

**TRANSCRIPTIONAL REGULATION IN THE DEVELOPMENT OF THE
CEREBELLUM AND CEREBELLAR GRANULE CELLS**

by

Peter Zhang

B.Sc., The University of British Columbia, 2006

M.Sc., The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(MEDICAL GENETICS)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2017

© Peter Zhang, 2017

Abstract

The cerebellum is critical for motor functions such as coordination, precision and accurate timing of movement; as well as non-motor functions including cognitive and emotional processes. During cerebellar development, genesis and fate decisions of cerebellar neural precursors are controlled by genetic networks consisting of transcription factors and their downstream targets. The objectives of my thesis are: 1) to construct the transcriptional network of the developing cerebellum based on gene expression and gene regulation; 2) to investigate the temporally regulated usage of alternative promoters in cerebellar development; 3) to generate a cerebellar granule cell specific transcriptome dataset to identify genes that are dynamically expressed, or significantly enriched in cerebellar granule cells during development; and 4) to study the roles of a newly discovered, dynamically expressed transcription factor - Kruppel-like Factor 4 (Klf4) in cerebellar granule cell development.

Taking advantage of high-throughput next generation sequencing technology, we used HeliScopeCAGE, which combines single molecule fluorescent sequencing technology (Helicos) and Cap Analysis of Gene Expression (CAGE), to generate a new transcriptome time series for cerebellar development. We were able to discover hundreds of gene regulators that are important for cerebellar development through differential expression and motif activity analyses. In addition, I analyzed the temporal shift of usage in alternative promoters of a gene, and found that different forms of gene products have distinct functions during cerebellar development. Furthermore, to study the granule cell-specific transcriptome, I used laser microdissection to isolate granule

cells from the cerebellum. Comparison of the granule cell transcriptome with the whole cerebellar transcriptome allowed me to identify genes that are dynamically regulated or significantly enriched in the granule cells. Lastly, I studied a mouse knock-out model of Klf4, a potentially key gene regulator from previous analyses, and found that Klf4 does, in fact, have an important role in early granule cell proliferation. This work also showed that Klf4 is involved in the regulation of other important granule cell transcription factors such as Pax6 in the cerebellum for the first time.

Preface

A version of Chapter 2 is under peer-review with the title “Identification of key regulators for cerebellar development using FANTOM5 time-course CAGE data”. I was one of the co-lead investigators with Thomas Ha and Anthony Mathelier. I was responsible for the design of research, sample collection, data analysis, biological validation and manuscript composition. In addition, Thomas Ha was involved in generating samples for the time series. The FANTOM consortium performed HeliScopeCAGE and data processing. Anthony Mathelier, Tyler Funnel, Wyeth W. Wasserman, Thomas Ha were involved in data analysis (Anthony Mathelier performed differential gene analysis and functional term analysis, represented by Fig 2.3 and 2.4). Thomas Ha was involved in biological validation experiments (including the in utero knock-down experiments, represented by Fig 2.6). Thomas Ha and Anthony Mathelier were involved in writing the manuscript. Dan Goldowitz was the supervisory author on this project and involved in the design of research, data analysis and manuscript preparation.

A version of Chapter 3 is under peer-review with the title “Relatively frequent switching of transcription start sites during cerebellar development”. I was one of the lead investigators with Emmanuel Dimont. I was responsible for the design of research, sample collection, data analysis, biological validation and manuscript composition. In addition, Thomas Ha and Douglas J. Swanson were involved in generating samples for the time series. The FANTOM consortium performed HeliScopeCAGE and data processing. Emmanuel Dimont developed the switching event detection algorithm

which is used to detect promoter switching events. Emmanuel Dimont and Winston Hide were involved in writing the manuscript. Dan Goldowitz was the supervisory author on this project and involved in the design of research, data analysis and manuscript preparation.

A version of Chapter 4 is in revision for publication with the title “Discovery of Transcription Factors Novel to Mouse Cerebellar Granule Cell Development through Laser Capture Microdissection”. I was the lead investigator responsible for the design of research, sample collection, data analysis, biological validation and manuscript composition. In addition, Thomas Ha and Douglas J. Swanson were involved in generating samples for the time series. Erik Arner, Michael de Hoon and the FANTOM consortium performed HeliScopeCAGE and data processing. Dan Goldowitz was the supervisory author on this project and involved in the design of research, data analysis and manuscript preparation.

A version of Chapter 5 has been published [Zhang, Peter et al. “Kruppel-Like Factor 4 Regulates Granule Cell Pax6 Expression and Cell Proliferation in Early Cerebellar Development.” Ed. Chunming Liu. PLoS ONE 10.7 (2015): e0134390. PMC. Web. 13 Dec. 2016.]. I was the lead investigator, responsible for the design of research, data collection, data analysis and manuscript composition. In addition, Thomas Ha, Matt Larouche and Douglas J. Swanson were involved in data analysis and manuscript edits. Dan Goldowitz was the supervisory author on this project and involved in the design of research, data analysis and manuscript preparation.

Table of Contents

| | |
|---|-------------|
| Abstract | ii |
| Preface | iv |
| Table of Contents..... | vi |
| List of Tables..... | xii |
| List of Figures..... | xiii |
| List of Abbreviations..... | xv |
| Acknowledgements..... | xvii |
| Chapter 1 : Introduction | 1 |
| 1.1. Cerebellar development overview | 2 |
| 1.1.1. Developmental neurogenetics of the cerebellum..... | 2 |
| 1.1.2. Cerebellar germinal zones..... | 3 |
| 1.1.3. The development of cerebellar granule neurons | 7 |
| 1.2. Transcriptome analyses of cerebellar development..... | 8 |
| 1.2.1. Cerebellar Development Transcriptome Database (CDT-DB) | 8 |
| 1.2.2. Cerebellar Gene Regulation in Time and Space (CbGRiTS) | 9 |
| 1.2.3. Advances in genomic assays..... | 11 |
| 1.3. Development of the high-throughput technology – HeliScopeCAGE | 11 |
| 1.3.1. Functional Annotation of Mammalian Genomes (FANTOM5)..... | 13 |
| 1.3.2. Generation of mammalian enhancer atlas and discovery of enhancer RNA | 14 |
| 1.3.3. Incorporation of time course data in FANTOM5 | 15 |
| 1.4. Novel study on the switching of alternative transcriptional start sites usage during cerebellar development..... | 17 |
| 1.5. Investigation of granule cell transcriptome with laser capture microdissection.... | 18 |

| | |
|--|-----------|
| 1.6. Kruppel-like factor 4 as an example of novel regulator during cerebellar and granule cell development | 19 |
| 1.7. Four main chapters, Chapter 2 – 5, and their project aims | 21 |
| Chapter 2 : Identification of key regulators for cerebellar development using FANTOM5 time-course CAGE data | 22 |
| 2.1. Introduction | 22 |
| 2.2. Materials & methods | 25 |
| 2.2.1. Sample preparation..... | 25 |
| 2.2.2. Sample quality | 25 |
| 2.2.3. Differential expression analysis | 25 |
| 2.2.4. Functional enrichment analysis | 26 |
| 2.2.5. Gene selection for experimental validation..... | 26 |
| 2.2.6. Design of shRNA and in utero transfection..... | 27 |
| 2.2.7. qRT-PCR for confirmation of expression profiles | 28 |
| 2.2.8. Immunofluorescence for in utero knockdown validation..... | 29 |
| 2.3. Results | 31 |
| 2.3.1. Time-course CAGE data collection and quality control | 31 |
| 2.3.2. Differential expression of cerebellar transcripts | 35 |
| 2.3.3. Functional annotation of differentially expressed genes over cerebellar development..... | 38 |
| 2.3.4. Selection of transcription factors as potential master regulators of cerebellar development..... | 40 |
| 2.3.5. Scrt2, Rfx3 and Atf4 as a master regulator of the cerebellum development. | 42 |
| 2.4. Discussion | 46 |
| 2.4.1 Overview of the FANTOM5 cerebellar time course | 46 |

| | |
|--|-----------|
| 2.4.2 Time course data reveal the expression landscape of cerebellar development transcriptome | 47 |
| 2.4.3 Novel key regulators in cerebellar development | 49 |
| 2.4.4 Concluding Remarks | 50 |
| Chapter 3 : Relatively frequent switching of transcription start sites during cerebellar development..... | 51 |
| 3.1. Introduction..... | 51 |
| 3.2. Materials & methods | 54 |
| 3.2.1. Mouse colony maintenance and breeding | 54 |
| 3.2.2. Tissue processing | 54 |
| 3.2.3. Quality assessment | 55 |
| 3.2.4. Transcriptome library generation by HeliScopeCAGE | 55 |
| 3.2.5. TSS switch detection | 56 |
| 3.2.6. Gene Ontology analysis for gene with crossover switching events | 59 |
| 3.2.6. In silico validation of gene expression with established databases and experimental validation with gene structure prediction and quantitative real-time PCR..... | 59 |
| 3.3. Results | 61 |
| 3.3.1. Overview of promoter switch events during cerebellar development | 61 |
| 3.3.2. Distribution of TSS switching events in cerebellar transcriptome | 64 |
| 3.3.3. Gradual increment in the number of crossover TSS switching events over developmental time..... | 67 |
| 3.3.4. Gene Ontology analysis for genes with the most significant crossover TSS switching events | 70 |
| 3.3.5. Validation of promoter switching events..... | 72 |
| 3.4. Discussion | 80 |

| | |
|--|-----------|
| 3.4.1. High prevalence of alternative TSSs in mammalian genomes | 80 |
| 3.4.2. Temporal regulation of alternative TSS associated with developmental processes in the cerebellum..... | 81 |
| 3.4.3. Alternative TSS as post-transcriptional control during cerebellar development | 83 |
| 3.4.4. Functional importance of alternative TSS during cerebellar development | 84 |
| 3.4.5. Conclusion..... | 85 |
| Chapter 4 : Discovery of transcription factors novel to mouse cerebellar granule cell development through laser capture microdissection | 86 |
| 4.1. Introduction..... | 86 |
| 4.2. Materials & methods | 89 |
| 4.2.1. Mouse colony maintenance and breeding | 89 |
| 4.2.2. Tissue processing and sectioning | 89 |
| 4.2.3. Laser capture microdissection | 92 |
| 4.2.4. Quality assessment | 93 |
| 4.2.5. Transcriptome library generation by HeliScopeCAGE..... | 93 |
| 4.2.6. Bioinformatics analysis..... | 94 |
| 4.2.7. Prediction of motif activity with Motif Activity Response Analysis (MARA).... | 94 |
| 4.2.8. In silico validation of gene expression with three online databases | 95 |
| 4.2.5. Quantitative real-time PCR..... | 96 |
| 4.3. Results | 97 |
| 4.3.1. Overview of the dataset | 97 |
| 4.3.2. Temporally regulated granule cell transcripts..... | 97 |
| 4.3.3. Granule cell enriched transcripts | 99 |
| 4.3.4. Temporal and spatial confirmation of gene expression using in situ hybridization databases | 100 |

| | |
|--|------------|
| 4.3.5. Gene Ontology analysis for granule cell enriched genes | 107 |
| 4.3.6. Motif Activity Response Analysis | 108 |
| 4.3.7. Experimental validation | 112 |
| 4.4. Discussion | 114 |
| 4.4.1. The cerebellar granule cell precursor transcriptome | 114 |
| 4.4.2. Confirmation of known genes in cerebellar granule cell development | 117 |
| 4.4.3. Genes with limited information pertaining to cerebellar development | 118 |
| 4.4.4. Discovery of novel transcription factors that may be involved in granule cell development..... | 120 |
| 4.4.5. Summary | 121 |
| Chapter 5 : Kruppel-like factor 4 regulates granule cell Pax6 expression and cell proliferation in early cerebellar development | 122 |
| 5.1. Introduction..... | 122 |
| 5.2. Materials & methods | 126 |
| 5.2.1. Expression analysis and transcription factor binding site prediction | 126 |
| 5.2.2. Klf4 colony maintenance and breeding | 127 |
| 5.2.3. Histological methods and analysis | 128 |
| 5.2.4. Assays for cell proliferation and cell death..... | 129 |
| 5.2.5. Real-time PCR..... | 130 |
| 5.3. Results | 131 |
| 5.3.1. Discovery of Klf4 as a regulator of Pax6 in the cerebellum..... | 131 |
| 5.3.2. Characterization of Klf4 expression | 131 |
| 5.3.3. Pax6 in the developing cerebellum following the elimination of Klf4 expression | 135 |
| 5.3.4. Klf4's roles in the developing cerebellum..... | 138 |
| 5.3.5. Investigation on functional redundancy in the Klf family..... | 142 |

| | |
|--|------------|
| 5.4. Discussion | 146 |
| 5.4.1. Klf4 regulates Pax6 expression | 146 |
| 5.4.2. Klf4 as a regulator of cell proliferation | 148 |
| 5.4.3. Temporal specificity of Klf-null phenotype | 150 |
| 5.4.4. Comparison of the Klf4-null with the Pax6-null cerebellum..... | 151 |
| 5.4.5. Conclusions..... | 152 |
| Chapter 6 : Discussion | 153 |
| 6.1. Overview of objectives..... | 153 |
| 6.1.1. Utilization of cerebellar HeliScopeCAGE time series toward understanding neurological disorders: Rett Syndrome | 153 |
| 6.2. Construction of transcriptional network in cerebellar development..... | 155 |
| 6.2.1. Role of promoter activity and enhancer activity in gene regulation during cerebellar development..... | 156 |
| 6.3. Extensive alternative promoter usage during cerebellar development..... | 159 |
| 6.4. Temporally regulated and granule cell enriched transcription factors | 162 |
| 6.4.1. The advantage of LCM to obtain enriched EGL samples..... | 163 |
| 6.4.2. Limitations of LCM | 164 |
| 6.4.3. Application of LCM in transcriptome studies | 165 |
| 6.5. Klf4 as an important transcriptional regulator during cerebellar development ... | 166 |
| 6.6. Future directions in the study of cerebellar development utilizing our HeliScopeCAGE database | 169 |
| 6.6.1. Comparative bioinformatic analyses of the FANTOM5 (Zenbu) and CbGRiTS databases | 169 |
| 6.6.2. Discovery of novel genetic elements and their regulatory interactions | 169 |
| 6.7. Concluding remarks | 172 |
| References | 173 |

List of Tables

| | |
|---|-----|
| Table 1.1. A summary of currently available mouse cerebellar time-course databases .. | 8 |
| Table 3.1. Comparison of TSS switching events during cerebellar development with other FANTOM5 datasets | 63 |
| Table 3.2. Top 20 genes with highest numbers of TSS switching events | 67 |
| Table 3.3. Distribution of crossover TSS switching events across time in cerebellar development (N=9,767)..... | 69 |
| Table 3.4. Cerebellar expression patterns of genes with most significant switching events from the in situ database, Genepaint and ABA | 79 |
| Table 4.1. External germinal layer (EGL) gene transcripts that show temporal regulation. | 98 |
| Table 4.2. External germinal layer (EGL) cell enriched transcripts (>2x expressions). 100 | |
| Table 4.3. External germinal layer (EGL) cell temporally regulated transcription factors (L-L) and EGL enriched transcription factors (L-C)..... | 102 |
| Table 4.4. Expression of EGL expressed TFs in wild-type mice compared with expression in the Atoh1 KO at CbGRiTS microarray database (http://cbgrits.org). | 104 |
| Table 4.5. Gene Ontology analysis of granule cell enriched gene..... | 108 |
| Table 4.6. Twenty-six motifs that showed significant change in motif activity (shown in z-value >1.7) during cerebellar development discovered with MARA | 111 |
| Table 5.1. The expression of Klf family members in mouse cerebellum..... | 143 |

List of Figures

| | |
|---|-----|
| Figure 1.1. Schematic illustration of the development of mouse cerebellum | 6 |
| Figure 2.1. Overview of FANTOM5 HeliScopeCAGE transcriptome dataset..... | 33 |
| Figure 2.2. Clustering analysis of the cerebellar development time-points. | 34 |
| Figure 2.3. Differentially expressed genes along the time-course. | 37 |
| Figure 2.4. EnrichR analysis of differentially expressed genes in cerebellar development | 39 |
| Figure 2.5. Validation of HeliScopeCAGE expression for six differentially expressed transcription factors with qRT-PCR..... | 41 |
| Figure 2.6. Phenotypic changes observed by Scrt2 knock-down during cerebellum development..... | 44 |
| Figure 3.1. A schematic diagram of alternative transcription start sites (TSSs) and the classes of TSS switching. | 58 |
| Figure 3.2. Overview of TSS switching events during cerebellar development..... | 62 |
| Figure 3.3. Distribution TSS switching events in different genes during cerebellar development..... | 66 |
| Figure 3.4. GO Analysis for genes significant ($p < 0.05$) for crossover switching at all time points (left) and at three selected time points (right) | 71 |
| Figure 3.5. Alternative TSSs in glypican 6 (Gpc6) and experimental validation of its non- crossover switching events with Real-time PCR..... | 74 |
| Figure 3.6. Alternative TSSs in acidic nuclear phosphoprotein 32 family, member A (Anp32a) and experimental validation of its crossover switching events with Real-time PCR | 76 |
| Figure 3.7. Alternative TSSs in contactin associated protein-like 2 (Cntnap2) and experimental validation of its crossover switching events with Real-time PCR..... | 78 |
| Figure 4.1. Experimental schematic diagram and EGL tissue collection with laser capture microdissection | 91 |
| Figure 4.2. In situ hybridization expression pattern (from ABA) of three genes that were found to be significantly enriched in the EGL LCM material. | 106 |
| Figure 4.3. Nine genes with significant changes in motif activity during cerebellar development discovered with Motif Activity Response Analysis. | 110 |

| | |
|--|-----|
| Figure 4.4. Eight genes significant in the differential expression analysis and their quantitative real-time PCR validation at E13, E15 and E18..... | 113 |
| Figure 5.1. Klf4 expression in the cerebellum and its co-expression with Pax6 | 134 |
| Figure 5.2. Pax6 is down-regulated in Klf4-null cerebellum at E13.5 and E15.5..... | 136 |
| Figure 5.3. Quantification of Pax6 cell number and expression down-regulation in Klf4-null cerebellum..... | 137 |
| Figure 5.4. Effects of Klf4-knockout on cell death and/or cell proliferation in the developing cerebellum | 140 |
| Figure 5.5. Klf4 has dual effects on the proliferation of epithelial cells in the cerebellum at E15.5..... | 141 |
| Figure 5.6. The expression levels of genes involved in complementary cell proliferative pathways in the Klf4-null with real-time PCR..... | 144 |
| Figure 5.7. The expression levels of genes involved in alternative cell proliferative pathways in the Klf4-null with real-time PCR..... | 145 |

List of Abbreviations

| | |
|---------------------|--|
| CAGE | Cap analysis gene expression |
| CbGRiTS | Cerebellar gene regulation in time & space |
| CDT-DB/Braintx | Cerebellar development transcriptome database |
| DE | Differential Expression |
| E# | Embryonic day # |
| EEL | Enhancer element locator |
| EGFP | Enhanced green fluorescent protein |
| EGL | External germinal layer |
| FANTOM | Functional annotation of mammalian genome |
| GC/GCs | Granule cell(s) |
| GO | Gene ontology |
| HeliScopeCAGE/hCAGE | A technology that coupled CAGE with Helicos sequencing |
| IGL | Internal granular layer |
| LCM | Laser capture microdissection |
| MARA | Motif activity response analysis |
| NE | Neuroepithelium |
| NTZ | Nuclear transitory zone |
| P#/N# | Postnatal/Neonatal Day # |
| PC/PCs | Purkinje cell(s) |
| RL | Rhombic lip |
| Sey | Small eye |
| shRNA | Short hairpin RNA |

| | |
|----------|-----------------------------------|
| TF/TFs | Transcription factors(s) |
| TFBS | Transcription factor binding site |
| TSS/TSSs | Transcription start site(s) |
| UBC/UBCs | Unipolar brush cell(s) |
| VZ | Ventricular zone |

Acknowledgements

I would like to express my sincere gratitude to my research supervisor Dr. Dan Goldowitz for his continuous support of my Ph.D. research, and for his patience, motivation, and immense knowledge. His guidance helped me in my Ph.D. research and writing of this thesis.

I would like to thank the rest of my thesis committee: Dr. Wyeth Wasserman, Dr. Tim O'Connor, and Dr. Christian Naus, for their insightful comments which incited me to improve various aspects of my research and writing of this thesis.

