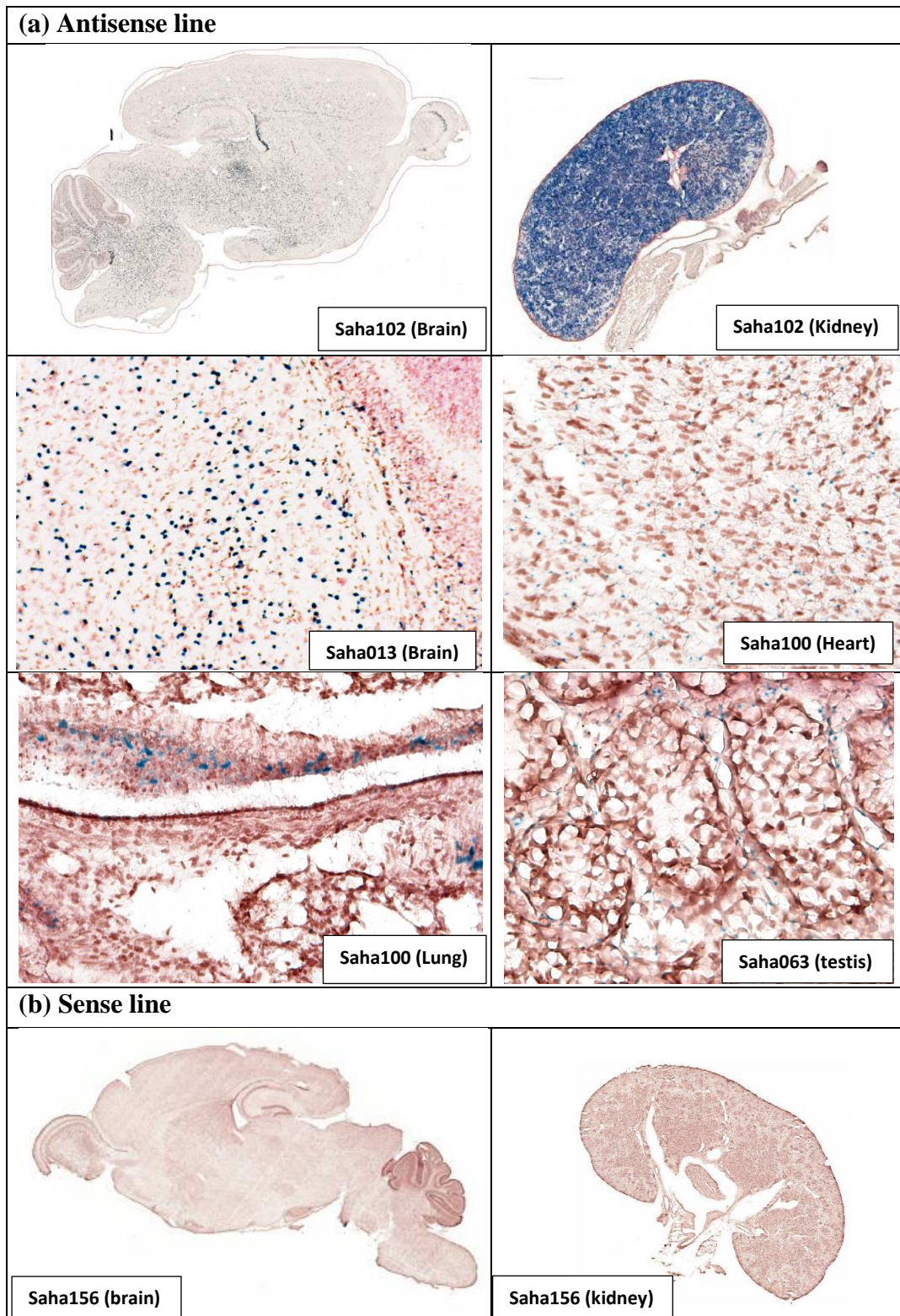
In the Rosa26 antisense line, a high X-gal expression was observed in all organs. Approximately, more than 10% of cells in all organs showed to possess positive staining. However, exceptionally, in few cases in lungs and spleen had low expression. In the heart, similar to the random lines, non-nuclear localized positive signals were observed.

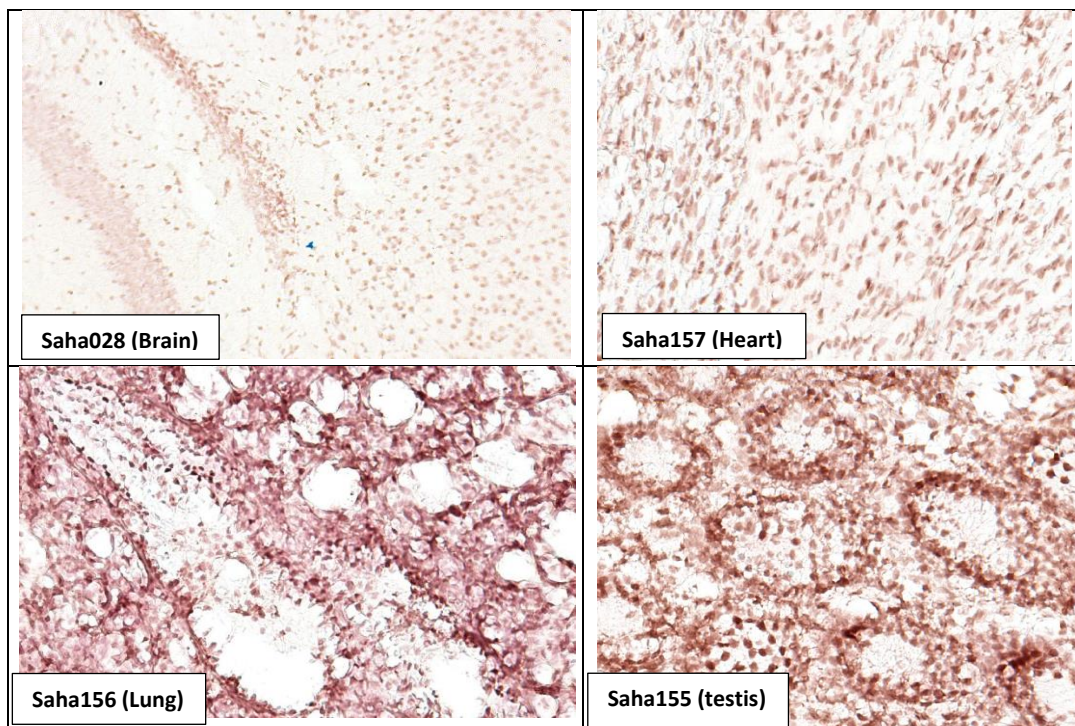
In this Rosa26 sense line, most of the expressions are undetectable. However, a few detectable positive signals have been identified in brain and kidney of neonatal time points. Similar to the neonatal situation, at the adult stage, detectable signals were identified in the brain, lungs, and testis. Although, the organ with the highest detectable positive signals was carrying a significantly low percentage of positive cells (less than 0.1%). In both of the lines, approximately, there is a 1000 fold difference in the level of expression between Rosa26AS and Rosa26S. The expression (**Fig. 3.4**) was compared between two different time points i.e. the neonates and adults of Rosa26AS and Rosa26S lines.



**Figure 3.4** The difference of expression of two gene-targeted lines at two different time points. **(a)** The orientation-dependent difference of expression, **(b)** age-dependent difference of expression

## Sample images:





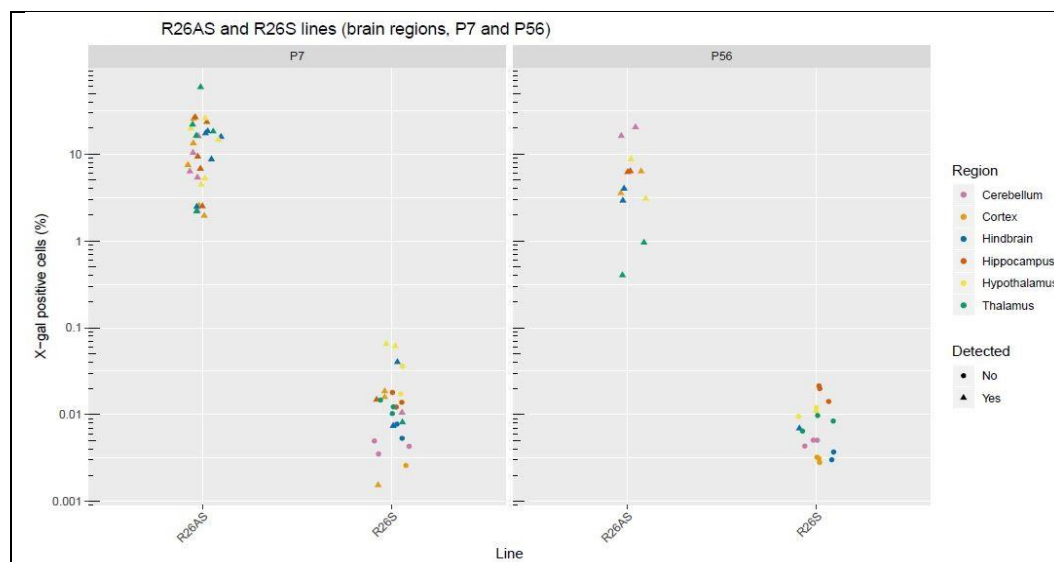
**Figure 3.5** Sample images from Rosa26 anti-sense (a) and sense lines (b). (Note: some regions are zoomed in to show the contrast/pervasiveness of the expression).

### 3.4.3 There is developmental time-dependent brain expression in brain regions of Rosa26 antisense line

In the Rosa26 anti-sense line, for both of the time points, expressions were very high (**Fig. 3.6**). In an average, ~13.35% cells are X-gal positive at neonatal time point. Among these values, the thalamus has the leading expression. In neonatal time point, unlike others, the cerebellum has statistically significant (p value= 0.037968) difference between both time points of neonatal and adult within the given set of sample number. On the other hand, in the adult time point, the number of positive cells decreased to less than 10% averagely (6.65%). At the same time, the thalamus values specifically fell to approximately 1% (0.68%).