I would like to thank my fellow lab mates J. Yeung, J. Cairns, S. Tremblay, A. Poon; T. Ha, D. Swanson, M. Larouche, J. Wilking, G. Mak; D. Rains and J. Boyle for technical supports, stimulating discussions and for all the fun we have had during my Ph.D. study. I would like to thank F. Lucero Villegas for animal management; Also, I would like to thank my best friend J. Heppner for patiently going over this thesis and correcting countless grammatical errors.

Last but not the least, I would like to thank my family - my parents and my beloved wife for supporting me throughout my Ph.D. study and my life in general.

Chapter 1 : Introduction

The cerebellum is essential for motor control such as movement coordination and balance[1] and non-motor roles - such as sensory perception, cognition and social behavior[2-4]. The cerebellum has been found in all vertebrates, and its basic structures are strikingly conserved among fish, reptiles, birds and mammals[5], suggesting the cerebellum's critical role in basic brain function. Several complex genetic disorders of the nervous system are associated with cerebellar development, including Joubert syndrome (characterized by congenital malformation of the brainstem, cerebellum, and the cerebellar peduncles[6]), Dandy-Walker malformation (characterized by a severe hypoplastic cerebellar vermis and an enlarged fourth ventricle[7, 8]), and Ponto-cerebellar hypoplasia (characterized by the progressive atrophy of cerebellum[9]). Connection of the cerebellum with the diverse phenotypes of these diseases further indicates the essential and multifaceted role of the cerebellum in the central nervous system (CNS). During normal cerebellar development, genesis and fate decisions of cerebellar neural precursors are controlled by genetic networks consisting of transcription factors and their downstream targets. The primary objective of my thesis is to identify transcriptional regulators that are important for the development of the cerebellum.

1.1. Cerebellar development overview

1.1.1. Developmental neurogenetics of the cerebellum

Development of the cerebellum begins during the early embryonic stages in vertebrates[1]. The cerebellum, as does the CNS, arises from a homogenous sheet of epithelial cells, known as the neural plate, induced to differentiate during gastrulation. The neural plate edges thicken and roll up on the antero-posterior axis, and eventually close dorsally to form the neural tube. This process involves the formation, at the anterior portion of the neural tube, of the three primordial vesicles; the prosencephalic, mesencephalic and rhombencephalic vesicles, which later develop into the forebrain, midbrain and hindbrain, respectively. The rhombencephalic vesicle consists of a rostral region (metencephalon) and a caudal region (myelencephalon). The dorsal region of the metencephalon has been identified as the origin of the cerebellum by Wilhelm His (1890), who further hypothesized that from this region the cerebellum develops as a bilateral organ, which would subsequently fuse at the dorsal midline in a rostral-to-caudal direction, to form a single primordium[10]. This theory has since become supported by numerous research groups[11-13].

The anatomic developmental processes described above are dependent on regulatory control by key transcription factors. Previous cellular fate mapping and transplantation studies have revealed that the cerebellum is specified by the expression of Homeobox protein Hox-A2 (Hoxa2) in posterior part of Rhombomere 1, and Orthodenticle homeobox 2 (Otx2), in anterior portion of the Rhombomere[14].

The expression of Otx2 is subsequently regulated by members of the Wnt[15] and fibroblast growth factor (FGF) families, including Wnt1, Fgf8, and Fgf15[16-18]. Homeobox genes - En1 and Pair-box gene Pax2, 5 and 8 are also required to establish the cerebellar territory through a gradient of gene expression[19, 20].

1.1.2. Cerebellar germinal zones

The production of cerebellar neurons (neurogenesis) occurs in two locations: the neuroepithelium (NE) along the roof of the fourth ventricle and the rhombic lip (RL) at the posterior aspect of the cerebellum (Figure 1.1). The NE gives rise to the principal output neurons of the cerebellar cortex known as Purkinje cells in addition to the inhibitory neurons of the cerebellar nuclei, and more than half a dozen types of cerebellar interneurons, including Golgi, basket, and stellate cells[21-23]. Neurogenesis of GABAergic neurons of the cerebellum occurs in the NE and is dependent on the coordinated action of several transcription factors. The transcription factor, Ptf1a, is required for the specification of the cerebellar GABAergic neuron progenitors (including Purkinje cells and interneurons) in the cerebellar neuroepithelium[24, 25]. Ptf1a has been suggested to be one of the most important factors for the specification of NE-derived neurons, and mutations of this gene in humans have been associated with cerebellar agenesis[26]. In the mouse, the earliest cerebellar progenitors exit the cell cycle in the NE at approximately embryonic (E) day E11 as they differentiate to generate the progenitors of the GABAergic Purkinje cells[27]. In addition, the cerebellar NE generates the precursors of cerebellar nuclear neurons, marked by the expression of the

transcription factors Lhx2/9, Meis 1/2, and Irx3[28]. Starting at E11.5, Purkinje cell (PC) precursors, identified by expression of the LIM transcription factors Lhx1 and Lhx5, migrate radially along the emerging glial fiber system to form a superficial zone – the Purkinje plate[28].

In addition to GABAergic cerebellar neurons, the NE also give rise to non-neuronal cells such as the Bergmann glia, marked by the expression of Growth and Differentiation Factor 10 (Gdf10,[29, 30]) starting at E13.5. These Gdf10-positive Bergman glial progenitors actively migrate toward the pial surface and then inwardly to the Purkinje cells, finally situating in the Purkinje cell layer[31]. In postnatal cerebellar development, Bergmann fibers act as essential guide rails for the migration of granule cells (GCs)[32], the elaboration of PC dendrites[33] and stabilization of GC-PC synaptic connections[34].

The second germinal zone forms along the posterior aspect of the cerebellar primordium, which generate the glutamatergic cells of the cerebellum including the granule neurons[35], the unipolar brush cells[21, 36] as well as a subpopulation of neurons of the cerebellar nuclei[37]. Differentiation of neurons from this zone are specified by the basic helix-loop-helix (bHLH) transcription factor Atoh1/Math1[37-39], the zinc finger protein Zic1 [40], and expressing markers such as Meis1, Pde1c, and Pcsk9[28].

Fate mapping experiments have confirmed that a portion of Math1-positive progenitors in the rhombic lip (the cerebellar nuclei progenitors) migrate dorsally along the surface of cerebellar primordium and later become a subpopulation of the cerebellar nuclear neurons[28, 37]. The anterior rhombic lip also generates neuronal

precursors that migrate ventrally, where they form the lateral pontine nucleus, cochlear nucleus, and hindbrain nuclei of the cerebellum[14, 28, 37, 41-43].

The unipolar brush cells (UBCs) are also produced within the rhombic lip during the late embryonic and perinatal periods[44], through specific expression of the transcription factor Eomesodermin/Tbr2[45, 46]. After neurogenesis, a sub-population of UBCs migrate toward the internal granular layer (IGL) via a narrow channel between the developing cerebellar cortex and the ventricular zone, then through the cerebellar white matter [21]. Most UBCs reach the IGL by postnatal day 10 (P10) in mouse and begin maturation throughout the first postnatal month upon arrival[47]. In addition, other Tbr2+ UBCs migrate along the ventricular zone toward the brainstem to enter the cochlear nuclei – the secondary location in which UBCs mature[48].

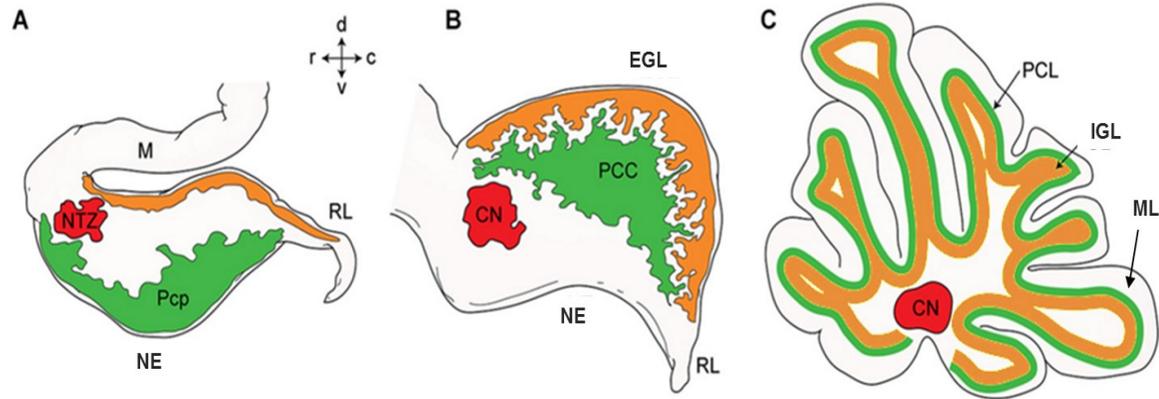


Figure 1.1. Schematic illustration of the development of mouse cerebellum

Cerebellar neurons arise from two distinct germinal zones: the neuroepithelium [NE, also known as the ventricular zone (VZ)] and rhombic lip (RL). The *Ptf1a*-expressing NE gives rise to GABAergic cerebellar nuclear (CN) neurons, Purkinje cells (PC, *labeled in green*) and GABAergic inhibitory interneurons (IN). The *Math1*-expressing RL produces glutamatergic CN neurons (*labeled in red*), granule cells (GC *labeled in orange*) and unipolar brush cells (UBC). Mb, midbrain; Rp, roof plate.

A) Early Embryonic Stage (E11-13). M: Mesencephalon, RL: Rhombic Lip, NE: Neuroepithelium, Pcp: Purkinje cells precursor, NTZ: Nuclear Transitory Zone

B) Late Embryonic Stage (E16-E18). EGL: External Germinal/Granular Layer, RL: Rhombic Lip, NE: Neuroepithelium, PCC: Purkinje Cell Clusters/plate, CN: Cerebellar Nucleus

C) Adult Stage (P21+). IGL: Internal Granular Layer, PCL: Purkinje Cell Layer, ML: Molecular Layer, CN: Cerebellar Nucleus

Figure adapted from [49]

1.1.3. The development of cerebellar granule neurons

In murine development, beginning at E12.5 rhombic lip progenitor cells migrate onto the dorsal surface of the cerebellar primordium to form the external granule layer (EGL)[23, 50]. The progenitor cells in the EGL are highly proliferative and give rise to the abundant cerebellar granule cells[38, 51].

Molecular genetic studies have demonstrated that Math1[35] and MycN[52] are required for granule cell specification, and expansion of the pool of granule cell precursors (GCPs) in the early postnatal period of development. Atoh1, also known as Math1, is the most critical and well-studied factor in rhombic lip development. The importance of Math1 is evident through KO studies showing abolishment of the granule cell progenitors[35]. Studies of transcriptional regulation of Math1 have shown that Zic1 binds a conserved site within the sequence of the Math1 enhancer region, and represses Math1 transcription by blocking the autoregulatory activity of Math1[53].

Starting at E12.5, the granule cells in the EGL, which can be clearly marked by the TF Pax6[54], undergo extensive proliferation regulated by multiple molecule, such as the Wnt1[55] and Shh[56], that are secreted from granule cells and Purkinje cells, respectively. Shortly after birth, granule cell precursors begin to express the differentiation marker NeuroD1, and undergo a major modification of GC morphology[57]. The mature GCs switches from tangential migration to radial migration from the EGL toward the ML. During this process, the characteristic 'T-shaped' parallel fiber forms, and the cell soma of GCs start migrating across the ML towards the IGL[58]. The continuous GC differentiation and radial migration during

postnatal development results in the expansion of the IGL and gradual disappearance of the EGL. At around P21, the EGL disappears and all granule cells complete their migration into the IGL[1] Although the cerebellum takes up only 10% of total brain volume, it contains roughly half of all neurons in the brain due to the large number of the granule cells within the IGL. In humans, approximately 45 billion granule neurons are produced during cerebellar development, out of approximately 110 billion neurons in the brain[59].

1.2. Transcriptome analyses of cerebellar development

Table 1.1. A summary of currently available mouse cerebellar time-course databases

| Thesis Section | 1.2.1 | 1.2.2 | 1.3 – 1.5 |
|------------------------|---|---|---|
| Project Name | Cerebellar Development Transcriptome Database | Cerebellar Gene Regulation in Time and Space | Functional Annotation of Mammalian Genomes Phase 5 |
| Project abbreviation | CDT-DB | CbGRiTS | FANTOM5 |
| Year of establishment | 2008 | 2015 | 2015 |
| Platform | Microarray | Microarray | HeliScopeCAGE |
| Development time frame | E18 – P56 | E11 – P9 | E11 – P9 |
| Number of time points | 8 | 13 | 12 |
| Web address | http://www.cdtdb.neuroinf.jp | http://www.cbgrits.ca | http://fantom.gsc.riken.jp |
| Reference | [60] | [61] | [62] |

1.2.1. Cerebellar Development Transcriptome Database (CDT-DB)

The cerebellum is a relatively simple, anatomically discrete, and well-studied part of the mammalian brain. Thus, it is an exceptional model for studying the central nervous system. One of the first cerebellar developmental databases was the Cerebellar Development Transcriptome Database (CDT-DB <http://www.cdtdb.brain.riken.jp>, see Table 1.1 above), published in 2008[60]. It was constructed to understand the genetic basis underlying cerebellar circuit development through spatial-temporal gene expression data on a genome-wide

basis. This resource-rich database consisted of in situ hybridization, GeneChip, microarray and RT-PCR data at 8 perinatal and postnatal time points in mouse (E18, P0, P3, P7, P12, P15, P21 and P56), which were integrated into a web-based knowledge resource, with links to relevant information at various database websites. Based on gene ontology analysis of the most differentially expressed genes, CBT-DB identified at least 186 synapse-related genes during first three weeks of birth[60]. In addition, microarray experiments revealed more than three hundred transcription factors that were differentially expressed during postnatal cerebellar development[60]. Altogether, the CDT-DB provides a unique informatics tool for mining both spatial and temporal patterns of gene expression in developing mouse cerebellum.

1.2.2. Cerebellar Gene Regulation in Time and Space (CbGRiTS)

Another microarray database – Cerebellar Gene Regulation in Time and Space (CbGRiTS, www.cbgrits.ca see Table 1.1 above) was completed in 2015[61], is one of the largest transcriptome databases in cerebellar developmental research consisting of microarray data from embryonic times (every 24 hours from E11 to E19) as well as postnatal time points (every 72 hours from P0 to P9) with over 15 million gene expression measurements from two inbred strains of mouse [C57BL/6J (B6) and DBA/2J (D2)][61]. Furthermore, CbGRiTS includes transcriptome data from fourteen recombinant inbred mice between B6 and D2 and three mutant lines of mice whose mutant genes are known to target cerebellar granule cells - Math1 KO, Pax6 KO and the meander tail mutant[61]. The authors of CbGRiTS describe it

as “an accessible, multi-level, and interactive platform to explore and examine gene expression patterns of developmentally regulated genes in time and space and an outstanding resource for functional investigation of gene regulatory networks in cerebellar development.” It allows the integration of the “time series microarray resource, bioinformatics analyses, morphological data, 3D modeling and other tools such as graphing, gene list function, and link-out utility to other anatomical ISH databases in a single web-based database. Together, these features and tools implemented in the CbGRiTS database are a powerful resource for neurodevelopmental biologists to explore data, generate hypotheses and test these hypotheses relative to the genetic underpinnings of cerebellar development” [61].

Since its establishment, CbGRiTS has become a stepping stone in understanding the function of key regulators in cerebellar development. For example, CbGRiTS contained data that showed in the Pax6 KO, cerebellar expression of the gene *wntless* (*Wls*) was upregulated at times when its expression normally decreases in the wild-type cerebellum[63]. Immunocytochemical analysis showed that this was likely due to an expansion of *Wls*-expressing cells into regions that are normally colonized by Pax6-expressing cells, indicating a negative interaction between *Wls*-expressing cells and Pax6-positive cells. These findings led Yeung et al. to discover the compartmentalization of the cerebellar rhombic lip by the mutual inhibitory regulation between Pax6 and *Wls*[63]. In addition to findings from up-regulation of *Wls*, the down-regulation of *Tbr1* and *Tbr2* in the Pax6 KO, as found in CbGRiTS, led to new discoveries on the essential roles of Pax6 on the

survival of glutamatergic cerebellar nuclear neurons and the neurogenesis of the UBCs, respectively[64].

1.2.3. Advances in genomic assays

While databases such as CDT-DB and CbGRiTS continue to serve as exceptional resources to provide insights in genetic basis of normal (and abnormal) cerebellar development, they are limited by the drawbacks of the decade-old assay - the microarray[65] – that was used to generate the transcriptome data. While being a powerful tool in high throughput quantitative measurement of gene expression, probe-based microarray is limited by its requirement of large amount of high-integrity RNA, high level of probe-binding artifact, and most importantly, inability to study novel transcripts[66]. These weaknesses of microarray have recently been overcome by innovative next generation sequencing technologies such as pyrosequencing (454[67]), sequencing by synthesis (Illumina[68]), sequencing by ligation (Solid[69]), and single molecule fluorescent sequencing (Helicos[70]). Our objective has been to elucidate novel and important regulators in cerebellar development, as well as to validate previously described regulators discovered by CbGRiTS, with the application of these powerful next generation sequencing technologies.

1.3. Development of the high-throughput technology – HeliScopeCAGE

In 2012, we started a collaborative project with the members of the FANTOM 5 consortium at the RIKEN Omics Inst in Yokohama, Japan. The FANTOM 5 project had a goal of elucidating the transcriptome for every mammalian cell type[71]. To

succeed in this ambitious project, the FANTOM consortium developed a powerful high throughput technology that combined the 5' sequence capture technique (Cap Analysis Gene Expression, or "CAGE") and next generation sequencing technology (Helicos) – HeliScopeCAGE[70]. In this workflow, CAGE is used to capture the biotin-labeled 5' cap of mRNA with streptavidin-coated magnetic beads[72], allowing the measurement of expression level of mRNAs transcripts. Utilization of 5' cap mRNA also allows identification of the promoter that was used to produce the captured mRNAs[73]. Initially, 454 sequencing was adopted to the CAGE procedure to allow for greater throughput than array-based platforms allowed (deepCAGE)[74]. For our project, a more modern technology, Helicos, was used in conjunction with CAGE. Helicos is a platform which images the extension of individual cDNA molecules using a universal primer and fluorescently labeled nucleotides[70]. Fragments of cDNA molecules are first hybridized with universal primers, then, fluorescently-tagged nucleotides are incorporated one at a time. Each nucleotide is also reversibly chemically modified such that further synthesis is blocked until an image has been captured. The specific fluorescent tag of each nucleotide is identified in the image, and the tag is then cut away enzymatically, as well as reversal of the terminating nucleotide modification, allowing the synthesis process to proceed. This process is repeated until the entirety of all fragments have been sequenced[70]. We used the HeliScopeCAGE platform to sequence the first 26bp of all mRNAs from hundreds of murine and human tissues and genetically modified cell lines. Bioinformatic tools and statistical methods such as ANOVA, Helmert (a method that determines first significant change in gene expression[75]) were then

used to analyze the HeliScopeCAGE datasets to identify candidate genes for granule cell development by comparing the gene expression of granule cells at different developmental time periods.

Similar to the microarray platform utilized in CbGRiTS, CAGE data produced by next generation sequencing would offer quantitative expression information for the transcriptome of the developing cerebellum. As a more advanced technology, CAGE also offered unbiased detection of novel transcripts, lack of background and saturate expression, increased specificity and sensitivity and easier detection of low-abundance transcripts. More importantly, joining the FANTOM project allowed us to explore many aspects of cerebellar development that were not possible with CbGRiTS including: 1) The 5' cap sequence captured by CAGE allowed us to explore unique aspects of enhancer and promoter usage regarding to transcriptional regulations. 2) The collaborative work of FANTOM consortium allowed us to compare gene expression across the majority of cell types and tissues which enables classification of ubiquitous and cell-type-specific transcriptional regulators. 3) CAGE's ability to work with small amount of tissues compared with microarray allowed us to submit a separate, laser-captured granule cell time course to investigate granule cell specific transcriptional regulators.

1.3.1. Functional Annotation of Mammalian Genomes (FANTOM5)

In the first phase of the FANTOM5 project, the consortium constructed single molecule CAGE (Cap Analysis Gene Expression) profiles across a collection of 573 human primary cell samples and 128 mouse primary cell samples[71]. This data set

is complemented with profiles of 250 different cancer cell lines (of 154 distinct cancer subtypes), 152 human post-mortem tissues and 271 mouse developmental tissue samples. Due to this extensive tissue coverage, the FANTOM5 project has identified and quantified the activity of at least one promoter for more than 95% of annotated protein-coding genes in the human reference genome - the remaining promoters are probably expressed in rare cell types or during windows of development or states of cellular activation that are not readily accessible and remain to be sampled[71]. The FANTOM5 data highlighted 430 transcription factors which are key regulators to build cell-type-specific regulatory network models. The findings of these TFs and their tissue-specific promoter and enhancer binding sequences greatly extend the data on regulatory genetic elements generated by ENCODE[71].

1.3.2. Generation of mammalian enhancer atlas and discovery of enhancer RNA

While exploring the CAGE-based transcription start site (TSS) atlas across vast number of tissue types, the FANTOM consortium have observed that well-studied enhancers often have CAGE peaks from enhancer regions[76]. These CAGE tags showed a bimodal distribution flanking the central expression peak, with divergent transcription from the enhancer and produce small, uncapped RNAs that are similar to leaky CAGE tags from mRNAs promoters[76].

Furthermore, it is found that enhancer RNAs have many features that distinguish them from mRNAs including (1) sense and anti-sense of enhancer RNAs are equally

sensitive to exosome degradation compared to sense mRNAs that have a longer half-life than their antisense counterpart, (2) enhancer RNAs are short, located to the nucleus, not spliced or polyadenylated unlike mRNAs, , and (3) enhancer RNAs have downstream polyadenylation and 5' splice motif frequencies at the genomic background level while functional mRNAs do not contain pre-mature termination codon and 5' splice sites [76]. The construction of an enhancer atlas would allow one to explore the functional association of genetic regulators and their targets during developmental processes such as cell proliferation and differentiation.

1.3.3. Incorporation of time course data in FANTOM5

As part of the second phase of FANTOM5[62], we generated HeliScopeCAGE libraries from whole mouse cerebella at 12 time points (embryonic days 11-18 at 24 hour intervals, and then every 72hrs until postnatal day 9) in order to identify active transcription factor networks in developing mouse cerebellum (<http://fantom.gsc.riken.jp/>, see Table 1.1 above). Initial analyses indicated that nearly half of all genomic transcripts identified by HeliScopeCAGE were active during cerebellar development, and identified a large set of cerebellar development enriched start sites. We also identified a discrete set of cerebellar development specific start sites active only in this tissue. This study also examined the temporal pattern of start site usage of key developmental genes and cell markers such as Wnt1, Ptf1a, Rora and Pvalb, which validated the approach. Comparison of the HeliScopeCAGE expression pattern of genes over time with previous microarray datasets indicated a high correlation for this subset of key developmental genes –

suggesting that HeliScopeCAGE was performing well (See Chapter 2). Identification of multiple transcription start sites for single genes, and differential temporal usage of promoters, to discover correlations between promoters used, developmental time point, and biological processes during cerebellar development was a subsequent goal undertaken in this dissertation (See Chapter 3).

Including our cerebellar time series, the collaborative FANTOM5 project consists of 19 human and 14 mouse time courses covering a wide range of cell types and biological stimuli[62]. According to the FANTOM consortium, FANTOM5 successfully expands “the set of known human and mouse core promoters from the FANTOM5 body-wide steady-state atlas to 201,802 and 158,966, and the set of transcribed enhancers to 65,423 and 44,459. Of all identified core promoters in human and mouse, 51% and 61% varied significantly in expression in at least one time course. By using a large-scale comparative analysis across many different tissues and time courses, and simultaneously sampling expression at gene promoters and enhancers, it is revealed that enhancer transcription is the most common rapid transcription change occurring when cells initiate a state change” followed by the transcription of protein-coding transcripts, reflecting the binding of transcriptional regulators to the cis-acting enhancer prior to transcription initiation [62].

For our part of the collaboration focusing on cerebellar development, the results of gene expression analysis (Chapter 2), promoter usage analysis (Chapter 3) and motif analysis (Chapter 6) illustrated the utility and power of time course data towards uncovering developmental genes and pathways involved in specific cell

populations of mouse cerebellum. With other resources such as CbGRiTS (Cerebellar Gene Regulation in Time and Space[61]), the FANTOM5 cerebellar development time course CAGE data and its analyses provide a unique resource for functional investigation of gene regulatory networks in mammalian brain development. In addition, the CAGE approach allowed for identification of alternative promoter usage of cerebellar genes, leading to further studies specifically addressing this issue.

1.4. Novel study on the switching of alternative transcriptional start sites usage during cerebellar development

The cerebellum has also been the focus of two extensive genome-wide gene expression studies (CbGRiTS[61] and CDT-DB[60]). However, detailed information on temporally regulated promoter usage of developmentally regulated genes is largely lacking. As noted above, CAGE allows the identification of alternative transcriptional start sites (TSSs) and the corresponding promoters for a single gene[77-80]. The usage of a TSS can be specifically measured by the concentration of CAGE-tagged mRNA produced from that TSS. When two distinct TSSs are used at a single time point, the one with higher expression is considered the “dominant” TSS. Understanding of how the distribution of TSS changes under various developmental stages can shed light on the regulation and function of different gene isoforms. Such analyses would give insight into temporally-specific transcriptional regulatory events (such as epigenetic modification, alternative splicing, and TSSs),

as well as hinting at post-transcriptional regulation (such as transcript localization, translation efficiency, and protein modification).

My hypothesis was that differential promoter usage is important for gene regulation and gene function during cerebellar development. In 2014 Dimont et al. developed a novel method for systematically detecting and characterizing TSS switching events across tissues[81]. Dimont et al. used a bootstrapping technique to estimate the likelihood of a switching event. Here, we analyzed our cerebellar time series to identify novel TSS switching events during cerebellar development. By taking advantage of the FANTOM5 collaboration with our cerebellar developmental time course, we have built a comprehensive TSS switching dataset followed by in silico and biological validation. These TSS switching events were predicted to produce temporally-specific gene transcripts and protein products that may play important regulatory and functional roles during cerebellar development.

1.5. Investigation of granule cell transcriptome with laser capture microdissection

The normal development of granule cells is the result of precise regulation of a set of transcription factors on their downstream targets. Our objectives are to identify TFs that are important for granule cell development, and to understand how these factors function during normal granule cell development. We used laser capture microdissection (LCM), a technique that isolates specific cell types of interest from discrete regions of tissue[82-84], to obtain enriched populations of granule cells from 3 distinct early-stages (E13, E15 and E18) of mouse cerebellar

development. These time points were chosen to study the cerebellar and granule cell transcription for several known important developmental processes during cerebellar granule cell development: E13 to study the early point of granule cell specification and formation of EGL; E15 to study the active stage of granule cell proliferation and tangential migration along the EGL; and E18 to study the granule cell transcriptome at the last day of embryonic development before the granule cell's differentiation and radial migration into the IGL.

By comparing the gene expression between EGL and whole cerebellum samples from E13, E15, and E18, we identified genes that are highly enriched in the EGL cells. Furthermore, with TF binding site analysis (aka, motif analysis), we identified over 100 transcription factors that are potentially key regulators in granule cell development during embryonic stages. Temporal profiles of the transcription factors identified through these methods could reflect important regulatory processes underlying cell specification, differentiation, proliferation or migration events that occur in the life of a granule cell.

1.6. Kruppel-like factor 4 as an example of novel regulator during cerebellar and granule cell development

Among the candidate regulators of granule cell development, we identified Kruppel-like factor 4 (Klf4) from three independent analyses: 1) its motif activity shifts from positive regulation to negative regulation in our whole cerebellum time series; 2) it is up-regulated in the EGL cells in our LCM time series at E13; and 3) it

is a transcriptional regulator of Pax6 discovered by the software Enhancer Element Locator (EEL)[85].

Klf4 belongs to the Kruppel-like factor family, which contains three C-terminal C2H2-type zinc fingers that bind DNA. The name “Kruppel-like” comes from its strong homology with the *Drosophila* gene product Kruppel, an important gene in segmentation of the developing embryo. Klf4 is one of the four genes necessary to create an induced pluripotent stem cell and has been extensively studied for its role in cell proliferation, differentiation and survival in multiple cell types[86], and its association with Pax6 has been documented in corneal development[87, 88]. Although the mechanism of Klf4 in the self-renewal of the stem cells remains unclear, it has been proposed that it may function to maintain cell proliferation[89] or inhibit apoptosis[90]. The function of Klf4 in brain development has been studied through myc-activated overexpression, where cell proliferation and differentiation was observed to be inhibited, along with defects in cilia genesis leading to hydrocephalus[91].

In the cerebellum, Klf4 has been identified as a cancer suppressor gene frequently inactivated in medulloblastoma[92] – a tumor that oftentimes originates from cerebellar granule neurons. However, the role of Klf4 in normal cerebellar development remains unknown. Here we hypothesized that Klf4 may play an important role during cerebellum development.

1.7. Four main chapters, Chapter 2 – 5, and their project aims

Overall, my research in cerebellar development can be divided into two parts. The first part, which consists of Chapter 2 and 3, focused in the transcriptome of the whole cerebellum. Chapter 2 aimed to identify the key transcriptional regulators during cerebellar development through differential gene expression analyses and biological validation. Chapter 3 aimed to study an important aspect of cerebellar transcriptome - the differential usage of alternative promoters during cerebellar development.

The second part of my investigation, consisting of Chapter 4 and 5, focused on the transcriptome of cerebellar granule cells. Chapter 4 aimed to identify granule cell-enriched genes by comparing the granule-cell specific transcriptome with the whole cerebellar transcriptome by using a relatively purified granule cell precursor population obtained through LCM. Chapter 5 aimed to study a knock-out of *Klf4* to define its role in cerebellar granule cell development. This study serves as a detailed validation of potential transcription factors important for cerebellar development and set an example on how future projects can be sprouted from our transcriptome studies.

Chapter 2 : Identification of key regulators for cerebellar development using FANTOM5 time-course CAGE data

2.1. Introduction

Brain development requires intricately controlled expression of specific gene regulatory networks across time. Despite recent developments in genomics technology, large-scale transcriptome analyses across time in neural development is limited. The cerebellum is a less complex, anatomically discrete and well-studied part of the mammalian brain that lends itself to such an analysis. The FANTOM5 Consortium has recently provided a novel examination, *en masse*, of the genomic factors that are in play as cells and tissues differentiate or transition from one state to the next in both human and mouse time course data[1]. HeliScopeCAGE employed by the FANTOM5 Consortium combines the CAGE (Cap Analysis of Gene Expression) protocol and Helicos sequencing to produce direct, high-precision measurement of gene expression based on 5' end sequence of mRNA[2]. Included in this dataset was tissue from the mouse cerebellum that was examined at 24 hour intervals throughout its development beginning at E12 (embryonic day 12) to P0 (postnatal day 0) and at 72 hour intervals from P0 to P9. These developmental time windows span most of the important neurodevelopmental events such as cell specification, emergence from the cell cycle, differentiation, migration and maturation in major neuronal types of cerebellum including the cerebellar granule cells, Purkinje cells, cerebellar interneurons, and cerebellar nuclear neurons.

The limited number of neuronal types that occupy the cerebellum originate

from two separate zones: the rhombic lip and the ventricular neuroepithelium. The rhombic lip (RL) gives rise to the excitatory neurons of the cerebellum that include the glutamatergic cerebellar nuclear neurons, the granule cell precursors, and the unipolar brush cells. The ventricular neuroepithelium gives rise to Purkinje cells and other GABAergic interneurons and cerebellar nuclear neurons[3-5]. The key transcription factors (TFs) Math1 and Pax6 are expressed in RL and the external germinal layer (EGL)[6-8]; and Ptf1a is expressed in the ventricular neuroepithelium[9]. Cerebellar granule cells undergo several developmental stages. They are specified at the rhombic lip around E12, then tangentially migrate along the surface of the cerebellum and establish the EGL. The granule cell progenitors are highly proliferative so that they generate the largest cohort of neurons in the brain[10-12]. In spite of past research on granule cell development [7, 13, 14], the understanding of the genetics behind cerebellar granule cells is being elucidated. By taking advantage of FANTOM5 cerebellar developmental time course analysis, we plan to identify the key regulators in cerebellar development with primary focus on cerebellar granule cells.

We used the mouse time course tissue from the FANTOM5 HeliScopeCAGE dataset, to investigate the key transcriptional regulators of cerebellar development. Comparing expression across time points, we found 4,763 mRNA transcripts temporally differentially expressed during cerebellar development. This set of identified transcripts showed enrichment for genes already known to be functionally associated to the cerebellum. We noted that the majority of the differentially expressed transcripts was observed at early time-points (E11 to E13) and after birth

(P0 to P3). Concordant with this observation is that there is also proportional enrichment of TFs that are differentially expressed during the same developmental time windows (E11 to E13 and P0 to P3). In addition, we have identified 534 TFs with dynamic expression patterns, representing candidate genes that regulate cerebellar development. To experimentally validate our CAGE data and bioinformatics predictions, 6 highly expressed TFs novel to cerebellar development were further studied using real-time PCR experiments and *in utero* knock down experiments. We found that Atf4, Rfx3, and Scrt2 knock-down embryos exhibited severe cerebellar defects, suggesting these TFs as critical regulators of the normal cerebellar development.

This work is part of the FANTOM5 project[62, 71]. Data downloads, genomic tools and co-published manuscripts are summarized at <http://fantom.gsc.riken.jp/5/>.

2.2. Materials & methods

2.2.1. Sample preparation

Mice were housed in a room with 12/12hr light/dark controlled environment. Embryos were obtained from timed pregnant females at midnight of the day when a vaginal plug was detected; this was considered embryonic day 0 (E0). Pregnant females were cervically dislocated and embryos were harvested from the uterus. The cerebellum was isolated from each embryo, pooled with littermates of like genotype, and snap-frozen in liquid nitrogen. 3-4 replicate pools of 3-10 whole cerebella samples were collected from 12 time points across cerebellar development (embryonic days 11-18 at 24 hour intervals and every 72hrs until postnatal day 9).

2.2.2. Sample quality

The standard TRIzol RNA extraction protocol [93] was used for tissue homogenization and RNA extraction. Bioanalyzer analysis was performed to check RNA quality. All RNA samples used for the time series achieved high RNA Integrity (RIN) Score. 34 out of 36 samples had RIN score of 9.7 or higher (10 being the best). The hierarchical clustering analysis (HCA) and principal component analysis (PCA) were performed according to methodology established previously[71].

2.2.3. Differential expression analysis

CAGE peaks, corresponding to experimentally derived transcription start sites

(TSSs), associated to refSeq transcripts (that is at most 50bp away from known refSeq TSSs) have been considered[71]. For each time-point t , we identified differentially expressed refSeq transcripts using the *edgeR* tool[94] when the adjusted p -value is below 0.01 by comparing the levels of CAGE signals at timepoint t with timepoint $t-1$.

2.2.4. Functional enrichment analysis

To investigate the functional association of differentially expressed genes, gene names associated to the refSeq transcripts predicted to be differentially expressed in at least one time-point of the time series have been provided to the enrichR tool[95] for a functional enrichment analysis. We used the enrichR tool through its API using the poster library of Python2.7. A term was considered as enriched if the enrichR adjusted p -value is < 0.01 . Visualization for the enrichment plots were constructed manually using Cytoscape 3.0.1 (Figure 2.4).

2.2.5. Gene selection for experimental validation

Transcription factors are the keys for genetics regulation and are important for cerebellar development as disruption of one TF could affect its downstream targets as well. Thus, TFs with most significant scores from bioinformatics analysis were considered for experimental validation. Genes with significant scores from multiple analyses were prioritized. Several screenings were conducted for potential

candidates: 1) They are expressed in the cerebellum according to ISH databases such as the Allen Brain Atlas. 2) If there is a knock-out (KO) mouse for the gene, the KO must display certain neurodevelopmental phenotypes (i.e. the genes whose KO “showed no apparent phenotypes” were discarded). 3) There were no published studies about the genes in cerebellar development. And lastly 4) The gene must have a max expression of >50tpm at one time during cerebellar development. All-time low expressers were excluded.

2.2.6. Design of shRNA and in utero transfection.

Four sequences of candidate shRNA were selected from different portions of each gene based on an open source algorithm at www.genscript.com/ssl-bin/app/sirna. Designed sequences were chemically synthesized as two complementary DNA oligonucleotides, annealed and ligated to the mU6 vector. Each construct was verified by sequencing. An in utero transfection technique was performed at E12.5 with designed shRNA construct plasmids co-introduced with an EGFP reporter expression construct into the neural epithelium of the IV ventricle using an in utero electroporation method described previously[96]. We used the Vevo 770 High-Resolution In Vivo Micro-Imaging System from VisualSonics for injection. Embryos were removed at 3 and 6 days post-transfection for histological processing and gene expression analysis for child cluster with immunohistochemistry and in situ hybridization analysis. For the three genes with cerebellar developmental defects (Atf4, Rfx3 and Scrt2), we collected 8 Atf4-transfected embryos out of 24 injections, 7 Rfx3-transfected embryos out of 23

injections and 11 Scrt2-transfected embryos out of 24 injections for phenotypical analyses.

2.2.7. qRT-PCR for confirmation of expression profiles

Cerebellar tissue was collected from the embryos and total cellular RNA was collected (TRIzol reagent, Thermo Scientific). Subsequently, cDNA was synthesized using oligo dT primers following the manufacturers protocol (High Capacity cDNA Reverse Transcription Kit; Applied Biosystems). Three biological replicates were analyzed for each of the 6 target genes at three time points (E13, E15, E18) using Applied Biosystems Fast SYBR Green Master Mix reagent and Applied Biosystems 7500 Real-time PCR system. PCR conditions were: 95 °C for 20s, 40 cycles of 95°C for 3s, and 60°C for 30s followed by 95°C for 15s, 60°C for 1min, 95°C for 15s and 60°C for 15s. Amplification of GAPDH and 18s rRNA were used as reference samples to normalize the relative amounts of cDNA between experiments. Expression profiles for each gene were calculated using the average relative quantity of the sample at each of the three time points.

Primer List

| Gene Name | Left Primer | Right Primer |
|-----------|-----------------------|----------------------|
| ATF4 | TCGGCCCAAACCTTATGACC | TGGCTGCTGTCTTGTTTTGC |
| INSM1 | TCCCCTACTCCCATTCCAGG | GGAGTCACAGCGAGAAGACC |
| MXI1 | CAAACCTCTCCTTCGCGTCCT | TTGAGAGCCGGTGTGACTC |
| PCPB1 | ACGGAAAGGAAGTAGGCAGC | CCCTCCGAGATGTTGATCCG |
| RFX3 | CCTGATCCGGCTGCTCTATG | TCGGTGTCTCTCCTGTCACT |
| SCRT2 | ACTCAGACCTCCTTCCCCTC | CCCCTCCGAAACCCTAGAGA |

2.2.8. Immunofluorescence for in utero knockdown validation

Whole cerebellar tissue was isolated from embryos and kept in 4°C 0.1M PBS. The tissue was then embedded in Optimal Cutting Temperature (OCT) compound and sectioned at 16µm. Samples were washed in 0.1M PBS for 2x5 minutes followed by 0.1M PBS-T for 5 minutes. After washing, samples were permeabilized and blocked using a blocking buffer (0.1M PBS, 10% normal goat serum, 0.3% BSA, 0.02% NaN₃). Samples were then incubated with a 1:200 dilution of primary antibodies overnight at 4°C, while control slides were incubated with blocking buffer. Following primary antibody incubation, all samples were washed with 0.1M PBS-T for 3x10 minutes. Samples were then incubated with a 1:200 dilution of secondary antibody and 1:1000 dilution of DAPI for 1 hour followed by

3x10 minute washes of 0.1M PB and 1x10 minute wash of 0.01M PB. Samples were then treated with FluroSave (EMD Millipore) and stored at 4°C. Samples were visualized using an inverted Axio Observer A1 microscope.

The primary antibodies used for immunofluorescence were: anti-SCRT2 (sc-85910, Santa Cruz Biotechnology), anti-ATF4 (sc-7583, Santa Cruz Biotechnology), anti-RFX3 (sc-10662, Santa Cruz Biotechnology), and anti-eGFP (ab290, Abcam Inc). Secondary antibodies used were Goat anti-Chicken Alexa Fluor 488 (A11039, Molecular Probes), Goat anti-Rabbit Alexa Fluor 594 (A31631, Molecular Probes), and Donkey anti-Goat Alexa Fluor 594 (90436, Jackson ImmunoResearch Lab. Inc.).