In Rosa26 sense line, most of the signals are undetectable (**Fig. 3.6**). At the neonatal time point, the hypothalamus was a region with high positive signals, but this region

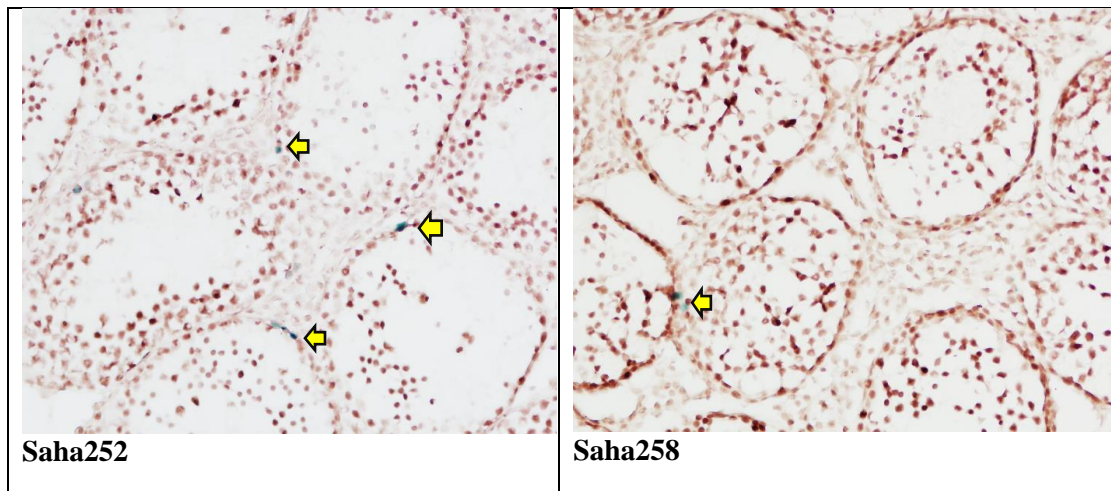
went undetectable in case of the adult time point. The hindbrain showed detectable range in both neonatal as well as in adult time points.



**Figure 3.6** The orientation-dependent expression differences of the transgene at two developmental time points.

### 3.4.4 Transgene in *Mov10L1* knock out mouse background did not show over-expression of promoter activity in testis of *Rosa26* sense line

To look for whether the hypermethylated transgene promoter can over-express promoter activity in the hypomethylated background. We observed that sense-line in *Mov10L1* KO background did not produce and signal. Only a few basal cells in the seminiferous tubules expressed promoter activity.



**Figure 3.7** Promoter activity in testis of Rosa26 sense line in a Mov10L1 KO background.

### 3.5 Discussion

To rule out the influence of integration site and variable copy number on the expression pattern of the transgene, we knocked it in single copy into a specific locus, Rosa26. At the outset, we had two choices of favorable chromatin loci for ubiquitous and stable expression, namely HPRT locus and Rosa26 locus. As the HPRT locus is located on the X chromosome, the expression is subject to random X-inactivation, or in other words, the expression would be guaranteed only in case of homozygous females. Moreover, the evidence is there that even in that case too, expression in certain tissues, like kidney and liver, are low and sometimes undetectable (Bronson *et al.*, 1996; Hatada *et al.*, 1999). Therefore we had to opt for the alternative choice i.e. Rosa26 locus. This has been widely used for favorable gene targeting site in mouse (Friedrich & Soriano, 1991; Srinivas *et al.*, 2001; Nyabi *et al.*, 2009). In addition to the use in mouse, this chromosomal site has also been used for traditional homologous recombination in humans and rats (Irion *et al.*, 2007; Kobayashi *et al.*, 2012). This also suggested that this site has limited inter-species



variability for stable expression. This ubiquitous expression also suggested that the transcriptional activity is less likely influenced by the chromatin configurations, which sometimes offer transcriptional repression through several regulatory elements in the flanking chromatin landscape. Whatsoever, we targeted our LacG transgene into this locus for a reliable expression pattern.

Similar to as described in Chapter 2, we stained the fixed tissues with X-gal from the mouse at two time points: neonatal and adult. At first, we attempted to keep the X-gal staining time same for the random lines and these two targeted lines. Upon staining the tissue for regular length of time, we observed a dark, heavy, diffusible stains only in the anti-sense line. It was unrealized until then that the blue compound as a product of the enzymatic reaction is, however, diffusible in fixed tissue and could eventually blur the distinctions between the nuclear and cytoplasmic signals (Sanes *et al.*, 1986; Gray *et al.*, 1988; Weis *et al.*, 1991). Overall, this the problem became apparent when we stained the samples from the Rosa26 antisense line with the optimized X-gal staining condition used in Chapter 2. The heavy blue stains darkened the sections, and almost completely masked the neutral red co-stains. Co-localization of both of the signals, where applicable, was needed to determine the total fraction of transgenic positive cells within the whole population of cells, which was determined by the total number of neutral red-stained cells. Therefore, strategically, we had to shorten the duration of X-gal staining for the Rosa26AS genotype so that we have distinct nuclear-localized signals to represent the total number of transgene positive cells.

In the revised protocol, we still observed a pervasive X-gal staining in most of the organs of heterozygous antisense animals, whereas we found almost no staining in the

cryosections of sense line. An estimation of the difference was 1000 fold between these two lines. Notably, the highest value of signal quantification if the Rosa26 sense line is 0.1%, whereas the estimated average signal quantification of Rosa26 sense line was around 10%. This is the novel approach where endogenous LINE-1 promoter was ever targeted into Rosa26 locus. Therefore, no data related with this transgene construct being targeted in this specific locus or other related loci was available to understand the mechanisms of silencing of the Rosa26 sense oriented promoter and the same of heavy staining of the anti-sense oriented promoter.

However, a similar study was performed using the same reporter in HPRT locus by Stringer and colleagues (Shaw-White *et al.*, 1993). There they observed a similar contrast between two orientations of LacZ transgene targeted into the HPRT locus. In that study, they targeted a LacZ gene, under control of an SV40 promoter, to the HPRT locus in ES cells. The expectation was that, since HPRT is a 'housekeeping gene' and is expressed constitutively in all cell types, LacZ would be ubiquitously expressed. The targeted ES cells, all of which expressed LacZ in culture, were injected subcutaneously into syngeneic strain 129 mice and allowed to grow into tumors containing multiple differentiated tissue types, which were then stained for beta-gal galactosidase activity. Targeted cell lines with LacZ in inverse orientation to the direction of HPRT gene transcription expressed high levels of beta-galactosidase in epithelial cells. However, targeted cell lines containing a transgene oriented in the same direction as the HPRT gene transcription did not express high levels of LacZ in any differentiated cell type. Analysis of transcripts suggested that this orientation effect may have been the result of transcriptional interference perpetrated by the HPRT gene promoter. Cell lines in which LacZ was oriented .in the same direction

as HPRT contained RNA transcripts that appeared to originate from the HPRT promoter and proceed through the transgene's promoter and the LacZ coding region. These RNAs would not be expected to be capable of producing an HPRT-beta-galactosidase fusion protein because of the presence of a stop codon in between the HPRT open reading frame (ORF) and the beta-galactosidase ORF. Some direct-orientation cell lines contained no detectable 3.3 kb transcript. It might be possible that the low abundance of the 3.3 kb transcript in these cell lines was due to read-through transcription interfering with transcription initiation from the transgene promoter (Shaw-White *et al.*, 1993). Notably, inconsistent with their result, in our result too, the quantified expression values from the sense line is lowest in compared to the randomly integrated lines (with intermediate values) and antisense line (with highest) values of expression.

In In this present study, we also predict that similar mechanisms of transcriptional interference might have taken place, completely ruling out the trivial possibility that the plasmid was constitutively defective for expression of beta-galactosidase because 3T3 cells transfected with this DNA were positive in the X-Gal assay. Overall, this might be the reason for how around 10% of all organs showed positive staining in the antisense line. Unstained cells in the organs of Rosa26 antisense, along with the lowly expressed organs, like lungs and spleen, might have different cell-specific or organ-specific regulations which restricted the transgene expression. These types of regulation can be due to tissue-specific and developmental stage-specific transcription factors, like auxiliary proteins and DNA-binding sequence-specific transcription factors or host epigenetic factors. Any of this kind of mechanism either have gone loosen or restricted at the cellular or organ levels mouse of either of the developmental time points. Matlik *et al.*, showed that there are a locus-

dependent and tissue-specific expression pattern of antisense promoter's activity in human *in vivo*. There, they have demonstrated that LINE-1 ASP antisense promoter (ASP) is active in a wide variety of normal human tissues, but LINE-1 ASPs at defined loci are not active in all tissues (Matlik *et al.*, 2006). There they explained that varies based on minimal sequence divergence and differences in their epigenetic state. In this case, we explain our case with the latter phenomenon as the probability of the former case is null.

It is also possible that the lack of expression in these cells could have been due to a silencer-effect exerted by the 5'-end of the Rosa26 segment juxtaposed to the transgene gene. However, the activity of such a hypothetical cis-acting a silencer would be necessarily conditional in two ways. Firstly, it would be inactive in cells in which LacZ expression was observed. Secondly, it would be inactive when located downstream of the transgene. In heart, we have not found any nuclear-localized signals like that we saw in chapter 2. However, we believe that these are real signals as we did not obtain any signals in transgene negative animals, but we cannot give any explanation to the cause of these signals of this shape.

Since we see the brain is an important site with adequate expression in random lines, we took a closer look at the brain in Rosa26 antisense and sense lines. In Rosa26AS, overall we observed a consistent high expression in both of the developmental time periods. However, compared to the neonatal time point, we observed a region-specific fall of expression, particularly in the thalamus at adult time point. This observation is just opposite to what we observed in the highly expressed line, LacG071, of random lines. However, it is difficult to conclude on this due to the small sample number. Here, we also observed an

overall fall in the total number of X-gal positive signals from neonatal time point to adulthood. This might be due to age-dependent phenomena combined with the intrinsic characteristic of this locus for this promoter. However, among all regions, the only age-dependent difference in expression in the cerebellum was showed significant statistically in the given set of data (however, this significance might not be staying upon increasing the sample numbers). This might tell us that some other factors might be involved in case of regulation of transgene in inverse orientation, which might determine which cells have high levels of beta-galactosidase, and which cells do not. As we know that different regions of the brain nurture more or less particular types of brain cells, and with the progression of age, these type of cells either differentiate into other types or degenerate. Therefore, this age-dependent the difference in expression in cerebellum might hint us regarding a cell type-specific expression, and also expression variegation, where expression of beta-galactosidase varied within a population of cells of one type. In the next chapter, we will show how we attempted to differentiate these cell types which particularly held an expression of the antisense-oriented transgene. On the contrary, as like most of the organs in the Rosa26S line, brain regions in the same line are mostly silent. Any occasion of expression in certain regions can be called as a rare cell-specific incident and could be explained appropriately by analyses at the single-cell level.

Human LINE-1 promoters are bidirectional, containing a sense promoter responsible for transcription within the LINE-1 element and an antisense promoter (LINE-1 -ASP) that can drive transcription of adjacent regions giving rise to transcripts composed partly of LINE-1 and partly of genomic sequence (LINE-1 chimeric transcripts (LCTs) (Speek, 2001; Cruickshanks & Tufarelli, 2009). Recent evidence suggests the existence of a causal

link between aberrant activation of individual LINE-1-ASP promoters and cancer development and progression (Weber *et al.*, 2010; Wolff *et al.*, 2010).