2.3. Results

2.3.1. Time-course CAGE data collection and quality control

As a part of the FANTOM5 Consortium effort to annotate the regulatory regions involved in gene expression in mammals, we have generated a time-series of whole cerebellum samples consisting of 8 embryonic and 4 postnatal time points from the C57BL/6 mouse that were subjected to CAGE experiments (Figure 2.1). Each time point contained 3 biological replicates that were pooled from as many as 5-20 whole cerebellum tissues. The ages of the embryonic samples were within a narrow window using timed-pregnant mating to minimize developmental noise (see Methods). Bioanalyzer analysis on the harvested total RNA samples revealed RNA integrity numbers (RIN) of 9.6 or higher for all samples except two samples from 4 postnatal time points (Figure 2.2A). The concordance among biological replicates was examined using a sample clustering approach (Figure 2.2B and C). Hierarchical clustering of the CAGE data at all 12 time points during cerebellar development showed a tight grouping of adjacent time points (Figure 2.2B). A principal component analysis revealed an orderly pattern of temporal trajectory from early embryonic to postnatal time points (Figure 2.2C). These results affirm the high quality of the CAGE data derived from our mouse samples.

In total, we have generated 36 HeliScopeCAGE libraries (12 time points with 3 biological replicates per time point) from whole mouse cerebella from E12 to P9. There is a total of 183,903,557 tag reads, with a minimum of 1,648,798 tag reads, a maximum of 9,020,041 tag reads, and a median of 5,455,656 tag reads. These tag

reads represent 25,207 unique transcription start sites (which would produce a unique transcript) or 20,027 unique genes. We used the unique transcripts for our expression analyses as different promoters of a single gene may have different expression patterns leading to products that have distinct function.

Biological significance of data from HCA from the validation component of HCA and PCA: Hierarchical clustering of the CAGE data at all 12 time points during cerebellar development indicated a transcriptional progression over time where the changes in early embryonic time points (E11 and E12) are distinctly separated from the expression profiles during late embryonic (E13-E18, P0) and postnatal (P3-P9) time points (Figure 2.2B and C). A principal component analysis (PCA) highlighted a similar relationship among timepoints, but in addition to the distinctness of E11 and E12, the postnatal times of P3-P9 formed a separate group. Both HCA and PCA highlight the seminal nature of early embryonic time-points for cerebellar development.

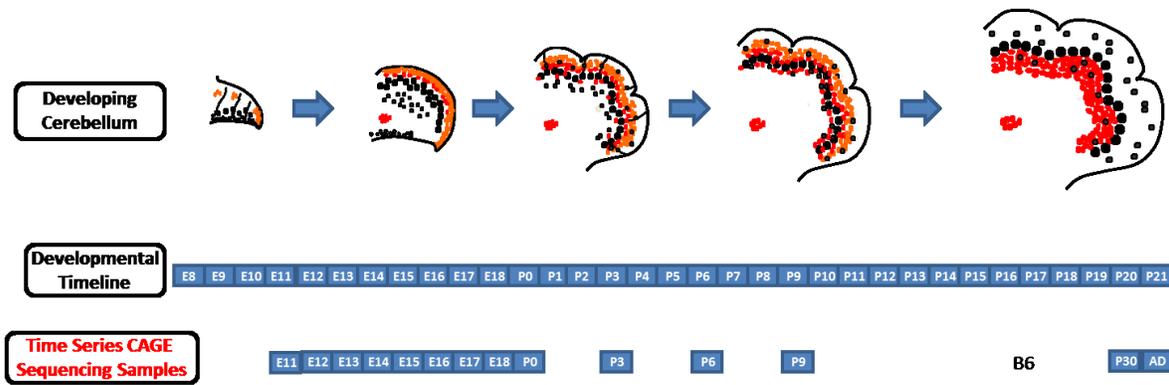


Figure 2.1. Overview of FANTOM5 HeliScopeCAGE transcriptome dataset.

Timeline and schematic diagram of cerebellar development over embryonic and postnatal development and the time points selected for time series microarray data collection. Black dots represent Purkinje cells. Grey dots represent interneurons. Orange cells represent granule cell precursors and red cells represent differentiated granule cells. E# - Embryonic Day #; P# - Postnatal Day #.

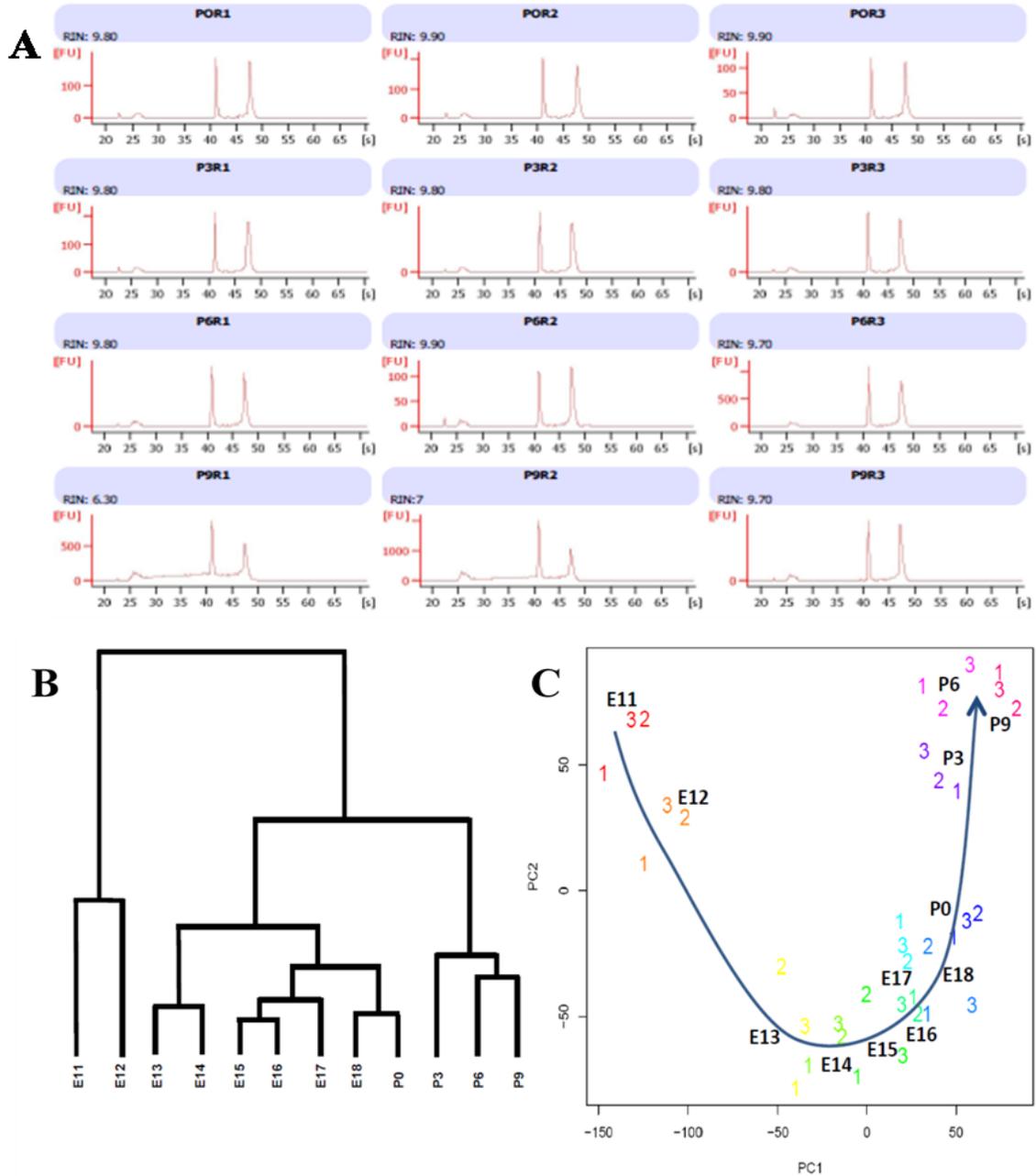


Figure 2.2. Clustering analysis of the cerebellar development time-points.

A) Bioanalyzer graph shows the RNA integrity of our cerebellar RNA samples. The two peaks in the center represent 18S and 28S rRNA respectively and a RIN score (max of 10) is measured by aggregating the electrophoretic trace of the RNA sample. **B)** Hierarchical clustering. Hierarchical clustering of cerebellar development

on time points show three major groupings. The very early embryonic time points (E11 and E12) are the most distant from other late embryonic (E13-E18, P0) and postnatal (P3-P9) time points. **C)** Principal component analysis (PCA). PCA on cerebellar development shows a clear temporal trajectory from early embryonic E11 to postnatal time point, P9.

2.3.2. Differential expression of cerebellar transcripts

We sought to use the dynamic expression of genes over developmental time to identify the genes involved in shaping the cerebellum in mice. The time-course CAGE data for the *in vivo* developmental cerebellar tissue series provides an opportunity to identify regulatory transitions. We hypothesized that significant differential expression of a set of genes may reflect the activity of important transcriptional regulators. Since CAGE data reveals the active transcription start sites / promoters with a quantitative measure of intensity, one can determine the transcripts differentially expressed between successive time-points. We used the *edgeR* package to extract the mRNA transcripts showing a significant differential expression between two successive time-points in the time-series ($p < 0.01$, see Material and Methods) [94].

Figure 2.3 plots the number of transcripts that are predicted to be differentially expressed at each time-point (when compared to the previous one). In aggregate, we predicted 4,763 differentially expressed (DE) mRNA transcripts in at least one time-point. We observed that E12 and E13 harbor the largest number of differentially

transcribed mRNAs when comparing time-points (1,571 and 3,404 DE transcripts, respectively; Figure 2.3). Moreover, we detected a peak in the number of transcripts between P0 and P3 (485 DE mRNAs). These results are in agreement with the hierarchical clustering analysis described above (Figure 2.2B). When looking specifically at transcription factors (TFs), we find that the number of DE TFs along the time-series mimics what we observed when considering all mRNA transcripts (Figure 2.3).

As cross validation, we compared the differentially expressed transcripts (defined as a minimum of 2-fold change in expression level at $p < 0.05$ significance) between our CAGE library and CbGRiTS - our previously established microarray database. We found 1,832 DE transcripts (7.3% out of 25,207 total transcripts) from the CAGE library and 469 DE transcripts (1.0% out of 46630 total transcripts) from CbGRiTS microarray database. The tag-based sequencing technology employed in CAGE might be more sensitive and detected more DE transcripts (7.3% in CAGE vs. 1.0% in microarray) with larger amplitude of change (maximum of 50-fold change in CAGE vs. 12-fold change in microarray). We found 213 DE transcripts (see Supplementary Table 2.1) that are shared between the two datasets, which account for 45.4% of DE microarray transcripts and 11.6% of DE CAGE transcripts, respectively. For all of the 213 shared DE transcripts, the changes in expression are the same direction between the two datasets, with no exception (i.e. there is in no case where a gene showed increase in CAGE data, but decrease in CbGRiTS, or vice versa; see Supplementary Table 2.1). This suggests that the two transcriptome-wide expression dataset is highly coherent.

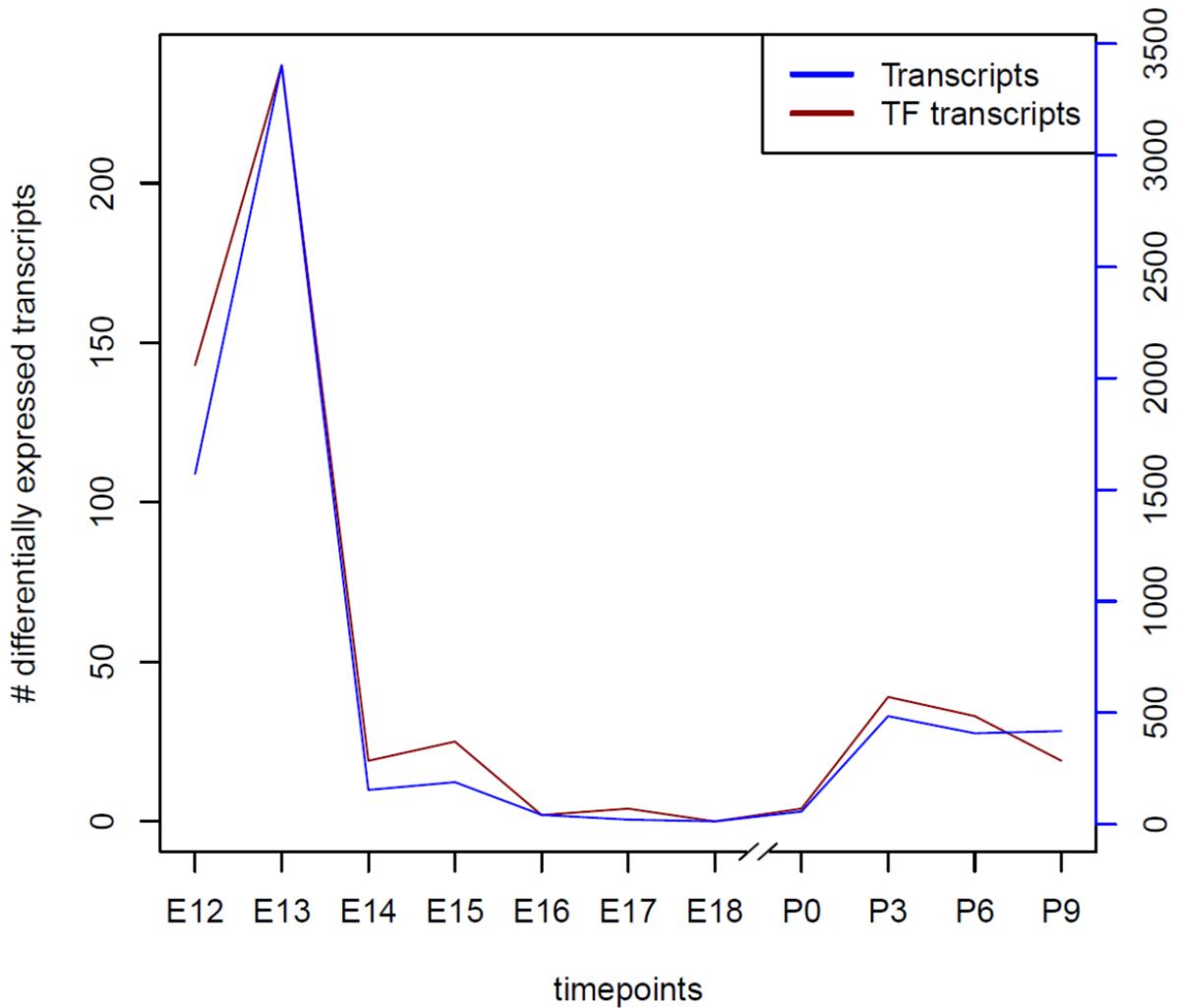


Figure 2.3. Differentially expressed genes along the time-course.

The number differentially expressed mRNA transcripts (blue line, right y-axis) predicted by edgeR[94] at the different time-points of the time series (x-axis). The number of transcripts associated to transcription factors are also provided (red line, left-y-axis).

*see the end of this document for **“Supplementary Table 2.1. 213 Differentially expressed (DE) transcripts that are shared between 1832 DE CAGE tags and 469 DE CbGRiTS probes at E12 – E13.”**

2.3.3. Functional annotation of differentially expressed genes over cerebellar development

We further considered the set of 4,763 differentially expressed mRNA transcripts for a functional enrichment analysis using the enrichR tool[95]. As expected, we detected several terms related to the cerebellum, or more generally the brain, as significantly enriched for DE mRNAs (Figure 2.4). For instance terms associated to neurons, synapses, or nervous system are enriched when considering gene ontology biological processes (Figure 2.4A) and cellular components (Figure 2.4B). The most enriched terms from the mouse gene atlas (MGA) is the 'cerebellum' ($p=1.9 \times 10^{-12}$) and the terms 'cerebral cortex prefrontal' ($p=1.5 \times 10^{-6}$) and 'cerebral cortex' ($p=1.7 \times 10^{-5}$) are also enriched (Figure 2.4C). Furthermore, the enrichment of terms in other brain regions than the cerebellum, that are prefrontal and cerebral cortex, indicates the conserved nature of the molecular determinants of brain development. Indeed, from the set of 3,885 genes associated with the DE mRNAs, 208 are associated to the MGA term 'cerebellum', which is associated with 428 genes. Only 22 of the 208 genes are also in the MGA term 'cerebral cortex' (associated with 349 genes) and only 15 are also in the MGA term 'cerebral cortex prefrontal' (associated with 344 genes). Hence, we are observing DE of mRNAs specific to cerebellum but also specific to other parts of the brain such as cerebral cortex and cerebral cortex prefrontal.

Complementary to the quality control analyses described in the first section of this chapter, it validates the set of DE transcripts derived for CAGE data as important regulators for the development of the cerebellum. These results highlight

2.3.4. Selection of transcription factors as potential master regulators of cerebellar development

We selected candidate transcription factors for further investigation of their roles in cerebellar development based on their degree of expression, differential expression pattern in early time points, and novelty based on literature. First, we ranked all 1,218 TF transcript isoforms based on their maximum expression values over all time points. We went through the top 50 TFs and eliminated the ones with known phenotypes upon gene knockout. For example, well known genes such as Pax6 and Math1 whose roles in cerebellar development are well-established were excluded from the list. Six transcription factors were chosen for further biological validation experiments.

For quantitative validation, we performed real-time PCR on these six genes (Atf4, Insm1, Mxi1, Pcbp1, Rfx3 and Scrt2, see Figure 2.5). We observed a similar expression pattern between our CAGE data (Figure 2.5, left side) and RT-PCR data (Figure 2.5, right side). For example, Mxi1 and Insm1 showed activated expression at E15 and E18 from the cerebellar CAGE data and their qRT-PCR showed the similar pattern; while Scrt2 and Rfx3 showed reduced expression at E15 and E18 from the cerebellar CAGE data and their qRT-PCR reflected this pattern (Figure 2.5). The qRT-PCR results generally corroborate our cerebellar CAGE data, the exception being Atf4.

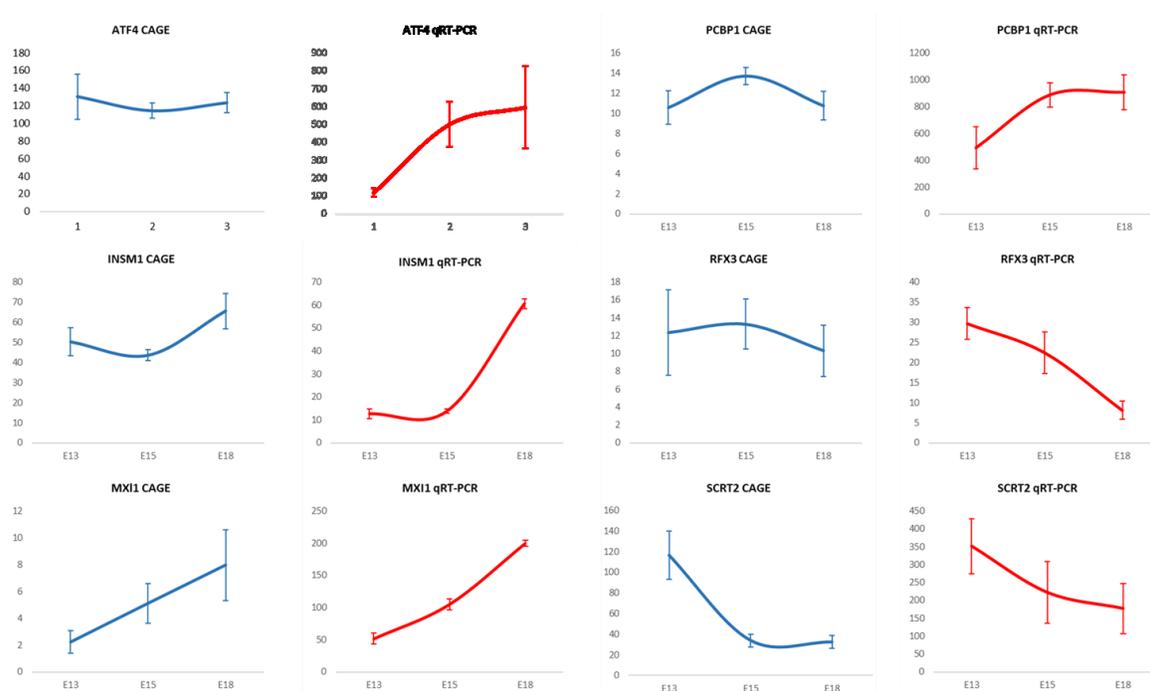


Figure 2.5. Validation of HeliScopeCAGE expression for six differentially expressed transcription factors with qRT-PCR

Quantitative real-time PCR validation for the expression of *Atf4*, *Insm1*, *Mxi1*, *Pcbp1*, *Rfx3* and *Scrt2*. Gene expression is measured – relative quantity (RQ, quantity relative to H₂O negative control) for qRT-PCR and tpm for CAGE genes were shown on y-axis. Expression at E13, E15 and E18 (x-axis) were sampled for prediction and validation groups to representative embryonic developmental stages. Expression pattern is used as indication for the accuracy of CAGE data.

2.3.5. Scrt2, Rfx3 and Atf4 as a master regulator of the cerebellum development

We have made shRNA knockdown constructs for these genes and tested the effect of knockdown on early cerebellar development. We injected shRNA constructs into the IVth ventricle, electroporated with EGFP plasmids at E12 and observed the phenotype at E18 (6 days later). Three of the six candidates, Atf4, Rfx3 and Scrt2, showed morphological perturbations of the developing cerebellum (Figure 2.6). Scrt2 showed the most striking phenotype with ventricular zone and EGL atrophy, lack of foliation at E18.5 and developmental delay (Figure 2.6A-D). With Scrt2 there was virtually no live EGFP positive cells in Scrt2 shRNA transfected samples at E18.5. The only EGFP positivity was found in processes, presumed to be dendrites or axons of degenerated cells; indicative of successful transfection. Atf4 and Rfx3 showed similar phenotype with tissue atrophy in the ventricular zone and developmental delay (Figure 2.6E and F), however, the atrophy was not as extensive as the phenotype of the Scrt2 knockdown. Interestingly, the Rfx3 knockdown retained several EGFP positive cells and its phenotype was the mildest among the three transcription factors that demonstrated a morphological phenotype when knocked down. This suggests the severity of phenotype may correlate with neuronal cell death.

To investigate the fate of the shRNA transfected cells, we harvested the embryos at earlier time points at E15.5. Morphologically there were no noticeable differences between the control and Scrt2 shRNA transfected group at E15.5 (3 days after transfection). There also seems to be no difference in number of

transfected proliferating cells (BrdU+) between the control and SCRT2 shRNA transfected group at E15.5.

Many of the cells stain positive for Calbindin, suggesting that a substantial portion of the transfected cells are Purkinje cells (Figure 2.6G-I). Knockdown of Scrt2 and Atf4 at E12.5 may interfere with the transfected cells' migration potential and the ones trapped in the cerebellum core may undergo apoptosis when they cannot join the Purkinje cell plate and receive necessary neurotrophic factors.

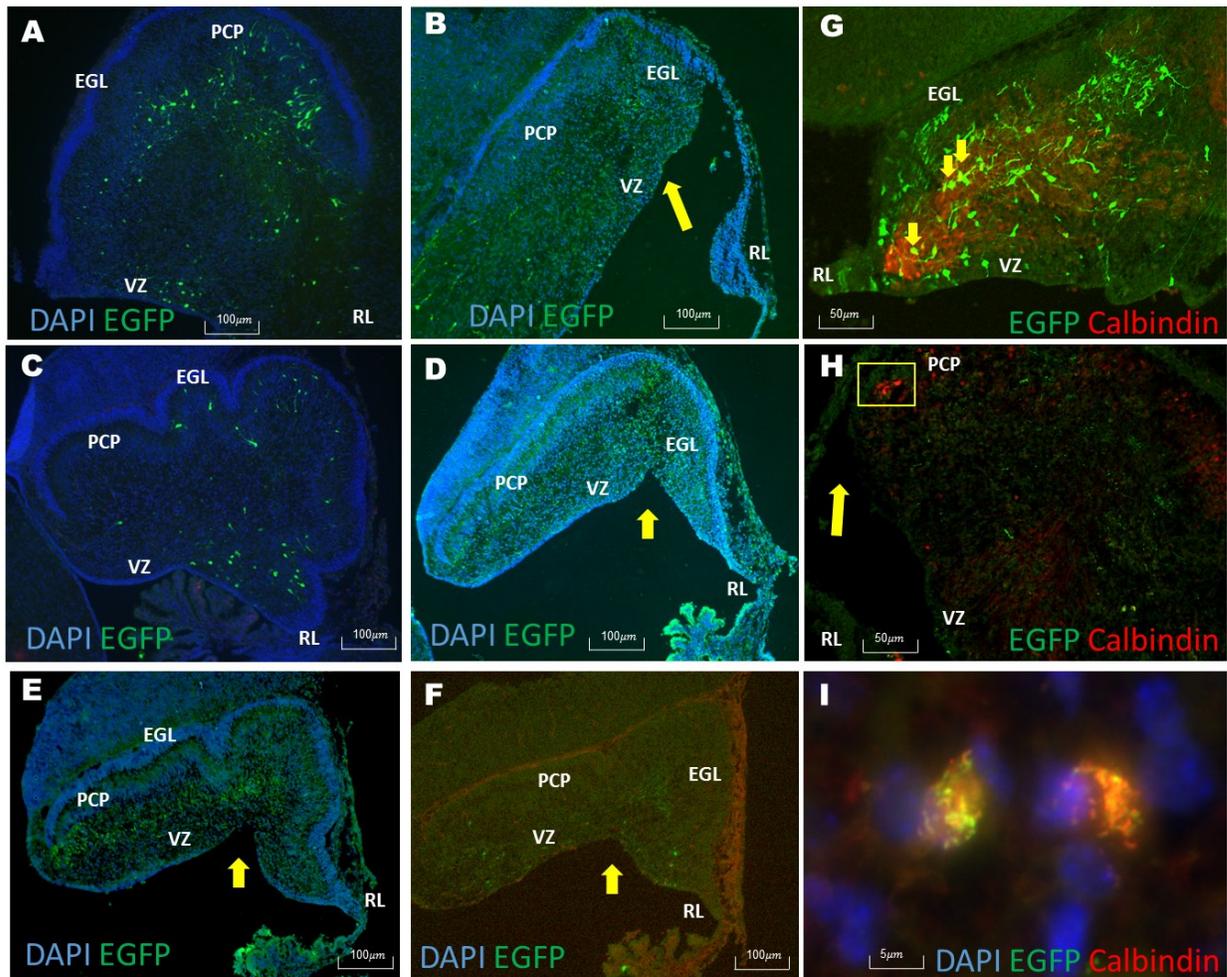


Figure 2.6. Phenotypic changes observed by Sct2 knock-down during cerebellum development.

The knockdown of Sct2, ATF4 and RFX3 showed phenotypic changes during cerebellum development. In utero transfection with shRNA against Sct2, ATF4 and RFX3 was performed at E12.5 embryos and harvested at E15.5 (G) and E18.5 (A-D and H-I). The harvested cerebellar tissue was sectioned and immunofluorescent staining with EGFP and Calbindin antibody was performed. A) Control EGFP transfection – lateral section. B) Sct2 shRNA transfection – lateral section. C)

Control EGFP transfection – medial section. D) Sct2 shRNA transfection – medial section. Note yellow arrow indicates location of developmental abnormality and tissue atrophy. The Sct2 shRNA transfected cerebella seems to exhibit developmental retardation such as lack of EGL foliation at E18.5. E) ATF4 shRNA transfection. F) RFX4 shRNA transfection. In utero transfection against ATF4 and RFX3 shRNA showed similar phenotype of tissue atrophy and developmental delay. However, the phenotype was less pronounced. G-I) Double color immunofluorescent staining with EGFP and Calbindin antibody on Sct2 shRNA transfected cerebellar tissue at E15.5 (G) and E18.5 (H and I). At E15.5 many of the EGFP+ Sct2 shRNA transfected cells are positive for calbindin staining as well (G). However, at E18.5, most of the EGFP+ Sct2 shRNA transfected cells disappeared and only their process (punctate EGFP signals) remained.

2.4. Discussion

2.4.1 Overview of the FANTOM5 cerebellar time course

The availability of large-scale CAGE data produced by the FANTOM5 consortium has been shown to provide an unprecedented opportunity to highlight the precise location of transcription start sites and the transcriptional events driving cellular differentiation and response to stimuli[62, 71]. Here, we generated mouse cerebellar CAGE data from 12 time-points to predict master regulators of cerebellar development. Out of the 33 FANTOM5 time courses (including 19 human and 14 mouse time courses), our cerebellar development time course is one of the two developmental time courses[62]. While the other FANTOM5 developmental time course – the visual cortex – consisted of mutant mouse tissues from 3 postnatal time points[62], our cerebellar time course is unique that it covered normal development of a CNS region from embryonic to postnatal time points. With the cerebellar time course, we first validated the quality of the data through orthogonal experiments and bioinformatics analyses. We hypothesized that TFs might act as key regulators in the development of the cerebellum as highlighted by our differential expression and TF binding motif bioinformatics studies. Finally, we experimentally validated that the TFs Scrt2, Rfx3, and Atf4 are important for cerebellar development since knock-downs introduced phenotypic disturbances.

2.4.2 Time course data reveal the expression landscape of cerebellar development transcriptome

Technologies that provide a genome-wide assessment of gene expression have yielded many insights into development. In particular, microarray technology has been applied to the analysis of gene expression during cerebellar development. Sato et al. have generated the Cerebellar Development Transcriptome Database (CDT-DB) sampling eight time points focused on postnatal cerebellum development (E18, P0, P3, P7, P12, P15, P21 and P56) using the Affymetrix microarray platform[20, 21]. This resource-rich database combined the microarray data with in situ hybridization, GeneChip, and RT-PCR data which were integrated into a web-based knowledge resource, with links to relevant information at various other websites.

A second cerebellar transcriptome database, constructed by Ha *et al.*, is Cerebellar Gene Regulation in Time and Space (CbGRiTS[22]). CbGRiTS contains over 300 Illumina microarrays that are derived from 12 developmental time points primarily focused on embryonic times (the same time points as assessed in the current study) from the C57Bl/6J and DBA/2J lines of mice. In addition, CbGRiTS includes transcriptome data from fourteen recombinant inbred mice constructed between B6 and D2 and three mutant lines of mice whose mutant genes are known to target cerebellar granule cells (Math1, Pax6, and the meander tail mutant[22]). CbGRiTS is an exceptional tool to investigate gene expression patterns over time and it provides several bioinformatic algorithms which allow the generation of genetic regulatory networks in cerebellar development. Furthermore, after the

exploration of a given gene in the CbGRiTS microarray database, one can link-out directly to the gene's corresponding pages in other anatomical databases[22].

While CDT-DB and CbGRiTS were among the largest transcriptome databases focusing on a single brain structure during mammalian brain development, bioinformatics analysis and identification of master gene regulators of brain development have been challenging due to lack of methods linking transcriptional activity with specific promoter use, an ability to compare multiple brain regions in the same platform, and an overall validation of these microarray-based datasets. The FANTOM5 HeliScopeCAGE data, utilizing the next generation sequencing technology, allows resolution of these shortcomings. However, even though the cerebellar developmental time course from FANTOM5 and CbGRiTS were generated utilizing entirely different technologies, In the two datasets (the cerebellum in FANTOM5 and CbGRiTS we observed that the expression patterns of transcripts and, especially, the significantly differentially expressed genes (Supplementary Table 1, at the end of this thesis) are very similar between FANTOM5 HeliScopeCAGE and CbGRiTS microarray datasets. This cross-validation between datasets provides a powerful tool, combined with bioinformatics analyses, to shed light into key transcriptional events and to highlight potentially new key master regulators in cerebellar development.

Furthermore, the hierarchical clustering analysis (HCA), which reveals the overall similarity and differences of expression profiles between samples, of all 12 developmental time points revealed a transcriptional landscape consisting of 3

stages of distinct expression in cerebellar development: an early embryonic stage consisting E12-E13, a late embryonic stage consisting E14-P0 and a postnatal stage consisting P3-P9. The trajectory observed in the PCA plot (Figure 2C) reveals the same transcriptional landscape. More interestingly, HCA of CbGRiTS data also showed a similar pattern with slightly different boundary time points - the late embryonic time stage consisted of E15-E19 in CbGRiTS instead of E14-P0 revealed by this study. We can hypothesize that key events in cerebellar development occur around E13, as well as and around the time of birth.

Indeed, E13 is the time point when cerebellar granule cells are first specified by the transcription factor *Math1* in the rhombic lip and initiate a series of developmental steps from their origins in the rhombic lip around E13 to the trans-migratory path to establish the EGL, to the highly proliferative and then migratory population at the postnatal time points that produce the largest cohort of neurons in the mammalian brain[10-12]. Thus, the current CAGE-based time course dataset provides an excellent resource to investigate the molecular changes that would precede, coincide or proximally follow time-specific cerebellar developmental events. The current dataset also can be used as a valuable reference point to gauge the tissue and cell-type specific expression of transcriptional network of the developing brain.

2.4.3 Novel key regulators in cerebellar development

The FANTOM 5 data led, using filtering steps, to the identification of a manageable set of six candidate genes to explore from a larger set of 200.

Interestingly, three of these six genes (Atf4, Rfx3, and Scrt2), when knocked down, yielded a perturbed phenotype. While all three TFs showed similar phenotypes, Scrt2 showed the most severe one. Compared to EGFP control, Scrt2 knockdown cerebellum showed marked ventricular zone and EGL atrophy, lack of foliation at E18.5 and developmental delay. Furthermore, there were virtually no live Scrt2 shRNA (+) and EGFP (+) cells at E18.5. This suggests that neuronal cell death may be the mechanism of tissue atrophy and developmental delay. Co-staining of transfected embryos with Calbindin, at E15.5 and E18.5, revealed double labeled cells suggesting that the major portion of the Scrt2 shRNA (+) and EGFP (+) cells are in fact Purkinje cells. Premature death of these Scrt2 shRNA (+) and EGFP (+)/Calbindin (+) Purkinje cells could explain marked ventricular zone and EGL atrophy, lack of foliation and developmental delay due to lack or reduced production of critical neurotrophic factors (such as Shh[26] and Bdnf[27]) that are known to be emitted by Purkinje cells. As Atf4 and Rfx3 showed similar phenotype as Scrt2, but in a milder form, these 3 TFs may play a role in the same signaling pathway in neural development.

2.4.4 Concluding Remarks

The current study represents an ideal marriage between state-of-the-art publicly genomics technology data and detailed understandings of brain development toward the identification of master regulators of cerebellar development. The FANTOM5 cerebellum time series CAGE data set is an accessible, high-quality transcriptome database for functional investigation of gene regulatory networks in cerebellar development.

Chapter 3 : Relatively frequent switching of transcription start sites during cerebellar development

3.1. Introduction

Alternative splicing can provide a large reservoir of transcriptional variants from the ~22,000 genes identified by the Human Genome Project[97]. The production of different isoforms due to the usage of alternative transcription start sites (TSSs), which was once considered as uncommon, has now been found in the majority of human genes [98, 99]. Alternative TSSs could be results of a gene duplication event followed by the loss of functional exons in the upstream copy and diversification of the two duplicated promoters. Alternative TSS usage can affect gene expression and generate diversity in a variety of ways. On the transcriptional level, alternative TSS could result in tissue-specific expression, temporally regulated expression, and the amplitude of expression[100]. On the post-transcriptional level, alternative TSS can affect the stability and translational efficiency of the mRNA[101]. Furthermore, alternative TSS can result in protein isoforms with a different amino terminus, which can lead to alterations in protein levels, functions, or subcellular distribution[102]. Therefore, the investigation of temporal switching of TSSs can provide insights into the regulation of different protein isoforms, and presumably their differences in function. One way to optimally identify differential use of isoforms is to examine transcriptional regulation over developmental time.

One high-throughput technique to survey gene expression at the transcriptome level is the Cap Analysis Gene Expression (CAGE) technique which

generates a genome-wide expression profile based on sequences from the 5' end of the mRNA[103]. In the FANTOM project, CAGE has been shown to identify different TSSs and the corresponding promoters for single genes [77-80]. With CAGE data, one can infer the TSS usage through the number of transcripts produced at that particular TSS. When more than one TSS is used at a single time point from a single gene, the TSS with highest expression is considered the “dominant” TSS. The understanding of how the TSS usage changes during development can shed light on how a single gene can function differently over developmental stages through temporally regulated alternative mRNA and protein isoforms.

The complexity of brain development requires intricately controlled expression of specific genes across time. The cerebellum is often used as a model in analyses of brain development due to its limited number of major cell types. These cells are positioned in spatially defined territories of the developing cerebellum. The cerebellum has also been the focus of two extensive genome-wide gene expression profiling of the developing cerebellum [60, 61]. Detailed information on temporally regulated promoter usage of developmentally important genes - which is still largely lacking - can provide valuable information on genome diversity. Moreover, different isoforms of these genes may be translated into distinct protein products that perform different tasks. Such analyses would give insight to the alterations made to the form of the final transcript, localization for transcription factors motif prediction, utilization, and associated regulatory network changes. Thus, in collaboration with the FANTOM5 project[62], we generated a CAGE dataset for the developing cerebellum

with 12 time points to study temporally-regulated gene expression and alternative TSS usage during cerebellar development.

TSS switching events across samples were systematically identified using the Silvapulle F_Q test, a statistical method for constrained hypothesis testing [81]. The F_Q test produces p-values to estimate significance of a switching event. We have applied the F_Q test to our cerebellar time series to identify novel TSS switching events during cerebellar development. Our hypothesis was that differential TSS usage can result in significant regulatory changes that underlie cellular events critical for cerebellar development and morphogenesis. By taking advantage of the FANTOM5 collaboration with our cerebellar developmental time course, we identified 48,489 novel TSS switching events, including 9,767 events in which the dominant TSS shifts over time. These TSS switching events were predicted to produce temporally-specific gene transcripts and protein products that can play important regulatory and functional roles during cerebellar development.

3.2. Materials & methods

3.2.1. Mouse colony maintenance and breeding

This research was performed with ethics approval from the Canadian Council on Animal Care and research conducted in accordance with protocol A12-0190. C57BL/6J mice were used in all experiments and were imported from The Jackson Laboratory (Maine, US) and maintained in our colony as an inbred line. To standardize the time of conception, timed pregnancies were set up. Every weekday at 10:00am, females were coupled with male; at 3:00pm, the females were checked for vaginal plugs and removed from their partners. The appearance of a vaginal plug was recorded as the day of conception (i.e. embryonic day 0) and embryos were collected at 10am on embryonic day 11-18 (E11-E18) every day and postnatal day 0-9 (P0-P9) every 3 days for a total of 12 time points in our cerebellar time series.

3.2.2. Tissue processing

On the day of embryo collection, the mothers were sacrificed and embryos were removed from the uterus in ice-cold RNase-free PBS. Cerebella were dissected from the head of the embryos, then pooled with littermates, and snap-frozen in liquid nitrogen. Three replicate pools of whole cerebella samples were collected at each time point. The standard TRIzol RNA extraction protocol [93] was used for tissue homogenization and RNA extraction.

3.2.3. Quality assessment

A Bioanalyzer (Agilent, Santa Clara, CA) was used to examine RNA quality. All RNA samples used for the time series achieved high RNA Integrity (RIN) scores above 9.0. The samples were sent to RIKEN Omics Center at Yokohama, Japan, as part of Functional Annotation of the Mammalian Genome 5 (FANTOM5) collaboration for CAGE analysis.

3.2.4. Transcriptome library generation by HeliScopeCAGE

CAGE is a technique that generates a genome-wide expression profile based on sequences from the 5' end of the mRNA. With CAGE, the first 27 bp from the 5' end of RNAs were extracted and reverse-transcribed to DNA [4]. The short DNA fragments were then systematically sequenced with the Helicos platform [14]. Each sequenced tag was then mapped to the reference genome to identify the transcription start site (TSS) of the gene from which it was transcribed. "Tag per million" (tpm) was used as a measure of the expression level of RNAs based on concentration – an expression of "10tpm" means that out of each million total transcripts, 10 were transcribed from the TSS in question. Alternative TSSs (illustrated in Figure 3.1a) can be detected when multiple CAGE tags are mapped to the same gene locus in the reference genome. Mapped CAGE tags can be clustered into promoter regions after thresholding to determine bona fide promoter regions in the genome. For this analysis, we use the list of promoter regions published by the FANTOM5 Consortium[71].