Our result highlighted that anti-sense promoter activity is thousand fold higher in compared to sense orientation being positioned in a permissive locus, like Rosa26. In type of cases, more active promoters (in this case Rosa26 promoter), in the upstream, may act as an alternate promoter to form a read-through transcript and ultimately leading to suppression of the sense oriented LINE-1 promoter to form transcript from ORF1 & 2 sequences. On the other hand, the anti-sense promoter can be unaffected. Overall, this reflected how the sense and antisense promoters, being inserted into a permissive locus may behave.

A probability of a certain level of methylation-dependent gene repression is always there, regardless of the chromosome position. In case of any hypomethylation-induced activation of LINE-1-ASP promoters can further drive the transcription of cancer-specific LINE-1 chimeric transcripts (LCTs) transcribed in the same (sense) or opposite (antisense) orientation with respect to the neighboring genes (Cruickshanks & Tufarelli, 2009). In another instance, it is evidenced by weber et al. that demethylation of a LINE-1 antisense promoter in the cMet locus impaired Met signaling through induction of illegitimate transcription (Weber *et al.*, 2010). Although the methylation of the promoters was not checked, yet it is believed that on being hypomethylated this antisense LINE-1-prompter can bring a synergistic effect in terms of intense transgenic expression. Knocking out the piRNA-DNA methylation pathway leads to DNA methylation at LINE-1 Promoters and thus it led to 70-fold increase in retrotransposition in postnatal germ-cell development in mouse with a 5'UTR-ORFeus transgene (Yang & Wang, 2016; Newkirk et al., 2017). Therefore, we bred a piRNA KO (Mov10l1<sup>-/-</sup>) mouse with LacG oriented in sense line.

Although we could observe a spermatogenic failure in the germ-lines. However, we could not see any increased expression of promoter activity (**Fig. 3.7**). This might indicate that prevalent RNA interference, but not the demethylation is responsible for the lower number of positive cells.

Overall, we found that there is an orientation-dependent expression of our transgene construct being targeted into the Rosa26 line. This finding is consistent to an earlier observation, however, in cell lines, where CMV the driven expression of the reverse tetracycline transactivator (rtTA) at the *Gt(ROSA)26Sor* locus was inferred to be more robustly expressed in the antisense orientation (Strathdee *et al.*, 2006). Besides showing an interesting expression of our transgene in Rosa26 antisense orientation, the non-expression from the sense orientation at the same time, also, reminded one of the limitations of gene targeting in Rosa26 locus i.e. transcriptional interference from upstream promoter sequences (which can be limited by the use of an insulator element). Other limitations of the general application of this method with respect to exogenous promoters have essentially limited by the transcriptional complexity of the *Rosa26* locus (Zambrowicz *et al.*, 1997) and of course a lack of systematic studies.

### **3.6 Conclusion**

To rule out the effect of the flanking chromosome site on the transgene expression, we were able to gene target our LacG transgene in two orientations into a specific locus, Rosa26, which is well known for a ubiquitous expression in different transgenic animals. In this case, we were able to show that the same transgene show contrasting orientation-dependent expression pattern in two lines. This again proves that the surrounding

chromosomal landscape was a prominent determining factor why promoter activity is repressed in randomly integrated lines and also provides the necessary evidence that endogenous LINE-1 promoter in anti-sense orientation might have a profound expression in the mammalian genome.

### 3.7 References

1. Al-Shawi R, Kinnaird J, Burke J, Bishop JO. (1990) Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Mol Cell Biol*; 10: 1192–8.
2. Awatramani R, Soriano P, Mai JJ. (2001) Dymecki S. An Flp indicator mouse expressing alkaline phosphatase from the ROSA26 locus. *Nat Genet.*; 29:257–9.
3. Bronson SK, Plaehn EG, Kluckman KD, Hageman JR, Maeda N, and Smithies O. (1996) Single-copy transgenic mice with chosen-site integration. *Proc. Natl. Acad. Sci. USA*; 93(17): 9067–9072.
4. Chen J, Rattner A, Nathans J. (2006) Effects of LINE-1 retrotransposon insertion on transcript processing, localization and accumulation: lessons from the retinal degeneration 7 mouse and implications for the genomic ecology of LINE-1 elements. *Hum Mol Genet*; 15(13):2146–2156.
5. Clark AJ, Bissinger P, Bullock DW, Damak S, Wallace R, *et al.* (1994) Chromosomal position effects and the modulation of transgene expression. *Reprod Fertil Dev.*; 6:589–598.
6. Cruickshanks HA, Tufarelli C. (2009) Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter, *Genomics*; 94: 397-406.
7. Dorer DR. (1997) Do transgene arrays form heterochromatin in vertebrates? *Transgenic Res.*; 6:3–10.
8. Ewing AD, Kazazian HH. (2011) Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. *Genome Res*; 21(6):985–990.
9. Feng YQ, Lorincz MC, Fiering S, Grealley JM, Bouhassira EE. (2001) Position effects are influenced by the orientation of a transgene with respect to flanking chromatin. *Mol Cell Biol.*; 21(1):298-309.
10. Friedrich G, Soriano P. (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.*; 5: 1513–1523.
11. Garrick D, Fiering S, Martin DIK, Whitelaw E. (1998) Repeat-induced gene silencing in mammals. *Nature Genetics*; 18:56–59.
12. Gray GE, Glover JC, Majors J, and Sanes JR. (1988). Radial arrangement of clonally related cells in the chicken optic tectum: Lineage analysis with a recombinant retrovirus. *Proc. Natl. Acad. Sci. USA*; 85: 7356-7360.
13. Han JS, Boeke JD. (2004) A highly active synthetic mammalian retrotransposon. *Nature*; 429(6989):314–318.



14. Han JS, Szak ST, Boeke JD. (2004) Transcriptional disruption by the LINE-1 retrotransposon and implications for mammalian transcriptomes. *Nature*; 429 (6989):268–274.
15. Hatada S, Kuziel W, Smithies O, Maeda N. (1999) The influence of chromosomal location on the expression of two transgenes in mice. *J Biol Chem.*; 274(2):948-55.
16. Irion S, Luche H, Gadue P, Fehling HJ, Kennedy M, et al. (2007) Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol*; 25: 1477–1482.
17. Kobayashi T, Kato-Itoh M, Yamaguchi T, Tamura C, Sanbo M, et al. (2012) Identification of rat Rosa26 locus enables generation of knock-in rat lines ubiquitously expressing tdTomato. *Stem cells and development*; 21: 2981–2986.
18. Levin HL, Moran JV. (2011) Dynamic interactions between transposable elements and their hosts. *Nature Rev Genet*; 12:615–627
19. Mao X, Fujiwara Y, Chapdelaine A, Yang H, Orkin SH. (2001) Activation of EGFP expression by Cre-mediated excision in a new ROSA26 reporter mouse strain. *Blood*; 97:324–326.
20. Martin DIK & Whitelaw E. (1996) The vagaries of variegating transgenes. *BioEssays*; 18: 919–923.
21. Matlik K, Redik K, Speck M. (2006) L1 antisense promoter drives tissue-specific transcription of human genes. *J Biomed Biotechnol* 2006:1–16.
22. Misra RP, Duncan SA. (2002) Gene targeting in the mouse: Advances in introduction of transgenes into the genome by homologous recombination. *Endocrine*; 19:229–238.
23. Newkirk SJ, Lee S, Grandi FC, Gaysinskaya V, Rosser JM, Vanden Berg N, Hogarth CA, Marchetto MCN, Muotri AR, Griswold MD, Ye P, Bortvin A, Gage FH, Boeke JD, An W (2017) Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proc. Natl. Acad. Sci. USA.*; 114 (28):E5635-E5644.
24. Nyabi O, Naessens M, Haigh K, Gembarska A, Goossens S, et al. (2009) Efficient mouse transgenesis using Gateway-compatible ROSA26 locus targeting vectors and F1 hybrid ES cells. *Nucleic Acids Res* 37: e55.
25. Sanes JR, Rubenstein JLR, and Nicolas J. (1986). Use of a recombinant retrovirus to study post implantation cell lineage in mouse embryos. *EMBOJ*; 12: 3133-3142.
26. Shaw-White JR, Denko N, Albers L, Doetschman TC, Stringer JR. (1993) Expression of the lacZ gene targeted to the HPRT locus in embryonic stem cells and their derivatives. *Transgenic Res.*; 2(1):1-13.
27. Soriano P. (1999) Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.*; 21:70–71.
28. Speck M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol.*; 21(6):1973-85.
29. Srinivas S, Watanabe T, Lin CS, Williams CM, Tanabe Y, et al. (2001) Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC developmental biology* 1: 4.
30. Strathdee D, Ibbotson H, Grant SG. (2006) Expression of transgenes targeted to the Gt(ROSA)26Sor locus is orientation dependent. *PLoS One.*; 1:e4.

31. Weber B, Kimhi S, Howard G, Eden A, Lyko F. (2010) Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. *Oncogene*; 29(43):5775–5784.
32. Weis, J, Fine SM, David C, Savarirayan S, and Sanes JR. (1991) Integration site-dependent expression of a transgene reveals specialized features of cells associated with neuromuscular junctions. *J. Cell Biol.*; 113: 1385-1397
33. Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, Yang AS, Jones PA, Liang G. (2010) Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. *PLoS Genet.*; 6(4):e1000917.
34. Yang F, Wang PJ. (2016) Multiple LINEs of retrotransposon silencing mechanisms in the mammalian germline. *Semin. Cell Dev. Biol.*; 59:118–125.
35. Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, *et al.* (1997) Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc. Natl. Acad. Sci. USA.*; 94:3789–3794.

## Chapter 4

### Cell-specific LINE-1 Promoter Activity

#### 4.1 Abstract

The mammalian brain is composed of different types of cells, which can be classified based on their functions. Previous chapters showed brain-region specific promoter activities. This possibly indicated a physiological role of the LINE-1 activity in the mammalian brain. It is assumed that identification of the cell types, carrying the promoter activity would be helpful to reveal the role of LINE-1 on neurophysiology. Therefore, this chapter aimed to identify the major brain cells that hold the promoter activity in terms of transgene expression. In this study, six different brain cell-markers were used to detect several neuronal, macroglial, microglial, and stem cell types. A fluorogenic substrate was used to replace the X-gal compound for detecting the promoter activities. At the same time, an automated, quantitative signal detection approach was also implemented for the detection of the percentage of brain cell types holding the promoter activity in a specific brain region. Although all the brain markers were successfully detected, yet a contradictory background problem due to the use of fluorogenic substrate halted the progress. An alternative approach can be realized to complete the entire plot.

#### 4.2 Introduction

It has been a while that we are aware that retrotransposition activity presents a formidable threat to the host genome. They are involved in causing several heritable and inheritable diseases in mammals, including humans (reviewed in Saha & An, 2019).

A systematic spatial and temporal control of gene expression is an essential part of biological processes, like proliferation, apoptosis, development, differentiation and aging. In a different type of cells, these regulations are specifically maintained by a class of gene regulatory elements, known as enhancers. These enhancers are a short sequence of approximately 1 kb long located either upstream, downstream or inter-introns, can harbor specific transcription factors to produce a cell-specific expression pattern (Levine, 2010; De Laat and Duboule, 2013). Despite our advanced knowledge in the retrotransposon field, our understanding is still extremely limited on how LINE-1 activity is determined by these type of cell-specific regulations – and at least, which cell types promote or suppress retrotransposition activity.

So far, we established that individual LINE-1 promoters exhibit both loci- and tissue-specific activation, which is also orientation dependent. This implied that LINE-1 promoter activity originates from some permissive loci and also suggesting that a new layer of cell-type-specific regulation that controls endogenous retrotransposons.

Over the past years, LINE-1–EGFP reporter system had been successfully utilized in animal models to elucidate the effect of LINE-1 mobilization in age-dependent neurogenesis, suggesting the brain as a hotspot for LINE-1 mosaicism (Coufal *et al.*, 2005; Muotri *et al.*, 2005; Garcia-Perez *et al.*, 2010). Our studies in the past two chapters showed a substantial expression at the brain level, and interestingly in specific brain regions. These prompted us to translate our approach to determine the brain cells specific to hold the promoter activity. Therefore, in the present study, we tried to determine which brain cell types mostly held the LINE-1 promoter activity is mostly expressed random line and other two different gene-targeted lines.

## **4.3 Materials and Methods**

### **Ethics statement**

The study was carried out under the strict accordance with the guidelines provided by National Institutes of Health in the Guide for the Care and Use of Laboratory Animals. The protocols were thoroughly approved by the Institutional Animal Care and Use Committees (IACUC).

### **Mice**

Same as Chapter 2

### **Tissue harvesting and preparation**

Same as Chapter 2

### **Cryosectioning**

Same as Chapter 2

Note: Tissues were sectioned in a thickness of 14 $\mu$ m for immunostaining.

### **Immuno-staining followed by containing Res-gal staining with secondary antibody**

Brain sections were obtained in 14 $\mu$ m thickness. Immersed in 1XPBS for 10 minutes. Rinsed with water briefly. Incubated at RT with 1xPBS with only 0.3% TritonX100 for 5 minutes 3 times. Blocked for 2 hours at RT with a blocking buffer, containing 0.3% Triton X100 in 1XBPS and appropriate serum. Except the case of using anti-mouse primary antibody (used 1% BSA), blocking buffer in all other cases was containing 5% donkey serum. Overnight incubated with a blocking buffer with a required ratio of primary antibodies. Incubated at RT with 1xPBS with only 0.3% Triton X100 for 5 minutes for 3 times. The secondary antibody in appropriate conc., containing 1mM Res-gal for samples

of Rosa26 antisense line and 2mM for the samples of random lines, was used to incubate at 37C for 1 hour. Incubated at RT with 1xPBS with only 0.3% TritonX100 for 5 minutes for 3 times. Incubated for 5minutes with DAPI. Incubated at RT with 1xPBS with only 0.3% TritonX100 for 5 minutes 3 times. Rinse with water. Mounted with 1%PVA before putting on coverslips. Note for the procedure to prepare 2<sup>nd</sup> dilution: Prepared 1M Res-gal stock in DMSO. The appropriate volume of Res-gal was added to a solution, containing 10M MgCl<sub>2</sub> and 0.3% of TritonX100 in 1XPBS. Vortexed vigorously for 2 minutes. Centrifuge for 1 minute at 14000 rpm. Pipette out the supernatant and add the appropriate volume of a secondary antibody. Did all steps in the dark as Res-gal is light-sensitive.

<b>Chemical/ Reagents</b>	<b>Manufacturer</b>	<b>Specification</b>
Res-gal (Resorufin Beta-D-galactopyranoside) MW 375.33	Marker Gene	Ref.: M0203 Lot: 286JJN009
<b>Primary Anti-bodies</b>		
Mouse anti-GFAP (1:1000)	Sigma	Ref: G3893 Lot: 056M4864V
Rabbit Anti-NG2 (1:500)	Millipore	Ref: Ab5320 Lot: 251778
Goat Anti Sox2 (1:100)	Santa Cruz	Ref: SC-17320
Rabbit Olig-2 (1:1000)	Millipore	Ref: AB9610 Lot: 2728398
Rabbit Anti-NeuN (1:1000)	Abcam	Ref: AB177487
Rabbit Anti-IBA-1 (1:150)	Wako	Ref: 019-19741
<b>Secondary antibodies</b>		
Donkey Anti-goat Alexa fluor 488 (1:1000)	Jackson Lab	Ref: A-705- 645 147
Donkey Anti-rabbit Alexa fluor- 488 (1:500)	Invitrogen A21206	Lot: 2045215
Donkey Anti-mouse Alexa fluor- 488 (1:1000)	Invitrogen A21202	Lot: 1305303

Donkey Anti-rabbit Alexa Fluor 594 (1:500)	Invitrogen A21207	Lot: 567297
4',6-diamidino-2-phenylindole (DAPI) (1:1000)	Thermo Scientific	Ref.: 62248
<b>Other reagents</b>		
Bovine serum albumin (heat-shock treated)	Fisher Bioreagents	CAS 9048-46-8
Donkey serum	Lampire Biological Laboratories	Cat.# 7332100 Lot: 13A29004
Dimethyl Sulfoxide (DMSO)	VWR Life Sciences	CAS 67-68-5 Lot: 19C2656019
Molecular Biology grade water	Hyclone	Cat SH30538-02 Lot: AAC200214

### Confocal microscopy

The confocal microscopy of the immunostained microscopic slides was taken with the help of Olympus FV1200 Scanning Confocal Microscope (x20 dry) based at the Functional genomics core facility at South Dakota State University. The images were analyzed by a software Fluoview Fv1000. Samples were excited at 488 nm (Alexa Fluor 488), 405 nm (DAPI), 568 nm (Res-gal) and the emission light was collected at 520 nm, 461 nm, and 559 nm for each of these channels, respectively. Z-stack images of approximately 4  $\mu$ m thickness were taken for each sample at 3  $\mu$ m step sizes. Each frame consisted of a 1024  $\times$  1024 pixel image, captured at a rate of 20  $\mu$ s/pixel.

The confocal images were further analyzed with the help of image J software (version 1.52p). In total, nine stacks from each of the three channels (with 3 stacks per channel) were stacked individually after adjusting for the parameters, like brightness and Z projection (type: "Max intensity"). Later, the three stacked images for each channel were merged and was saved as .tiff file. This image was fed in QuPath for the quantification of cells with signals.

**QuPath: quantification of signals**

Two image types in the program QuPath were used for quantification: brightfield (H-DAB) for X-gal stained tissues and fluorescence for immunofluorescence tissues. For the X-gal stained tissues, a different script was created for each tissue based on its specific characteristics. Each script was modified using a certain number of channels. Each channel denoted a specific command to either denote a cell as either positive, negative, or border based on the intensity of the hematoxylin or DAB. The mean, sum, and max intensity of these two parameters were optimized over a range of several tissues to accurately detect an X-gal positive cell from a negative. A percentage was then derived by taking the number of positive cells divided by the number of negative cells plus positive cells and then multiplied by 100. The cells detected as border were determined to be falsely detected cells and were not included in this equation. Immunofluorescence followed this same protocol, but channels were focused on FITC, DAPI, and Res-gal instead of hematoxylin and DAB.

**4.4 Results**

In this experiment, we attempted to co-stain five brain cell markers (NeuN, Olig2, NG2, GFAP & Sox-2), along with Res-gal and DAPI on the same section on the same tissue section. These tissue sections were from animals of random lines as well as Rosa antisense lines. N804 is an animal belonging to the highest expressing LacG071 line, and Saha063 and Saha100 are the animals that belonged to Rosa26 antisense lines.

Figures 4.1 to 4.5 showed how well the neural markers could detect the actual cell type in the presence of Res-gal and DAPI. Figure 4.6 shows the detection of microglial marker (IBA-1) with the help of immunostaining. This case is, however, free from Res-gal.

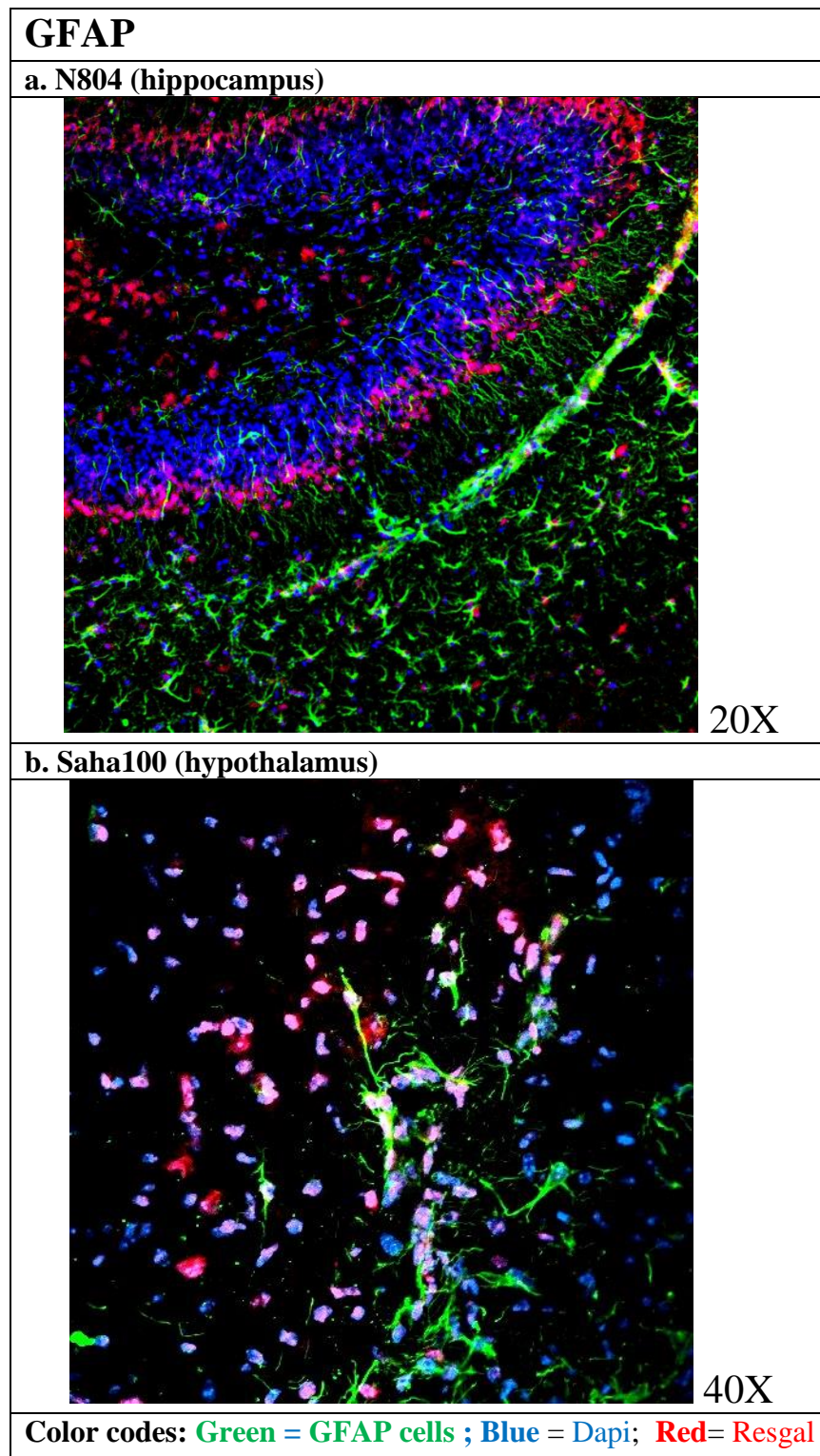


Next, we wanted to detect these three signals with the help of QuPath. And we could successfully code a script which perfectly detected different signals, belonged to different channels (RGB) (**Fig. 4.7a**). Figure 4.7b shows the concept for the quantification of colocalizing signals. Here only those cells colocalizing three colors are expressed in percentage value.