3.2.5. TSS switch detection

TSS switching events are detected by comparing the expression of transcripts from two TSSs of a single gene at two time points. The difference in expression level of the two TSSs is designated d_1 and d_2 at time point 1 and time point 2, respectively. The null hypothesis is that there is no switching for the two TSSs ($d_1=d_2$, see Figure 3.1b). A non-crossover TSS event is detected if one TSS is used more frequently at one time point compared to the other, but the same TSS is used dominantly at both time points ($d_1>d_2$, or $d_1<d_2$, both d_1 and d_2 same sign, Figure 3.1b). A crossover TSS switching event is detected if one TSS is used more frequently at one time point compared to the other, and that the dominant TSS switches at the two time points ($d_1>0$ and $d_2<0$ or $d_1<0$ and $d_2>0$, Figure 3.1b). In order to reduce potential confounding of TSS switching events by differential aggregate promoter expression between time points, candidate events were further limited to TSS pairs that do not change in overall mean expression between developmental stages being compared. The null hypothesis tested at this stage is that the mean TSS expression at the two time points is equal, and results were filtered out if the t-test adjusted p-value was < 0.1 .

In addition to the differences in expression (d_1, d_2), the results of TSS switching are represented using the FQ statistic[104] for each gene. The test of the null hypothesis of no differential crossover promoter usage corresponds to a test involving the FQ statistic, which is functionally similar to the ANOVA F-test. Exact p-values for this test are obtained as described in Silvapulle [104]. To our knowledge, the Silvapulle FQ test is the only statistical test available that was specifically

developed for testing hypotheses regarding qualitative interaction, and which we apply in the current study for testing the presence of crossover switching in gene promoter usage.

All P-values are adjusted for multiple comparisons using the Benjamini–Hochberg method to control the false discovery rate. The P-value of the FQ test was used as an indicator of significance for choosing biological validation candidates.

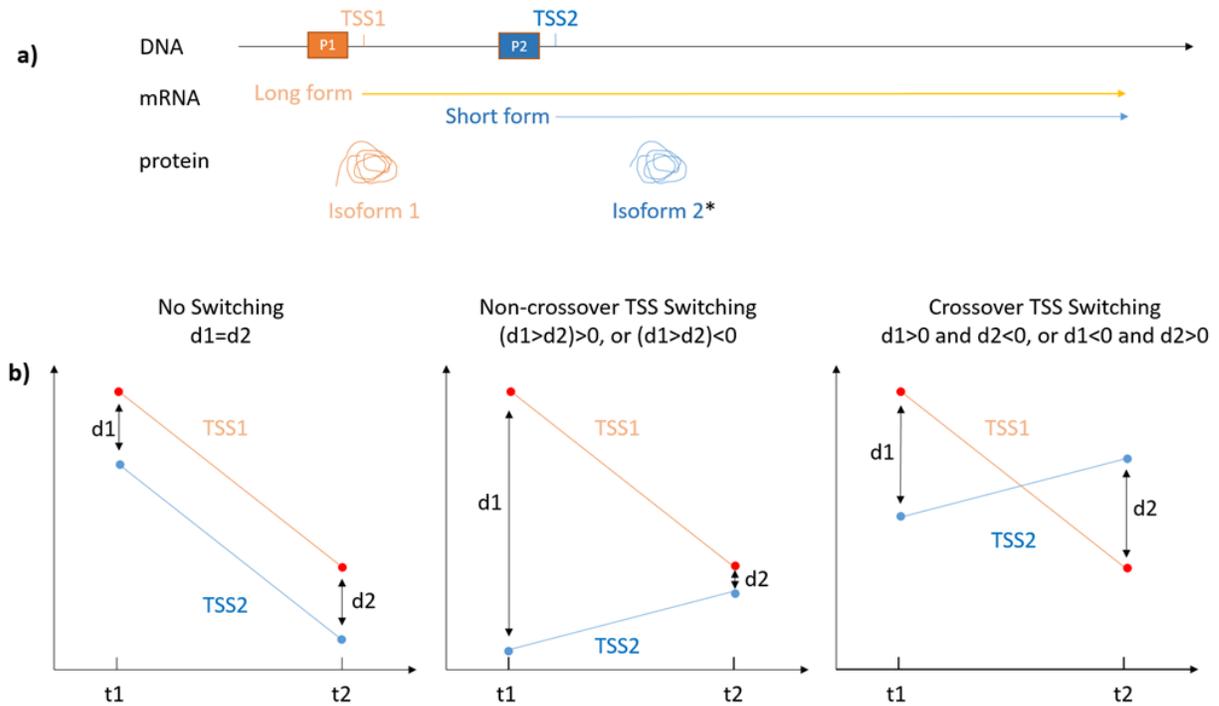


Figure 3.1. A schematic diagram of alternative transcription start sites (TSSs) and the classes of TSS switching.

- a) Alternative TSSs can generate different splicing variants that can be translated into different protein isoforms.

*the functional domains may be affected by alternative TSSs which results in functional diversity.

- b) Different outcomes comparing alternative TSS usage at two time points – no TSS switching, non-crossover TSS switching or crossover TSS switching.

Y-axis represents the quantitative measure of TSS usage measured by the expression level of its mRNA transcript.

X-axis represents the two developmental time points used in the comparison (t1 vs. t2).

3.2.6. Gene Ontology analysis for gene with crossover switching events

To identify cellular processes and molecular pathways in genes with crossover TSS switching events, we used Database for Annotation, Visualization and Integrated Discovery program (DAVID, <https://david.ncifcrf.gov/>[15]) to examine the gene ontology of genes with at least one crossover event with $p < 0.05$ in FQ test. Top 20 GO terms were used for overall analysis in crossover TSS switching genes during cerebellar development. Furthermore, for temporal functional analysis of crossover TSS switching events, top 20 GO terms were generated with DAVID for all events associated with three developmental time points – E13, E15 and P0.

3.2.6. In silico validation of gene expression with established databases and experimental validation with gene structure prediction and quantitative real-time PCR

We used online databases to examine the 20 genes with the lowest p-values. First, we used in situ resources - Genepaint (<http://genepaint.org/>[105]) and Allen Brain Atlas (<http://www.brain-map.org/>[106]) to examine the genes' expression in the cerebellum. Second, we examined the predicted mRNA structures from the two TSSs with the intron/exon database Aceview (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/>[107]) as well as functional domains of their protein products from protein domain database PhosphoSitePlus (<http://www.phosphosite.org/>[108]) to determine the potential effect of TSS switching events on biological function.

Three genes were chosen for further validation for TSS-specific quantitative real-time PCR for the validation of alteration in TSS usage at E12, E15 and P9. Cerebellar RNA was extracted from C57BL/6J mice at E12, E15 and P9 following the same procedure that were used for HeliScopeCAGE RNA collection. cDNAs were produced with random hexamers using the High Capacity cDNA Archive Kit (Applied Biosystems). cDNA products were diluted to 100 ng total RNA input. Sequences of the transcript of interest were loaded into Primer Express® software (Applied Biosystems). For each gene, an isoform-specific forward primer was designed for each of the long and short isoform, while the reverse primer aligns to a common sequence that is shared by both isoforms. Amplicon lengths were between 80 and 120bp. The qPCR was performed with the FAST SYBR Green PCR Master Mix (Applied Biosystems) on an ABI StepOne Plus Sequence Detection System (Applied Biosystems). All runs were normalized to the control gene, Gapdh. Three biological replicates were prepared for each gene target and three technical replicates were performed for each biological replicate. Gene expression was represented as relative quantity against the negative control which used water as the template (noted as “Relative Quantity vs. H2O” in figures). The results of Real-Time PCR were analyzed and graphed by ABI StepOne Plus Sequence Detection System (Applied Biosystems). The expression data were compared with the HeliScopeCAGE data.

3.3. Results

3.3.1. Overview of promoter switch events during cerebellar development

Our cerebellar time series, which consisted of transcriptome data from 12 time points, yielded a total of 183,903,557 CAGE tags that are mapped to 25,207 genes in the reference genome. We identified 48,489 TSS switching events (Figure 3.2a) in the cerebellar time series data that occur in 5,433 genes. These events are comprised of 38,722 non-crossover switching events (Figure 3.2b) that occur in 5,293 genes, and 9,767 crossover switching events (Figure 3.2c) that occur in 1,511 genes. 1,371 out of 1,511 genes (~91%) that have crossover TSS switching events also have at least one non-crossover switching event. This indicates that crossover TSS switching events are rarer and occur in fewer genes when compared to the non-crossover events.

When comparing the cerebellar TSS switching data to nine other tissues in the FANTOM5 dataset (see Table 3.1), our cerebellar development time series has the 3rd highest total number of TSS switching events (48,489) behind “Epithelial to mesenchymal” (132,661 events) and “Adipocyte differentiation” (66,087 events); and is the highest of the three samples derived from ectoderm [“Human iPS to neuron (wt) 1” and “Trachea epithelia differentiation”]. While the cerebellar development time series has less total events than “Epithelial to mesenchymal” and “Adipocyte differentiation” samples, it has a higher frequency of crossover TSS switching events - 20.1% vs 17.6% and 12.5%, respectively. Interestingly, when compared to 48,489

events found in the cerebellum, four out of the five remaining datasets had a higher percentage of crossover events but a much lower number of total switching events.

In conclusion, cerebellar development showed a high frequency in crossover TSS switching among datasets with a high number of total switching events.

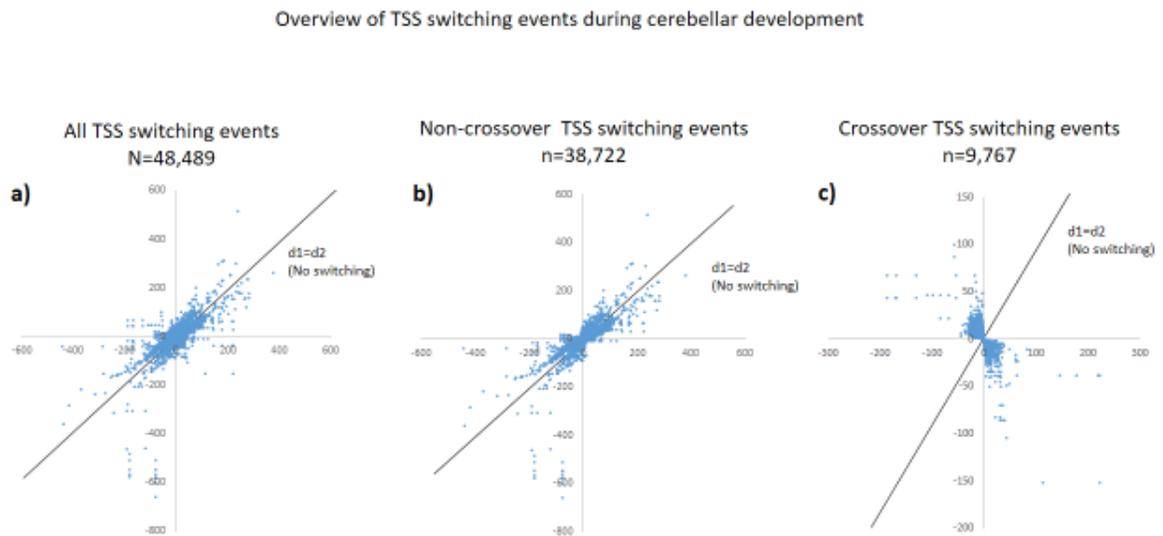


Figure 3.2. Overview of TSS switching events during cerebellar development

a) Overview of 48,489 TSS switching events during cerebellar development.

These events significantly deviate from the no-switching line (indicated by $d1=d2$) ($p<0.05$).

b) Overview of 38,722 non-crossover TSS switching events during cerebellar development

c) Overview of 9,767 crossover TSS switching events during cerebellar development

X-axis represents d1, which is the difference in expression between the two TSSs, measured in tags per million (tpm), at developmental time point 1 (t1), see Figure 1b for a graphic illustration.

Y-axis represents d2, which is the difference in expression between the two TSSs, measured in tag per million (tpm) at developmental time point 2 (t2), see Figure 1b for a graphic illustration.

Table 3.1. Comparison of TSS switching events during cerebellar development with other FANTOM5 datasets

TP# - number of time points in the time series; Switching # - total number of TSS switching events found in the dataset

Gene # - total number of genes with at least one TSS switching event

Column 6: % - TSS switching genes over all 25,207 genes

Non-Xover - total number of non-crossover TSS switching events found in the dataset

Column 8: % - percentage of non-crossover events over all switching events

Xover - total number of crossover TSS switching events found in the dataset

Column 10: % - percentage of crossover events over all switching events

| Time Series | Germ Layer | TP# | Switching# | Gene# | % | Non-Xover | % | Xover | % |
|-----------------------------------|------------|-----|------------|-------|------|-----------|------|-------|------|
| Cerebellar development | Ectoderm | 12 | 48489 | 5433 | 21.6 | 38722 | 79.9 | 9767 | 20.1 |
| Human iPS to neuron (wt) 1 | Ectoderm | 4 | 45069 | 6692 | 26.5 | 41302 | 91.6 | 3767 | 8.4 |
| Trachea epithelia differentiation | Endoderm | 19 | 8389 | 2458 | 9.8 | 6112 | 72.9 | 2277 | 27.1 |
| Adipocyte differentiation | Mesoderm | 16 | 66087 | 5996 | 23.8 | 57857 | 87.5 | 8230 | 12.5 |
| Epithelial to mesenchymal | Mesoderm | 21 | 132661 | 7004 | 27.8 | 109252 | 82.4 | 23409 | 17.6 |
| BMM TB activation IL13 | Mesoderm | 11 | 825 | 527 | 2.1 | 564 | 68.4 | 261 | 31.6 |
| AoSMC response to IL1b | Mesoderm | 10 | 192 | 159 | 0.6 | 129 | 67.2 | 63 | 32.8 |
| Macrophage response to LPS | Mesoderm | 23 | 32234 | 4557 | 18.1 | 22239 | 69 | 9995 | 31 |
| ES to cardiomyocyte | Mesoderm | 13 | 189 | 163 | 0.6 | 100 | 52.9 | 89 | 47.1 |
| Myoblast to myotube | Mesoderm | 9 | 21912 | 4249 | 16.9 | 18735 | 85.5 | 3177 | 14.5 |

3.3.2. Distribution of TSS switching events in cerebellar transcriptome

When we looked at the distribution of the 48,489 TSS switching events over the 5,433 genes, we found a majority of genes with few events and a minority of genes with many events. Thus, we found there are 1,534 (28% of TSS switching gene) genes with one TSS switching event; and only two genes with more than 800 switching events (Figure 3.3a). When we looked at the top 20 genes with the most TSS events (listed in Table 3.2), we observed that these genes account for 13.5% for all TSS switching events, or a total of 6,567 events. From Figure 3.3 (as well as Table 3.2), we can see that there are two outlier genes that have the largest number of TSS switching events for all 3 groups (all TSS, non-crossover and crossover, indicated by arrows in Figure 3.3a-c) - Frmd4a (FERM domain containing 4A) with a total of 852 TSS switching events and Ank3 (ankyrin 3) with a total of 801 TSS switching events (see Table 3.2). These two genes have more than twice the number of TSS switching events than the next closest gene, Abr (active BCR-related gene) with a total of 386 TSS switching events. The numbers of TSS switching events are more evenly distributed across the rest of the 18 genes with a higher frequency of switching (see Table 3.2) as the difference between each rank is less than 10% of the number of events in this group.

When comparing the distribution of crossover and non-crossover events, we found that crossover switching events are clustered in fewer genes when compared with non-crossover events. Since the frequency of non-crossover switching is about four times the number of cross-over (38,722:9,767 or 3.96:1), we would expect roughly a 4:1 ratio for non-crossover : crossover events for any given gene,

assuming an even distribution of both categories. Indeed, we observed roughly a 4:1 ratio for Ablim1 (204 non-crossover events and 50 crossover events) and Dlg2 (223 non-crossover events and 43 crossover events, Table 3.2). However, for the majority of the 20 genes with the greatest number of switching events, the frequency of crossover events is much higher than one fourth of the non-crossover counterpart, such as the two outlier genes mentioned above - Frmd4a (509 non-crossover events vs 343 crossover events) and Ank3 (464 non-crossover events vs 337 crossover events, Table 3.2). This un-even distribution of crossover events is also reflected by the lower abundance of genes with a low number of switching events – 3,052 genes have less than 3 non-crossover events (Figure 3.3b) and only 944 genes have less than 3 crossover events (Figure 3.3c).

In conclusion, we found that crossover events tend to cluster in a fewer number of genes when compared to the non-crossover counterpart.

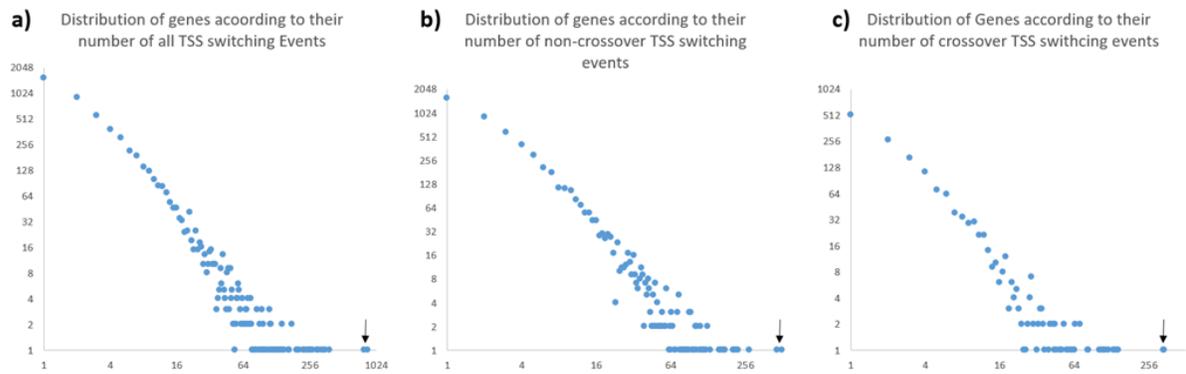


Figure 3.3. Distribution TSS switching events in different genes during cerebellar development

- a) Distribution 48,489 TSS switching events in genes during cerebellar development. Arrow points to the two genes with more than 800 switching events.
- b) Overview of 38,722 non-crossover TSS switching events in 5,293 genes during cerebellar development
- c) Overview of 9,767 crossover TSS switching events in 1,511 genes during cerebellar development

x-axis – number of TSS events occurs within one gene (log2 scaled)

y-axis – number of genes that have the number of TSS events indicated on the

x-axis

Table 3.2. Top 20 genes with highest numbers of TSS switching events

| | <u>Gene ID</u> | <u>All events</u> | <u>Non-crossover events</u> | <u>Crossover events</u> |
|----|----------------|-------------------|-----------------------------|-------------------------|
| 1 | Frmd4a | 852 | 509 | 343 |
| 2 | Ank3 | 801 | 464 | 337 |
| 3 | Abr | 386 | 275 | 111 |
| 4 | Ednrb | 356 | 211 | 145 |
| 5 | lqsec1 | 348 | 206 | 142 |
| 6 | Bcat1 | 329 | 221 | 108 |
| 7 | Pde4d | 308 | 176 | 132 |
| 8 | Ldb1 | 304 | 167 | 137 |
| 9 | Sorbs2 | 297 | 175 | 122 |
| 10 | Cnpy1 | 273 | 158 | 115 |
| 11 | Dlg2 | 266 | 223 | 43 |
| 12 | Ebf1 | 262 | 160 | 102 |
| 13 | Ablim1 | 254 | 204 | 50 |
| 14 | Zeb2 | 246 | 218 | 28 |
| 15 | Trim2 | 233 | 168 | 65 |
| 16 | Celf2 | 227 | 162 | 65 |
| 17 | Map2 | 226 | 170 | 56 |
| 18 | Itgb8 | 208 | 126 | 82 |
| 19 | Ank2 | 197 | 126 | 71 |
| 20 | Ptprg | 194 | 111 | 83 |

3.3.3. Gradual increment in the number of crossover TSS switching events over developmental time

Next, we focused in the temporal distribution of crossover TSS switching. When we look at a day-to-day change in promoter usage (E12 vs E11, E13 vs E12 etc, underlined in Table 3.3), TSS switching occurs evenly across cerebellar development from 13 events to 39 events - with the exception of the E13-E12

comparison (Table 3.3). There are 93 TSS switching events between E12 and E13 indicating a major shift in promoter usage at this developmental stage

To examine the general pattern of TSS switching during cerebellar development, we counted promoter switch events by developmental time points (Table 3.3).

Among the 12 data points in our time course, a total of 66 comparisons between two data points have been carried out to search for the switching of alternative TSSs (Table 3.3). Over the time series, there is a general incremental number of crossover switching events that are detected between two samples that are temporally distant. This most likely reflects the gradual shift of cerebellar transcriptome and TSS usage during development. There are rare exceptions to this pattern, for example, there are more switching events between E11 and E17 samples than found between E11 and E18 samples.