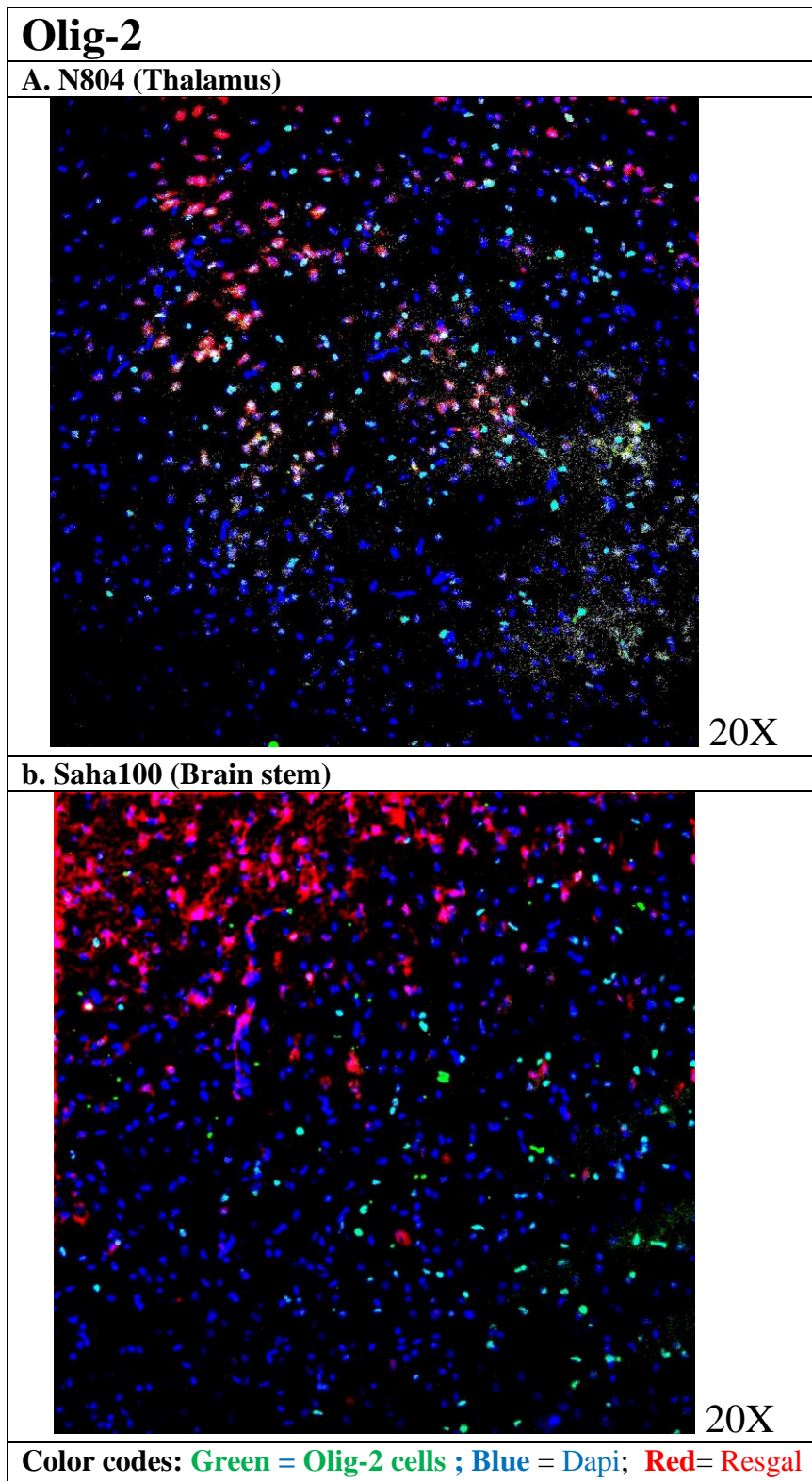
Figure 4.8 shows all types of colocalizing detections detection was necessarily containing DAPI signals in it. Here, different color codes are used to mark different types of detections.

Figure 4.9 (a) and figure 1.9 (b) shows the total percentage of cells showing colocalization of three signals at once for NeuN and Sox-2 markers, respectively.

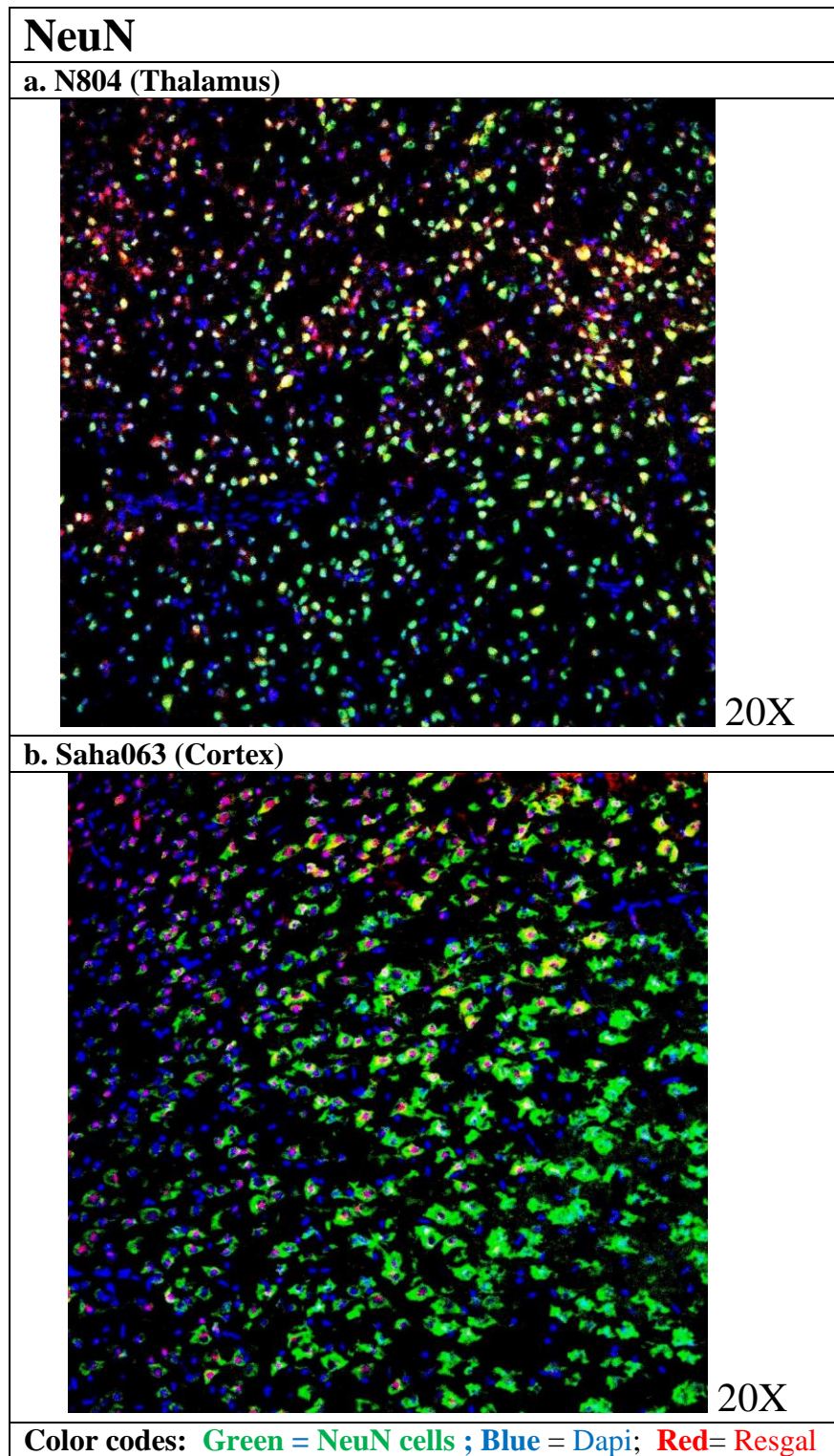
A sample contrast between two thalamic regions of two different lines (**Fig. 4.10**), Rosa26 antisense line and LacG071 line. It represented almost a similar number of cells stained with Res-gal. It also indicated that all the Res-gal signals were nuclear-localized.