Table 3.3. Distribution of crossover TSS switching events across time in cerebellar development (N=9,767)

Number of crossover TSS switching events that are found in adjacent time points are underlined.

For example, 93 in column 3, row 3 represents 93 crossover events found between E12 and E13

| | E12 | E13 | E14 | E15 | E16 | E17 | E18 | N0 | N3 | N6 | N9 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| E11 | <u>31</u> | 180 | 290 | 340 | 320 | 333 | 236 | 291 | 354 | 279 | 429 |
| E12 | | <u>93</u> | 159 | 200 | 209 | 238 | 162 | 228 | 274 | 225 | 381 |
| E13 | | | <u>21</u> | 59 | 99 | 97 | 76 | 118 | 190 | 180 | 327 |
| E14 | | | | <u>34</u> | 55 | 86 | 69 | 114 | 203 | 198 | 303 |
| E15 | | | | | <u>35</u> | 30 | 29 | 56 | 129 | 143 | 301 |
| E16 | | | | | | <u>29</u> | 23 | 53 | 103 | 113 | 226 |
| E17 | | | | | | | <u>13</u> | 39 | 58 | 91 | 204 |
| E18 | | | | | | | | <u>20</u> | 42 | 76 | 123 |
| N0 | | | | | | | | | <u>39</u> | 60 | 134 |
| N3 | | | | | | | | | | <u>25</u> | 76 |
| N6 | | | | | | | | | | | <u>17</u> |

3.3.4. Gene Ontology analysis for genes with the most significant crossover TSS switching events

To functionally annotate the genes that undergo significant crossover TSS switching, we used the Database for Annotation, Visualization and Integrated Discovery program (DAVID, [https://david.ncifcrf.gov/\[109\]](https://david.ncifcrf.gov/)) to examine the biological process and terms associated with crossover TSS switching genes. From 1,509 genes with 9,767 crossover TSS switching events at $p < 0.05$, we analyzed 20 gene ontology (GO) terms with the lowest p-value from the DAVID analysis (see Figure 3.4a). Terms associated with neuronal development, such as “neuron development”, “neuron projection” and “synapse” also showed up at high significance levels from DAVID analyses (Figure 3.4a).

We have found that the largest alteration in gene expression occurs at E13, E15 and P0 (manuscript in preparation) and were interested to determine the extent that crossover TSS switching plays a role in transcriptome diversity. When comparing crossover events at E13 with all other time points we find 1,440 significant ($p < .05$) events in 584 genes. When comparing crossover events at E15 with other time points we find 1,355 significant ($p < 0.05$) events in 582 genes. Finally, when comparing crossover events at P0 with all other time points we find 1,152 significant ($p < .05$) events in 506 genes. We used these gene lists as input to DAVID and the top 20 terms were selected for these temporal comparisons among the three time points (Figure 3.4b). We found that 7 terms (phosphoprotein, alternative splicing, splice variant, cytoplasm, neuron projection, cytoskeletal protein binding and cytoskeleton) were shared among each of the three time points. These 7 GO

terms were also found among the 8 most significant terms in the analysis with all genes discussed previously. We also observed that comparisons between shorter time spans yield more common GO terms –e.g., there are 5 terms shared between genes with crossover TSS events at E13 and E15, 1 term between E15 and P0 and no terms were common between E13 and P0. Lastly, the majority of GO terms unique to a given time point shared a common theme that may reflect active biological process occurring at the given time – e.g., four out of eight E13 terms were associated with cell motion and cytoskeleton; five out of seven E15 terms were associated with ion binding and six out of twelve P0 terms were associated with regulation of intracellular organization.

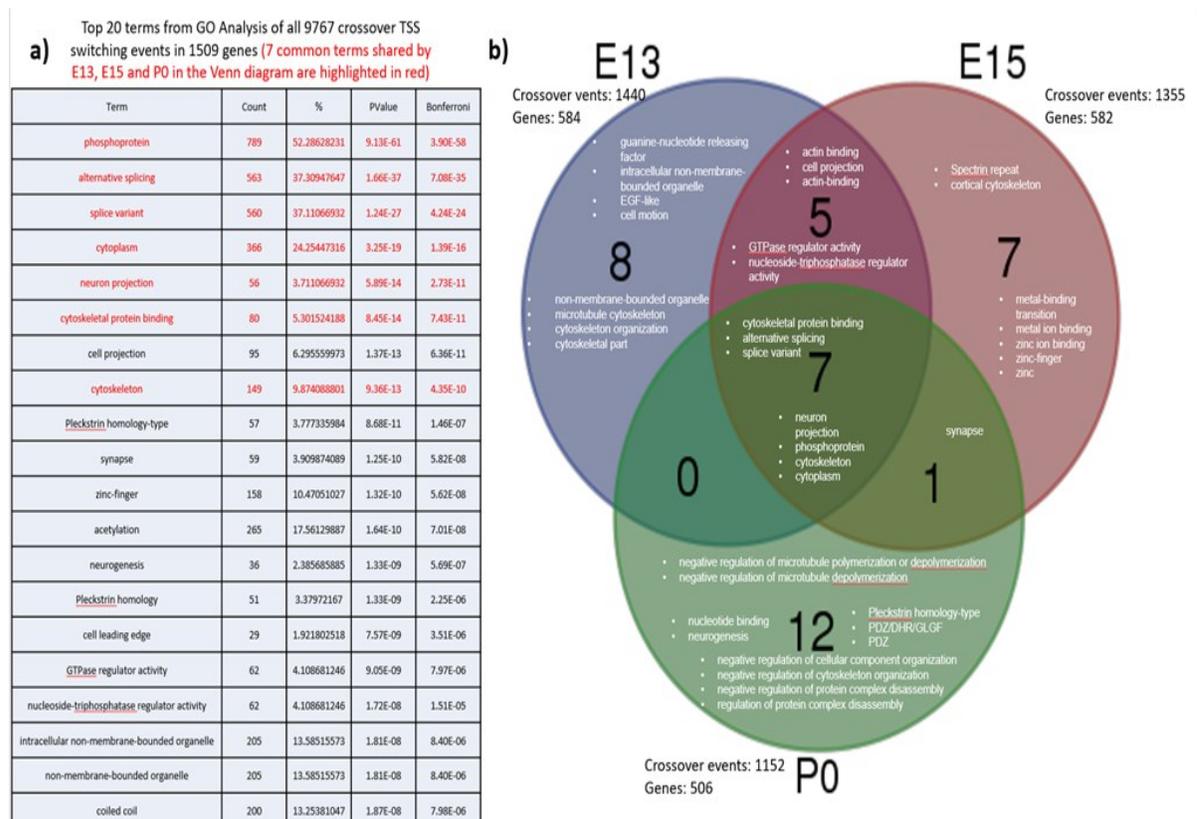


Figure 3.4. GO Analysis for genes significant ($p < 0.05$) for crossover switching at all time points (left) and at three selected time points (right)

- a) Top 20 terms from GO analysis of all 9,767 crossover TSS switching events in 1,509 genes. For column heading: “Term” is the GO term, “Count” is the number of genes associated with the GO term and “%” is the fraction of the number of genes associated with the GO term divided by the total input of 1,509 genes, “PValue” and “Bonferroni” represent the significance of the GO term.
- b) A Venn diagram comparing the top 20 GO terms from crossover TSS switching events between all samples and either E13, E15 or P0 samples.

3.3.5. Validation of promoter switching events

To further investigate the genes with the 20 most significant TSS switching events, we used the *in situ* hybridization expression database Genepaint (<http://www.genepaint.org/>) to examine their expression pattern in the cerebellum (summarized in Table 2.9). Three of these genes showed robust cerebellar expression (Gpc6, Anp32a and Cntnap2) and were chosen to demonstrate the potential biological roles of the TSS switching events during cerebellar development. First, their mRNA structures were obtained from the intron/exon database Aceview (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/>); then their protein structure for each isoform was obtained from protein domain database PhosphoSitePlus (<http://www.phosphosite.org>); finally, the TSS switching events for these three genes were validated with quantitative real-time PCR with promoter-specific primers.

When we investigate the role of the most significant TSS switching events, we found that some of the most significant events do not seem to affect protein sequence and may play roles in transcriptional or post-transcriptional regulation. One example we examined is Glypican-6(Gpc6) - a member of Glypican family that is found on the cell surface and plays important roles in cellular growth control and differentiation. The two TSS sites are 32bp apart in the genome and mRNA that originate from the two TSS sites differ in the first exon in the 5'UTR region (Figure 3.5a). The two forms of mRNA were predicted to be translated into the same protein isoform that contains 565 amino acids. The single glypican domain that makes up the majority of the peptide is not effected by the TSS switching event (Figure 3.5a). Therefore, the usage of alternative TSSs in Gpc6, which is expressed in the NE, NTZ and EGL in the cerebellum (Figure 3.5b), could play a regulatory role, such as temporally regulated expression, amplitude of expression, mRNA stability and mRNA translational efficiency. Our qRT-PCR data confirmed the TSS switching prediction (Figure 3.5c) and showed that it undergoes a non-crossover TSS switching between E15 (TSS2 is the dominant form and has >2 fold usage compared with TSS1) and P9 (TSS2 has slightly higher usage than TSS1, but remains as the dominant form, see Figure 3.5d).

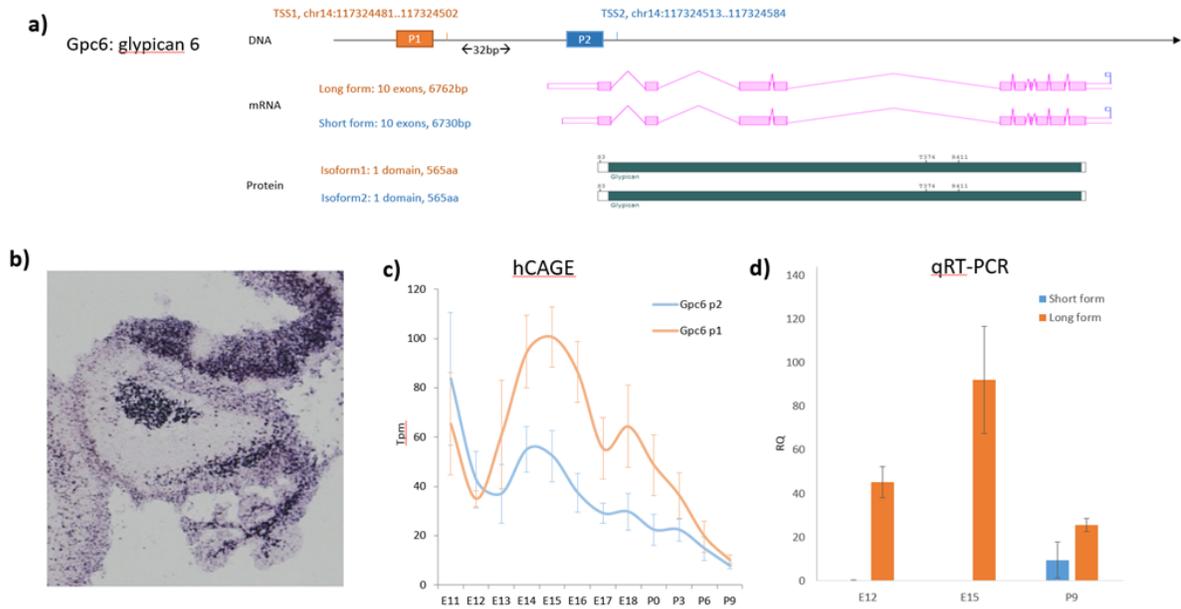


Figure 3.5. Alternative TSSs in glypican 6 (*Gpc6*) and experimental validation of its non-crossover switching events with Real-time PCR

a) Schematic DNA structure of *Gpc6*, alternative mRNA variants and un-altered protein structure

b) *in situ* expression of *Gpc6* in mouse cerebellum at E14.5 (from GenePaint)

c) HeliscopeCAGE expression data for the two alternative TSSs during cerebellar development

X-axis: time, from embryonic day 11 (E11) to postnatal day 9 (P9)

Y-axis: expression level measured in tpm (tags per million)

d) qRT-PCR expression data demonstrating a non-crossover TSS switching event between E15 and P9.

X-axis: time at E12, E15 and P9

Y-axis: expression level measured in RQ (relative quantity against H2O as negative control)

Some of the most significant TSS switching events occur between two TSSs that could produce protein isoforms with different N-termini, which may or may not affect the function of the protein isoforms. An example of this would be Acidic (leucine-rich) nuclear phosphoprotein 32 family member A (Anp32a) - a member of acidic nuclear phosphoprotein 32 kDa (ANP32) family (Figure 3.6). The two TSS sites are 328bp apart in the genome and mRNA that originate from the two TSS sites differs in the first exon in the 5'UTR region as well as the N-terminus of protein products. The first 12 amino acids of the long isoform were absent on the short isoform. Functional domains were not affected by the TSS switching event - both isoforms retained two LRR4 domains and a single NOP14 domain (Figure 3.6a). The difference at the N-terminus can lead to alterations in Anp32a's protein level, subcellular distribution or function in the EGL where it is strongly expressed (Figure 3.6b). As predicted (Figure 3.6c) and validated with our qRT-PCR data, Anp32a undergoes a crossover TSS switching between E12 (TSS9 as dominant form) and P9 (TSS 4 as dominant form, see Figure 3.6d).

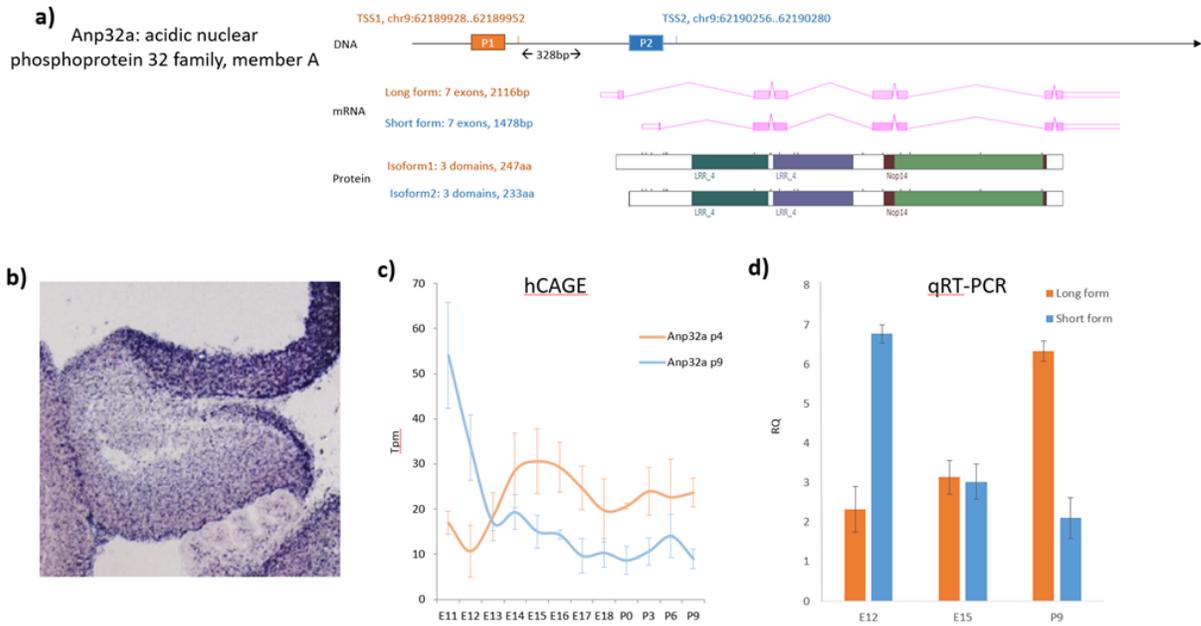


Figure 3.6. Alternative TSSs in acidic nuclear phosphoprotein 32 family, member A (Anp32a) and experimental validation of its crossover switching events with Real-time PCR

a) Schematic DNA structure of Anp32a, alternative mRNA variants and altered protein structure at the N-terminus

b) *in situ* expression of Anp32a in mouse cerebellum at E14.5 (from GenePaint)

c) HeliscopeCAGE expression data for the two alternative TSSs during cerebellar development

X-axis: time, from embryonic day 11 (E11) to postnatal day 9 (P9)

Y-axis: expression level measured in tpm (tags per million)

d) qRT-PCR expression data demonstrating a crossover TSS switching events between E12 and P9.

X-axis: time at E12, E15 and P9

Y-axis: expression level measured in RQ (relative quantity against H2O as negative control)

Lastly, among the genes with the most significant TSS switching events, we have discovered a crossover TSS switching event where protein function is highly affected in the Contactin-associated protein-like 2 (Cntnap2) – a gene encodes a member of the neuroligin family which functions as cell adhesion molecules and receptors in neurons. The two TSS sites are more than 2 million bp apart in the genome. mRNAs that originate from the two TSS sites differ by more than 6000 bp and consist of the first 20 exons of the long mRNA – only 4 exons at the 3' end of the long form mRNA are present in the short form (Figure 3.7a). The Cntnap2 protein, in its long isoform, contains 1400 amino acids and many functional domains including one F5/8 type C domain, two epidermal growth factor repeats domains, four laminin G domains and a TM domain. The short protein isoform of Cntnap2, which has 190 amino acids has only two of the eight functional domains remaining, the last laminin G domain and the TM domain (Figure 3.7a). In the Genepaint database, a probe specific to the long isoform of Cntnap2 was used, and it is indicated that the long isoform is primarily expressed in the rhombic lip of the cerebellum at E14.5 (Figure 3.7b). According to our prediction (Figure 3.7c) and qRT-PCR results, Cntnap2 undergoes a crossover TSS switching between E15 (TSS4 as dominant form) and P9 (TSS3 as dominant form, see Figure 7d). The highly differentiated protein isoforms of Cntnap2 suggest the gene's temporal shift in protein functions during cerebellar development where a truncated form is made specifically in the during early embryonic stages.

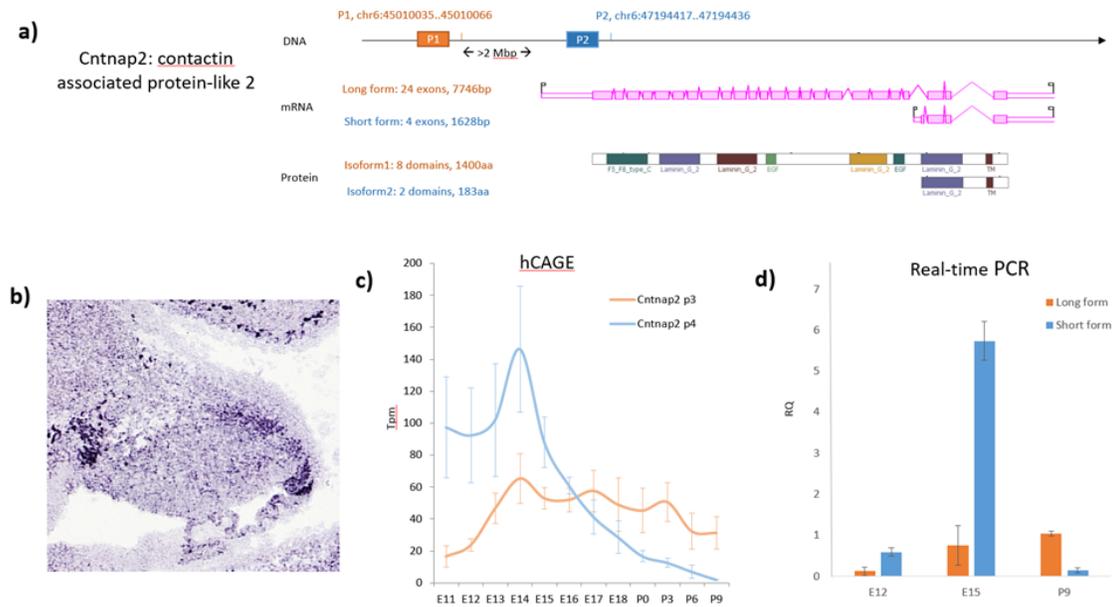


Figure 3.7. Alternative TSSs in contactin associated protein-like 2 (Cntnap2) and experimental validation of its crossover switching events with Real-time PCR

a) Schematic DNA structure of Cntnap2, alternative mRNA variants and truncated protein structure of the short isoform

b) *in situ* expression of Cntnap2 in mouse cerebellum at E14.5 (from GenePaint)

c) HeliscopeCAGE expression data for the two alternative TSSs during cerebellar development

X-axis: time, from embryonic day 11 (E11) to postnatal day 9 (P9)

Y-axis: expression level measured in tpm (tags per million)

d) qRT-PCR expression data demonstrating a crossover TSS switching events between E12 (as well as E15) and P9.