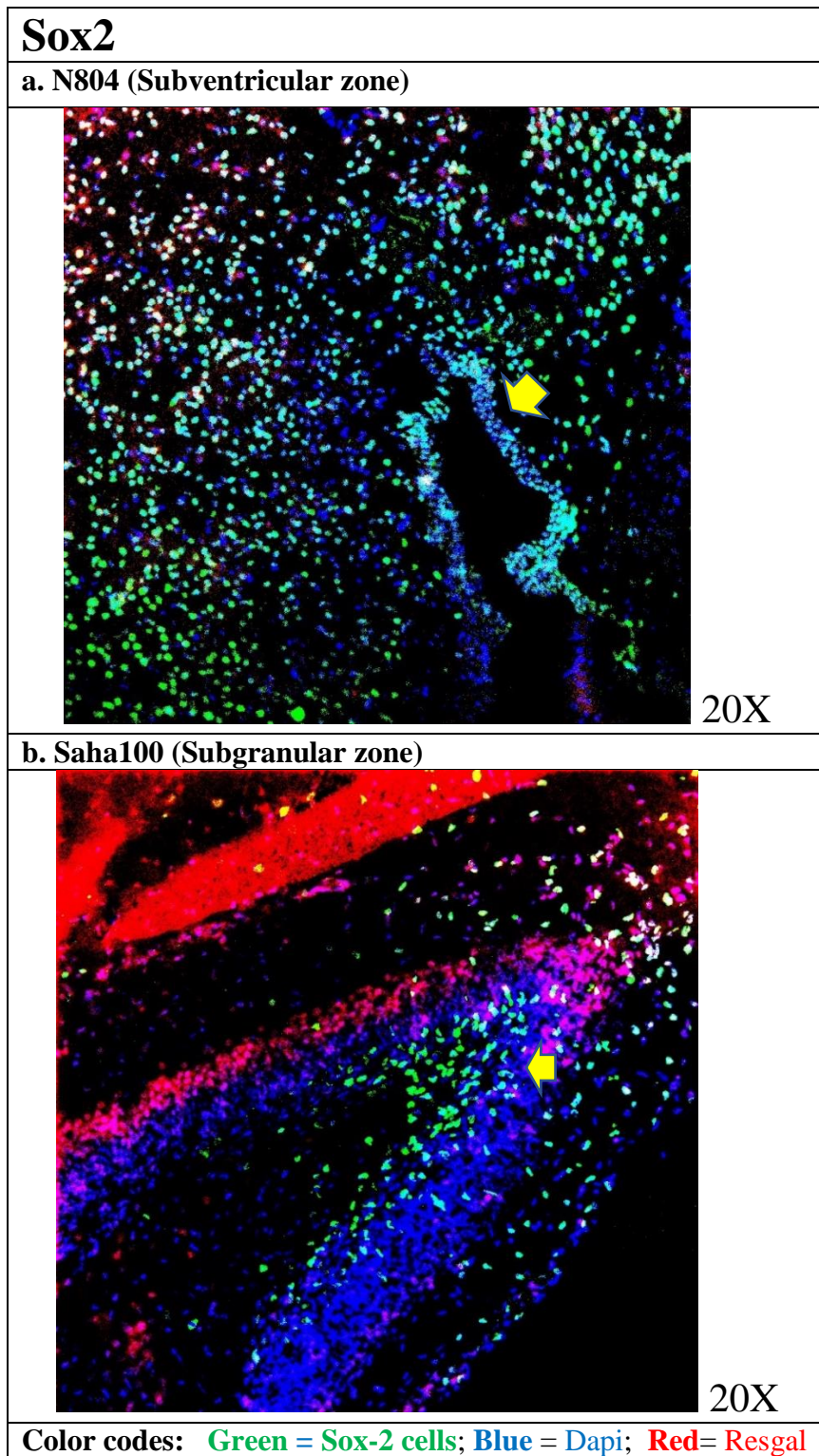
**Figure 4.1.** Detection of the GFAP expressing astrocytes (green) in the hippocampus (a) and hypothalamus (b) of the brain of animals belonging to LacG071 line and the Rosa 26 antisense lines, respectively. In these confocal images, Res-gal fluorescence (red) was found to be co-localized in the DAPI stained nuclei (blue) cells.



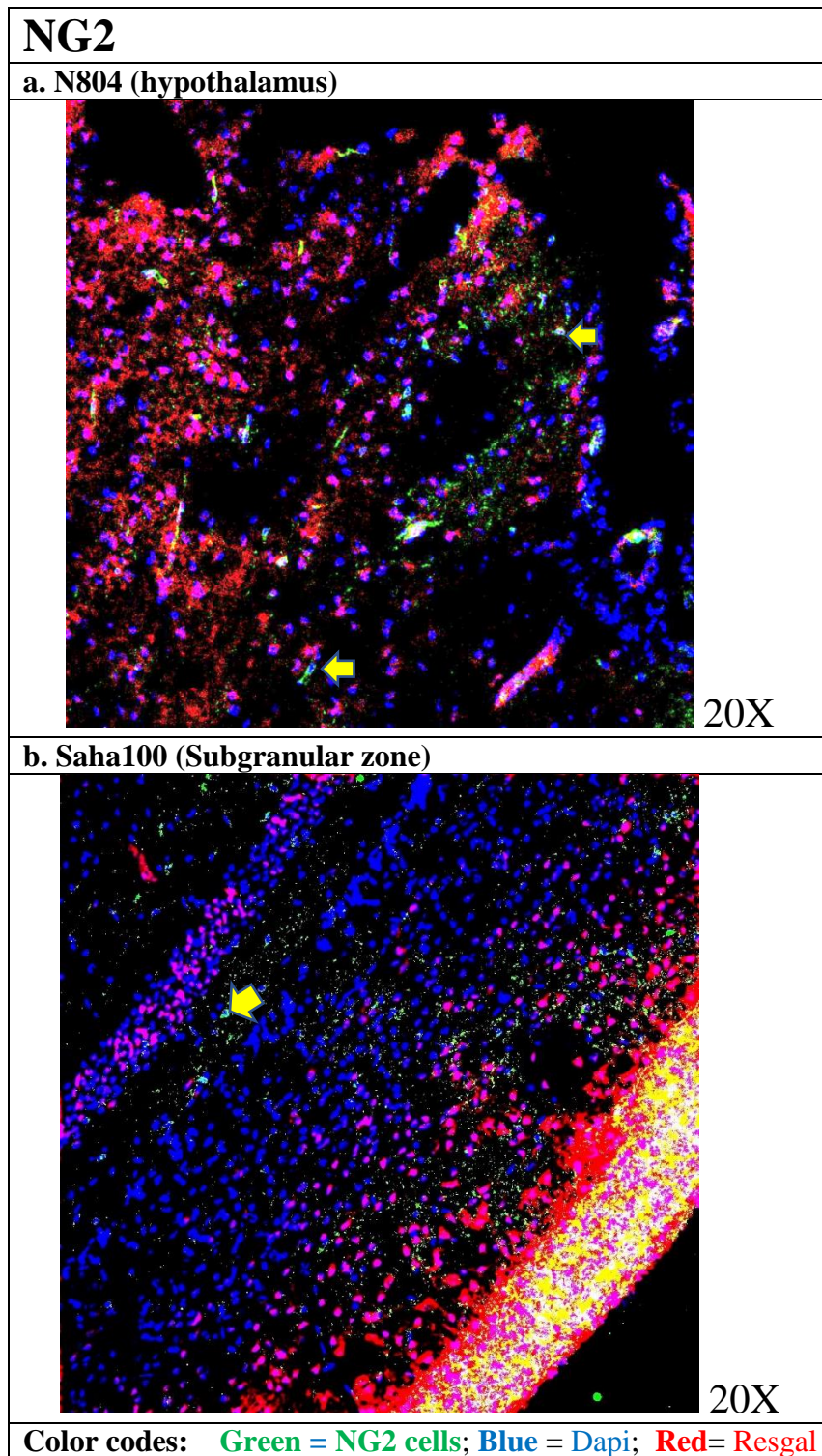
**Figure 4.2:** Detection of the Olig-2 expressing oligodendrocytes (green) in the thalamus (a) and brain stem (b) of the brain of animals belonging to LacG071 line and the Rosa 26 antisense lines, respectively. In these confocal images, Res-gal fluorescence (red) was found to be co-localized in the DAPI stained nuclei (blue) cells.



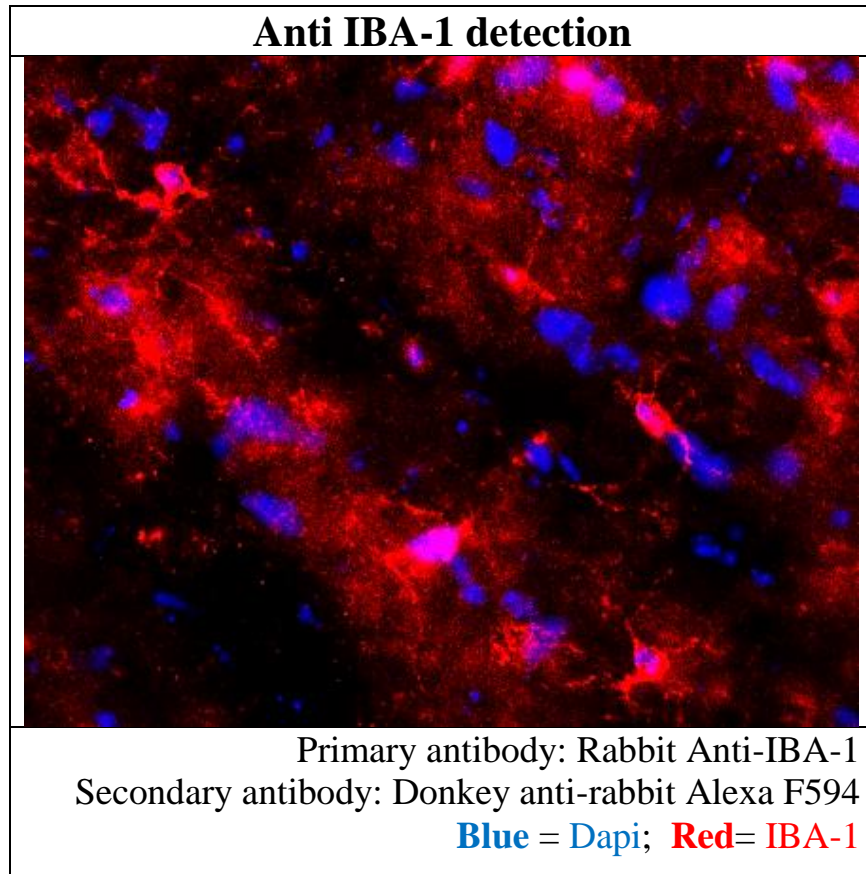
**Figure 4.3.** Detection of the NeuN expressing neurons (green) in the thalamus (a) and brain cortex (b) of the brain of animals belonging to LacG071 line and the Rosa 26 antisense lines, respectively. In these confocal images, Res-gal fluorescence (red) was found to be co-localized in the DAPI stained nuclei (blue) cells.



**Figure 4.4.** Detection of the Sox-2 expressing stem cells (green) in the subventricular zone (a) and subgranular zone (b) of the brain of animals belonging to LacG071 line and the Rosa 26 antisense lines, respectively. In these confocal images, Res-gal fluorescence (red) was found to be co-localized in the DAPI stained nuclei (blue) cells.



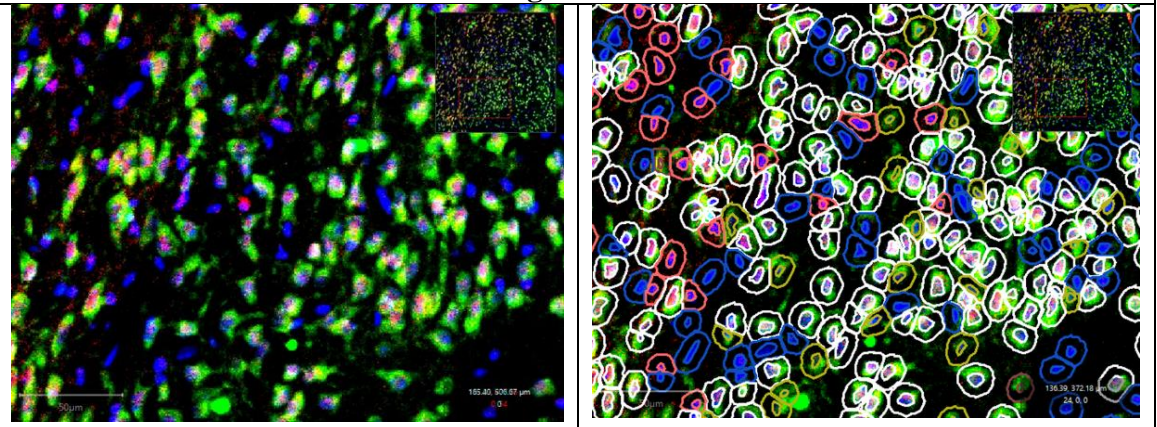
**Figure 4.5.** Detection of the NG2 expressing oligodendrocyte progenitor cells (green) in the hypothalamus (a) and subgranular zone (b) of the brain of animals belonging to LacG071 line and the Rosa 26 antisense lines, respectively. In these confocal images, Res-gal fluorescence (red) was found to be co-localized in the DAPI stained nuclei (blue) cells.

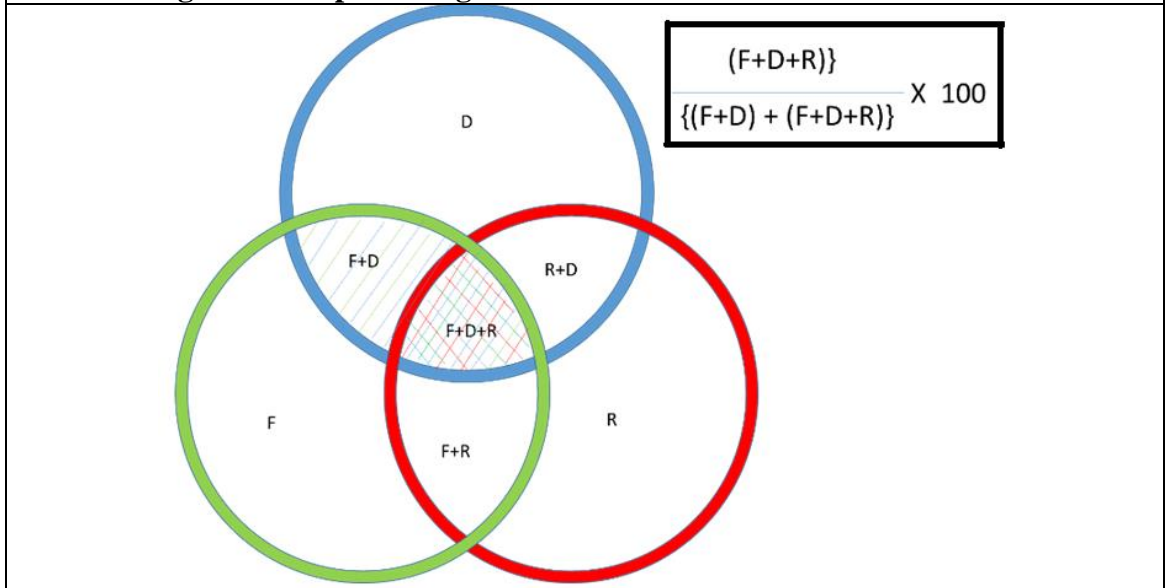


**Figure 4.6.** Detection of the IBA-1 expressing microglial cells (red) in the hypothalamus of the brain of the animal, Saha100, belonging to the Rosa 26 antisense lines. DAPI stained nuclei (blue) cells.

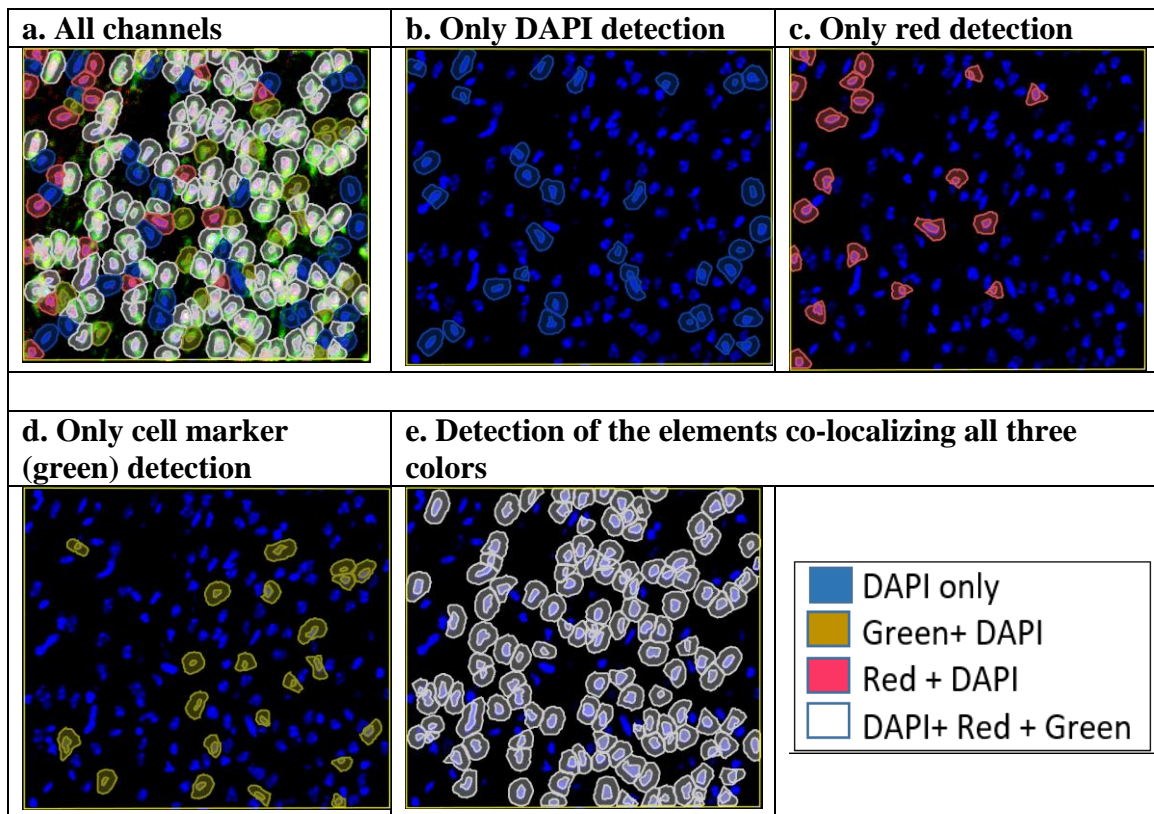
## QuPath detection

### a. Detection of immunofluorescence signals



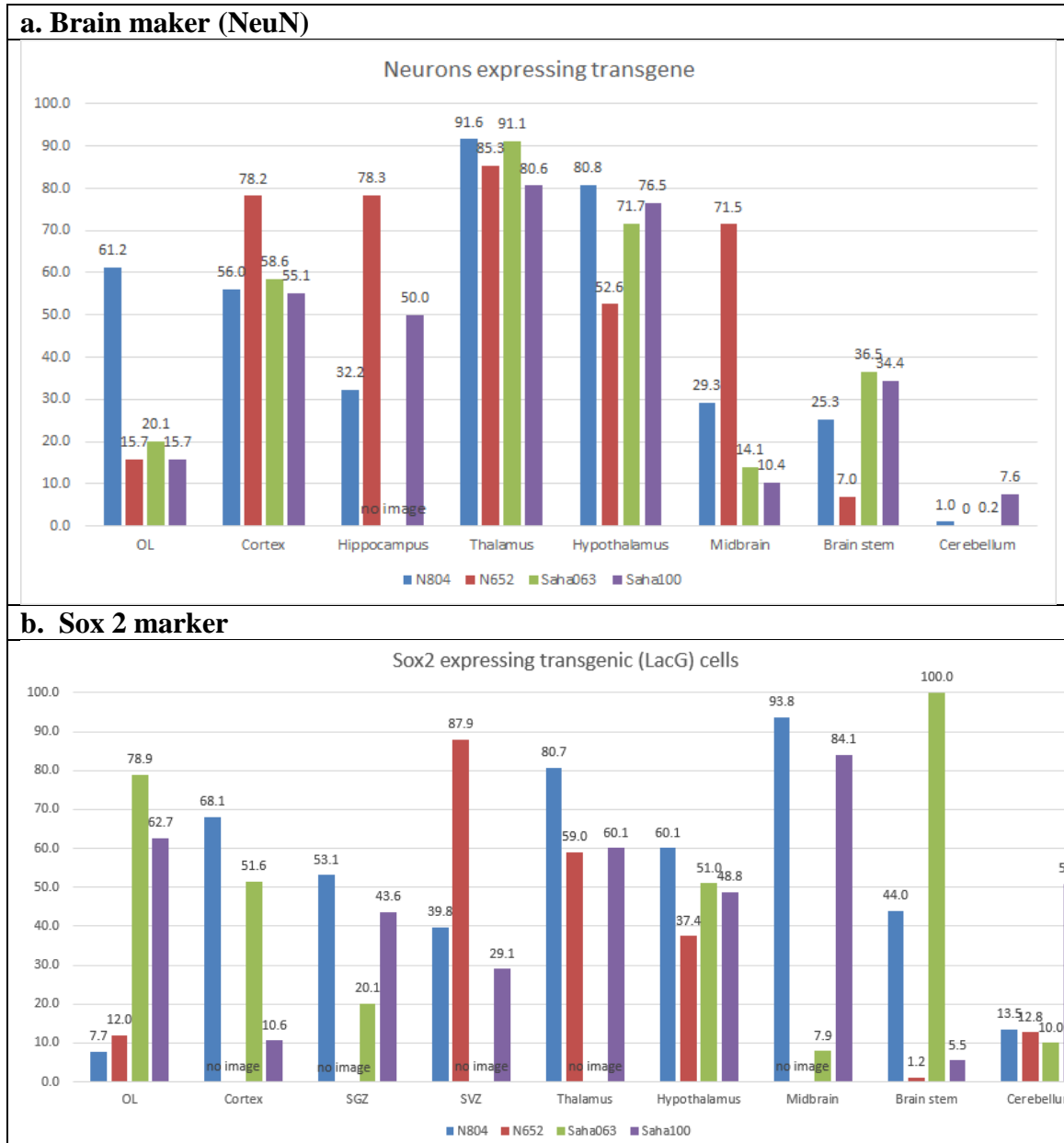
**b. Venn diagram conceptualizing the calculation**


**Figure 4.7. QuPath detection.** (a) A real-time detection of immunofluorescence signals & (b) Venn diagram conceptualizing the calculation of the percentage of transgene expressing brain cell markers. The calculation was made using a mathematical formula mentioned in the inset (where, F= number of cells expressing cell markers, D = DAPI stained nuclei, & R= Res-gal stained cells).

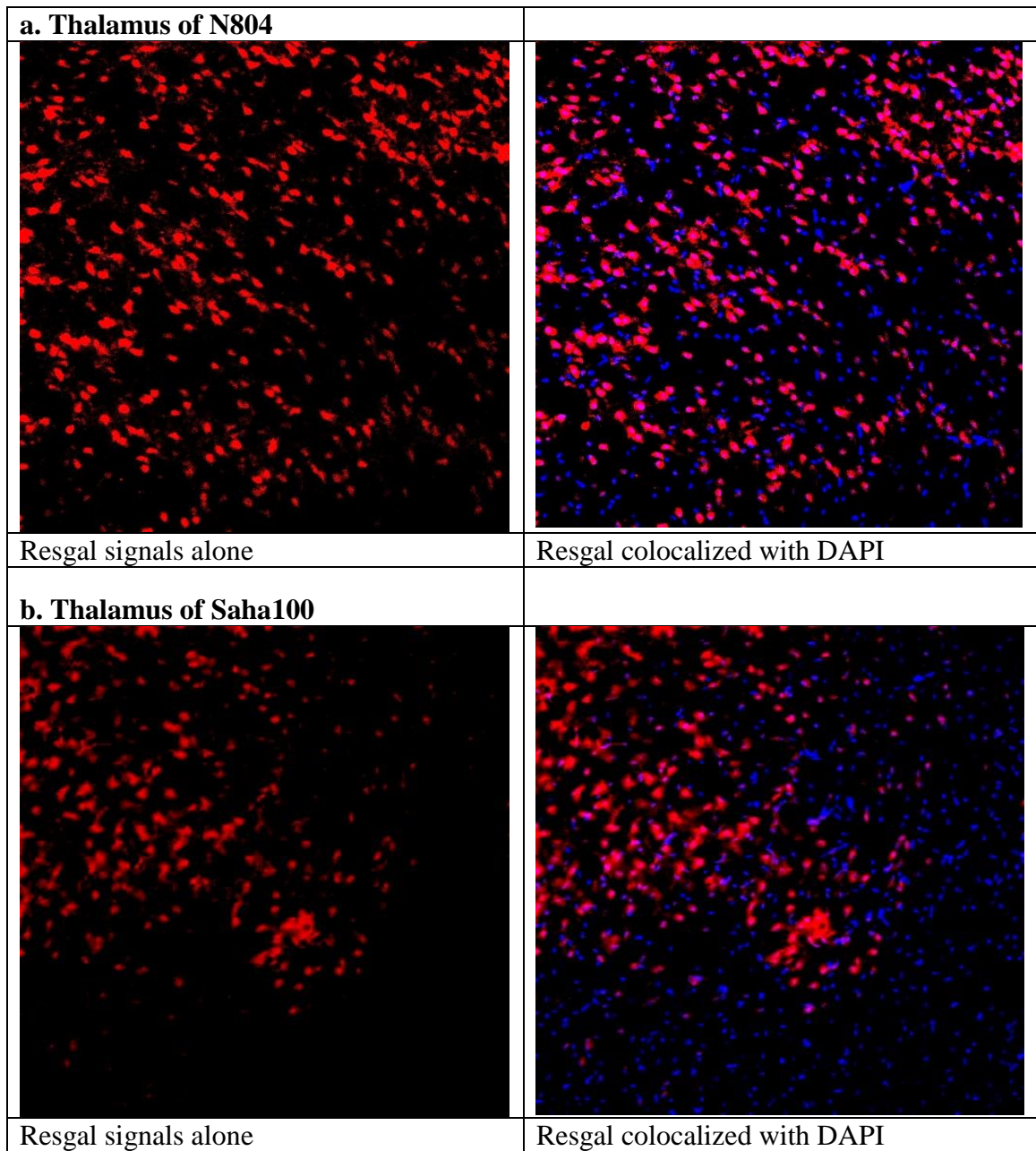




**Figure 4.8.** Effectiveness of QuPath enforced detection. Here, different detections denoted with multiple colors (a) have been segregated into specific condition combining individual colors (b-e).



**Figure 4.9.** Percentage of Res-gal positive cells in different parts of mouse brains expressing markers, (a) NeuN & (b) Sox2.



**Figure 4.10.** The difference in Res-gal signals between two thalamic regions of two different lines, (a) LacG071 & (b) Rosa26 antisense line.

## 4.5 Discussion

The primary idea of this chapter was to detect the transgene positive brain markers. Due to the presence of the LacG reporter cassette in our transgene, our transgene, if translated will translate into the beta-galactosidase enzyme, which can enzymatically break down many chromogenic and fluorogenic substrate. Resorufin  $\beta$ -D-galactopyranoside (or, Res-gal) is a fluorogenic substrate, which when breaks down form a fluorescence emitting substance. Here, we wanted to replace X-gal with this compound since both of them share the same mechanism for detection. Moreover, when we use it in conjunction with DAPI and a secondary antibody with a fluorophore attached to it. In a successful detection of primary antibody, we shall get the overall result in a fluorescent form, again provided that transgene is expressed and the product degrades the Res-gal.

For the *in situ* detection of the cells, samples from two different mouse line were chosen, especially those which shows promoter activity in terms of X-gal signals. Such mouse lines are Rosa26 line and not expressing LacG071 subline. We optimized an immunostaining protocol to use the Res-gal, which we had to incorporate in the final step of incubation with a secondary antibody.

We chose five different common brain markers for detection with the primary antibody. They include the neuronal marker (NeuN), glial markers (NG2, Olig2, and GFAP), and stem cell marker (Sox-2). All could be detected successfully, and their phenotypes were verified with the help of experts. With a similar protocol, we could also successfully detect IBA-1, a microglia marker. However, it had never been used in the presence of Res-gal (Fig. 4.6).

As expected, we obtained the Res-gal signals colocalizing in the nuclear mostly. Therefore, we wanted to detect the percentage of cells that are at the same time detected by cell markers and express Res-gal. These classes of cells are, in other words, transgene expressing specific brain cells.

And we took the help of a pathological quantifying software, QuPath to detect the fluorescence signals separately and effectively. We made generated the script in such a way that DAPI signal as essential criteria to be present for each of the co-localization and ruled out any combination not having DAPI signal excluded.

The images were taken in confocal microscopy from different brain regions. After image processing, we quantified the percentage of cells in each of the brain regions. The results of the analyses (**Fig. 4.9 a & b**) indicated a background problem for the Res-gal. It meant although the Res-gal signals were nuclear-bound, yet we observed the same abundance of signals in both of these mouse lines (**Fig. 4.10**). However, the X-gal data did not turn in with such a result. Therefore, it was concluded that though X-gal and Res-gal share similar degradation pathways by beta-galactosidase, there is a major difference in their sensitivity. It was possible that these substrates got broke down within a short period and had enough time to diffuse some other parts of the same section while being incubated.

Therefore, this technique can be optimized further with different duration of incubation or by finding an alternative substrate with lesser sensitivity.

## 4.6 References

1. Coufal NG *et al.* (2009) L1 retrotransposition in human neural progenitor cells. *Nature*; 460: 1127-1131.

2. De Laat W, Duboule D. (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*; 502: 499–506.
3. Garcia-Perez JL *et al.* (2010) Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature*; 466: 769-773.
4. Levine M. (2010) Transcriptional enhancers in animal development and evolution. *Curr. Biol.* ;20: R754–63.
5. Muotri AR *et al.* (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*; 435: 903-910.
6. Saha PS and An W. (2019) Recently Mobilised Transposons in the Human Genome. In: eLS. John Wiley & Sons, Ltd: Chichester.

## Chapter 5

### General discussion and future directions

In this project, we successfully characterized a few critical aspects of the LINE-1 promoter activity in our transgenic mouse model. We attempted to take a deep insight into the determinant factors, which would alter promoter activity in the various situation with regards to transgenic expression.

First, we were able to show that each integration site for the transgene brought about a distinct change of promoter activity either as stable or silencing. This position effect confirms the critical importance of the site of integration on the level of LINE-1 promoter activity. Surprisingly enough, this position effect, again, in turn, can be governed by changing the orientation of the cassette. This phenomenon could let us understand why some individuals are more susceptible to LINE-1 mediated disorder, while some are unaffected. In other words, this locus and orientation-dependent expression pattern also let us agree that permissiveness of loci and their transcriptional influence on the LINE-1 promoter can possess a risk for some LINE-1 mediated diseases. Therefore, evaluating locus-specific permissiveness can be a better biomarker than carrying out the same to assess global LINE-1 expression or methylation status.

Second, through our endeavor, it was possible to understand that some organs have no, low or high LINE-1 promoter activity – of course depending on the age of the host. To our best knowledge, this was the first approach of this kind to understand the *in vivo* organ-specific regulation with considering development-time a factor. The same technique can be

diversified to more number of organs of this mouse model; thus, we could get better organ-specific information. In addition, RNA and protein detection approaches can be included to validate our staining data. Studying the promoter activity in cancer/other disease model would be helpful to broaden our perspectives on the regulation of the LINE-1 activities in respective conditions.

Third, our current knowledge regarding the cell-types holding the retrotransposition events is limited. Still, it is also unknown when cells lose control to repress these “molecular parasites” in them and when they cannot. The attempt of understanding cell-specific regulation to the LINE-1 promoter activity can be continued. It might give us an insight into how the diverse population of different cell types, not only in the brain but elsewhere in the organ system can promote or repress LINE-1 activity.

The technique relies on a very simple staining technique. It is very handy and cost-effective, and also can be adopted by any laboratory in a short period. The effect of various physiological, pathological, chemical, or environmental factors on LINE-1 mobilization can be assayed. For example, the effect of different carcinogens or potential drug candidates can be screened for their roles in triggering or barring retrotransposition. It is strongly believed that with the help of our technique and mouse model in combination, a wide range of chemicals can be assessed within a short time.

Overall remark regarding this study is, this study may carry consistency with the previous ground-breaking works by Brouha *et al.*, 2003, Pizarro and Cristofari, 2016 and finally, Philippe *et al.*, 2016 (Brouha *et al.*, 2003; Pizarro and Cristofari, 2016; Philippe *et al.*, 2016). Although these works are strong in their context; however, they all commonly cross-talked about the locus-dependency with retrotransposition-competency. Here, we showed

*in situ* that indeed there is a locus-dependency - complicated by age, organ and orientation factors - for LINE-1 to be expressed in mammals.

## 5.1 References

1. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH. (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Sciences of the United States of America* 100:5280–5285.
2. Philippe C, Vargas-Landin DB, Doucet AJ, van Essen D, Vera-Otarola J, Kuciak M, Corbin A, Nigumann P, Cristofari G. (2016) Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *Elife*; 5. pi: e13926. doi: 10.7554/eLife.13926.
3. Pizarro JG, Cristofari G. (2016) Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic Cells. *Frontiers in Cell and Developmental Biology* 4:14.

-----