X-axis: time at E12, E15 and P9

Y-axis: expression level measured in RQ (relative quantity against H2O as negative control)

Table 3.4. Cerebellar expression patterns of genes with most significant switching events from the *in situ* database, Genepaint and ABA

N/E – not expressed or ineffective probe

NE – neuroepithelium

RL – Rhombic lip

EGL – external granular layer

NTZ – nuclear transitory zone

N/A – data not available

| Gene | Full Name | Genepaint |
|----------|---|---|
| DLG3 | discs, large homolog 3 | N/E |
| SLC12A5 | solute carrier family 12, member 5 | N/E |
| PDE4D | phosphodiesterase 4D | NE, interior cerebellum |
| IQSEC1 | IQ motif and Sec7 domain 1 | N/E |
| CNTNAP2 | contactin associated protein-like 2 | RL specific |
| CNPY1 | canopy 1 homolog | N/A |
| MAPK8IP1 | mitogen activated protein kinase 8 interacting protein 1 | specific cerebellar nuclei, spinal cord |
| DLGAP4 | discs, large homolog-associated protein 4 | widespread cerebellum |
| ANK3 | ankyrin 3, epithelial | interior cerebellum |
| CACNB4 | calcium channel, voltage-dependent, beta 4 subunit | N/E |
| ANP32a | acidic (leucine-rich) nuclear phosphoprotein 32 family | strong, EGL & NE specific staining |
| TMX3 | thioredoxin-related transmembrane protein 3 | N/A |
| APBB3 | amyloid beta (A4) precursor protein-binding, family B, member 3 | N/E |
| PRMT8 | protein arginine N-methyltransferase 8 | widespread cerebellum |
| EDNRB | Mus musculus endothelin receptor type B | strong NE specific staining |
| SEMA4G | sema domain 4G | widespread cerebellum |
| FBLN5 | fibulin 5 | N/E |
| ZRANB1 | zinc finger, RAN-binding domain containing 1 | N/E |
| ZBTB38 | zinc finger and BTB domain containing 38 | N/A |
| IBTK | inhibitor of Bruton agammaglobulinemia tyrosine kinase | N/E |
| GPC6 | glypican 6 | Strong NE, NTZ specific staining |
| HSPH1 | heat shock 105kDa/110kDa protein 1 | N/A |
| ZFP451 | Mus musculus zinc finger protein 451 | moderate EGL staining |
| GRAMD1B | GRAM domain containing 1B | N/E |

3.4. Discussion

3.4.1. High prevalence of alternative TSSs in mammalian genomes

In this study, we have identified 5,293 genes (~21% of a total of 25207 genes) that exhibit differential TSS usage during cerebellar development. These findings are in line with previous studies and indicate that TSS switching events are common and can play an important role in the diversity of the cerebellar transcriptome during development[110-112]. Furthermore, we have identified 9,767 crossover TSS switching events which suggests an alteration in the dominant TSS over time. Since the alternative mRNA isoforms could be translated into functionally different products, a crossover switching event suggests that one gene can play different roles at different time points in development.

Alternative usage of multiple TSSs of one gene is common in mammalian genomes. It is a key mechanism to increase mRNA and protein diversity since multiple mRNAs from a single gene can encode distinct protein isoforms with different functions (reviewed in [113]). Recent studies suggest that about half of the mouse genes have multiple alternative promoters[102, 114]. For example, alternative promoters have been identified in >20% of genes in ENCODE (<http://genome.ucsc.edu/ENCODE/>) regions[77]. Other genomic studies also found more than a quarter of human genes having multiple active promoters[115-117]. The complex transcriptional regulation of alternative promoter usage has been identified in several genes[113]. Furthermore, in some genes, such as tumor protein p53 (TP53) and guanine nucleotide binding protein (GNAS), alternative promoters

were shown to be activated or silenced[117]. However, the focus of previous studies has been the tissue-specific transcriptional regulation of alternative promoters; the temporal aspect of alternative promoter usage during cerebellar development has been overlooked. Our analyses focused on the switching usage of alternative promoter in the mouse cerebellum, and this is the first systematic study of alternative promoter usage in the development of the mouse cerebellum.

3.4.2. Temporal regulation of alternative TSS associated with developmental processes in the cerebellum

Alternative TSSs reflect different promoter regions that can be used for tissue-specific and/or temporal-specific expression. For example, albumin in hepatocytes has several cis-acting elements that recruit different sets of trans-acting factors, which enable spatial, temporal and dynamics regulation of the transcription of albumin mRNA[118]. In this study, we have identified 9,767 crossover TSS switching events in 1,511 genes. Thus, in ~20% of genes there is more than one promoter that is used dominantly during cerebellar development. Functional annotation analysis for these genes revealed GO terms that are expected to be associated with alternative promoter usage, such as “alternative splicing” and “splicing variants”, as well GO terms that point to processes where promoter switching might play a role during development, such as “phosphoprotein”, “cytoskeleton organization” and “neuron projection”. Phosphoproteins are involved in the post-translational regulatory process phosphorylation, in which a phosphate group is added to a peptide. The physical binding of phosphoproteins, such as Fas-

activated serine/threonine phosphoprotein (FAST), to regulators of alternative splicing has been evidenced by yeast two-hybrid screening and biochemical analyses[119]. Furthermore, the sensory, motor, integrative, and adaptive functions of neuron projections are associated with the development of a growth cone, which is composed primarily of an actin-based cytoskeleton[120]. One of the cytoskeleton remodeling genes, Disabled-1 (Dab1), has multiple isoforms, as a result of alternative splicing[121], that are activated by tyrosine-phosphorylation and play important roles in neuronal positioning by recruiting a wide range of SH2 domain-containing proteins and activates downstream protein cascades through the Reelin signalling pathway[122]. Deficiency in Dab1 pathway resulted in a delay in the development of Purkinje cell dendrites and dysregulation of the synaptic markers of parallel fiber and climbing fiber in the cerebellum[123].

The dominant TSS usually switches gradually over time so that only 3.7% of crossover TSS switching are detected at adjacent time points (357 of 9,767 events). However, more than a quarter of the changes at adjacent time points occur between E12-E13 (93 out of 357). This time period coincides with key developmental events such as cell specification, cell proliferation of granule cell precursors in the rhombic lip, as well as the initiation of cells migrating toward the anterior end of the cerebellum[35].

3.4.3. Alternative TSS as post-transcriptional control during cerebellar development

Alternative TSSs can produce distinct mRNA isoforms that have different RNA stability and translational efficiency of the mRNA isoforms. For example, Vascular Endothelial Growth Factor A (VEGF-A) mRNA stability is regulated through alternative initiation codons that are generated through usage of alternative promoters[124]. The role of Anp32a, which undergo a crossover switching, is not known in normal cerebellar development, but it is found to be involved in a variety of cellular processes in both nucleus and cytoplasm, including signaling, apoptosis, protein degradation, and morphogenesis [125] and it is associated with spinocerebellar ataxia type 1 by interacting with ataxin-1[126]. Moreover, Anp32a is known to be a key component of the inhibitor of acetyltransferase (INHAT) complex in the nucleus, involved in regulating chromatin remodeling or transcription initiation[127]. There are suggestions that Anp32a may play important roles in the brain as the level of Anp32a is increased in Alzheimer's disease and may be involved in the regulatory mechanism of affecting Tau phosphorylation and impairing the microtubule network and neurite outgrowth[128]. We found that two alternative forms of Anp32a are dominantly expressed at different developmental stages in the cerebellum. The long form has 12 additional amino acids on the N-terminus compared to the short form. This difference could alter ANP32A protein stability, distribution and function in the cerebellum.

Alternative TSSs can also be a means of producing mRNA isoforms with various mRNA stability and translation efficiency. Gpc6 is most abundantly expressed in the

ovary, liver, and kidney, with low level expression in the nervous system[129]. In mice, Gpc6 is critical to modulating the response of the growth plate to thyroid hormones[130]; while in human, mutations in the region where Gpc6 resides on Chromosome 13 are associated with defects in endochondral ossification and cause recessive omodysplasia[131]. In the CNS, glypicans (Gpcs) are expressed and secreted by the astrocytes [132]. This includes Gpc6, which is enriched in the cerebellum that regulates the clustering and receptivity of glutamate receptors of the excitatory cerebellar granule cells [132]. We found that the two mRNA isoforms of Gpc6 only differ in mRNA sequence, which undergoes non-crossover switching during cerebellar development, could affect its mRNA stability, translation efficiency, or its secretion by the cerebellar astrocytes.

3.4.4. Functional importance of alternative TSS during cerebellar development

Alternative TSSs can produce protein isoforms with distinct N-termini; this in turn would lead to alterations in protein function. An example would be the secreted and membrane-bound isoforms of mammalian Fos-responsive gene, Fit-1, that are generated and regulated by a pair of alternative promoters[133]. We found that during cerebellar development, the short form of Cntnap2 loses most of the functional domains present in the long form – with only the last laminin G domain retained. Cntnap2 has been found to play a role in the local differentiation of the axon into distinct functional subdomains[134]. The function of Cntnap2 short form during cerebellar development is still to be investigated, but the lack of most functional domains suggests its role as a transcriptional suppressor – through

mechanisms such as non-sense mediated decay[135]; or a functional competitor - the overlapping protein domains antagonizing each other through competing for the same domain binding region[136], for Cntnap2 long form counterpart during early development. During postnatal development, the short form of Cntnap2 ceases to be expressed and the long (and presumably fully functional) form is maintained at a steady level. Cntnap2 is strongly associated with autism spectrum disorders, shown in previous studies[137-139]. A knockout mouse for Cntnap2 targeted the gene's first exon and completely eliminated the expression of the long form[140], which caused abnormalities in body size, neuronal migration and activity, and behaviour. Thus the knockout has been used as an animal model for autism [141, 142]. However, the short form of Cntnap2 should be present in the knockout, and no attention has been directed to the expression of the short form in the knockout. A mutation targeted to the C-terminus would be required to reveal Cntnap2's overall function in considering both its long and short protein isoforms.

3.4.5. Conclusion

We analyzed the cerebellar developmental time course data from the FANTOM5 project and identified 9,767 TSS switching events with temporally specific dominant promoters. This is the first study to investigate the prevalence of alternative TSS usage during cerebellar development and their potential roles in transcriptional, post-transcriptional and functional regulation.

Chapter 4 : Discovery of transcription factors novel to mouse cerebellar granule cell development through laser capture microdissection

4.1. Introduction

The cerebellum is important for motor coordination and cognitive functions [1, 143, 144]. The granule cells are the most abundant neurons in the cerebellum; the mouse cerebellum contains about 100 million granule cells, which make up more than half of total number of neurons in the brain[145]. Granule cells have a unique developmental history. Starting at embryonic day 12 (E12), granule cell precursors, specified by the transcription factor (TF) Atoh1 [38, 43, 146], are generated in the rhombic lip of the cerebellum while the Purkinje cells and other types of cerebellar interneurons are born in the cerebellar neuroepithelium. These cells leave the rhombic lip and migrate along the outer surface of the cerebellum to form a region, almost entirely made up with granule cell precursors, called the external germinal layer (EGL). Throughout embryonic development, the granule cells in the EGL, which can be clearly marked by the TF Pax6[87], undergo extensive proliferation involving various molecular pathways, including the Wnt [147] and Shh [56] pathways. Shortly after their birth, granule cell precursors start to differentiate and express NeuroD1, which is specific for differentiated granular neurons [148, 149]. The differentiating granule cells grow axons that interact with overlying Purkinje cell dendrites, then continue to migrate inward past the Purkinje cell bodies to their final destination in the internal granule layer (IGL). Granule cells continue to differentiate and migrate into the IGL that results in the expansion of the IGL and gradual

disappearance of the EGL. At around post-natal day 21 (P21), the EGL disappears and all granule cells complete their migration into the IGL [150].

The normal development of granule cells is a result of precise regulation of a set of “driver” genes – the TFs and their downstream targets. Our objective was to identify these TFs that are important for granule cell development and to understand how these factors function during normal granule cell development. We used laser capture microdissection (LCM), a technique that can isolate specific cell types of interest from discrete regions of tissue [82, 84], to obtain pure populations of granule cells from 3 distinct early-stages (E13, E15 and E18) of mouse cerebellar development. This approach was used in an attempt to mitigate the expression noise from other cell types that reside in the cerebellum and to focus on studying the granule cells. We used the technique - HeliScopeCAGE that combines next-gen sequencing (Helicos) and Cap Analysis of Gene Expression (CAGE)[71], to generate transcriptome libraries for the developing granule cells. When constructing our time series, E13, 15 and 18 were chosen to study the cerebellar and granule cell transcription for several important developmental processes during cerebellar granule cell development: E13 to study the early point of granule cell specification and formation of EGL; E15 to study the active stage of granule cell proliferation and tangential migration along the EGL; and E18 to study the granule cell transcriptome at the last day of embryonic development before granule cell's differentiation and radial migration into the IGL. With EGL and whole cerebellum samples from E13, E15 and E18, we have identified 1,311 differentially expressed genes in the granule cells (1149 temporal regulated EGL genes and 196 EGL enriched genes, with 34

overlaps in the two sets), including 82 TFs. Furthermore, with TF binding site analysis (aka, motif analysis), we have identified 46 TF candidates that could be key regulators responsible for the variation in the granule cell transcriptome between developmental stages. Altogether, we identified 125 potential TFs (82 from differential expression analysis, 46 from motif analysis with 3 overlaps in the two sets) that may be important for development of the normal cerebellum. Temporal profiles of these TFs that we identify in the present may reflect important regulatory processes underlying cell specification, differentiation, proliferation or migration events that occur in the life of a granule cell.

4.2. Materials & methods

4.2.1. Mouse colony maintenance and breeding

All experimentation with animals was under an approved Canadian Council on Animal Care research protocol (A12-0190). C57BL/6J mice were imported from The Jackson Laboratory (Maine, US) and maintained as an inbred line through brother/sister mating. To standardize the time of conception for embryos, timed pregnancies were established. Each day at 10:00 AM, females were mated with male studs and at 3:00 PM, the females were removed from their partners and checked for the appearance of a vaginal plug. All plugs were recorded as embryonic day 0 (E0); pregnant mothers were sacrificed at 10 AM on embryonic days 13, 15 and E18 and the embryos were extracted from mothers' uterus in ice-cold PBS and subsequently used for whole cerebellum dissection or EGL laser capture microdissection.

4.2.2. Tissue processing and sectioning

The timed embryos were pre-divided into two groups: one to generate whole cerebellar transcriptome data; and the other for laser capture microdissection to generate cerebellar granule cell transcriptome data.

For the whole cerebellar transcriptome group, the cerebellum was dissected out from each embryo in ice-cold RNase-free PBS, pooled with littermates, and snap-frozen in liquid nitrogen. Three replicate pools of 3-10 whole cerebella samples

were collected from E13.5, E15.5 and E18.5. RNA was extracted using the Trizol RNA extraction kit (Invitrogen, Carlsbad, CA, USA).

For the laser capture microdissection series, whole heads of mouse embryos were collected and washed with ice-cold, RNase-free PBS. The heads were then horizontally embedded with Optimal Cutting Temperature (O.C.T.) compound and immediately frozen with liquid N₂ and kept at -80°C until ready for sectioning. Tissue was cryosectioned at 8 µm in a sterile, RNase-free environment. Ten glass slides were prepared to receive three sections per slide with the first slide receiving sections 1, 11, 21; the second slide receiving sections 2, 12, 22; and so on). The glass slides were then kept at -80°C for up to 7 days before they were used for laser capture microdissection. A representative slide from each series was stained with cresyl violet for histological verification of slides with cerebellum and EGL. See Figure 4.1a for a schematic diagram for the tissue processing for both the whole cerebellar and LCM groups.

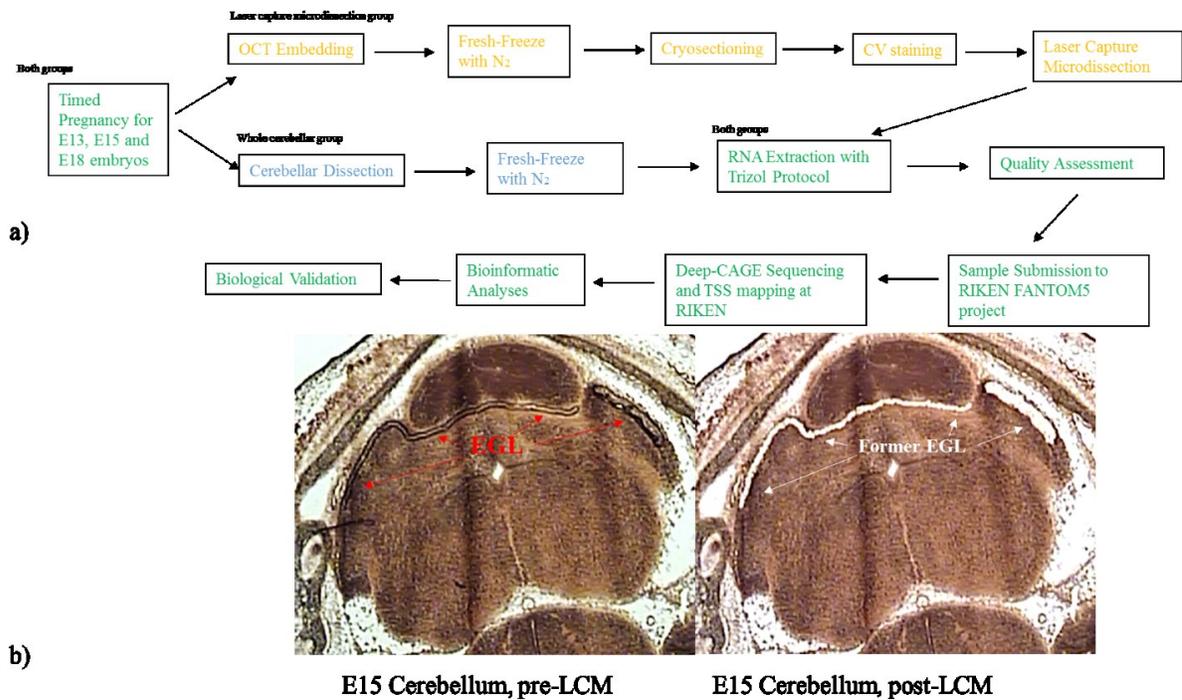


Figure 4.1. Experimental schematic diagram and EGL tissue collection with laser capture microdissection

- a) Schematic diagram of sample processing and analysis of LCM and whole cerebellum samples. Boxes with text in green are experiments shared by both groups; boxes with text in blue are experiments using whole cerebellar tissue; boxes with text in yellow are LCM experiments to isolate EGL tissue.
- b) Example of isolation of EGL using the E15 cerebellum with laser capture microdissection. The image on the left shows the EGL area manually outlined using the LCM microscope (red arrows point to the EGL region). The image on the right shows the same cerebellum with the EGL collected by an adherent LCM cap (white arrows point to the former EGL region).

4.2.3. Laser capture microdissection

To minimize RNA degradation, a rapid cresyl violet staining protocol was adapted and modified from the H&E staining described in [83], and all solutions were made with RNase-free conditions. The slides were first put in 70% EtOH for 10s and then submerged in 0.1% cresyl violet solution (made by dissolving 0.25g cresyl violet powder and 0.04g sodium acetate in 250ml RNase free water) for 60s. Then the slides were dehydrated by immersion in single solutions of 50%, 70% and 95% EtOH (10s each), and three times in solutions of 100% EtOH for 30s each. Finally, tissue was further dehydrated and cleared by processing slides in 3 rounds of Xylenes for 60s each round. The Veritas automated LCM system (Arcturus Veritas) was used to identify the cerebellar granule precursors in the EGL under 20x objective. We were able to identify the granule cells for their tiny size, high nuclear contents and their location in the densely populated EGL that are heavily stained by cresyl violet. The tissue containing granule cells were isolated by an infrared laser beam generated by the instrument and mounted onto a special cell adherent cap (CapSure Macro LCM Caps, Molecular Devices Cat# LCM0211). The captured cells were incubated in the lysis buffer (Ambion #4305895,) for 30 minute in a 42°C water bath and RNA from isolated EGL cells was extracted with Trizol RNA extraction kit (Invitrogen, Carlsbad, CA, USA). Images of pre- and post-LCM E15.5 cerebellum are shown in Figure 4.1b.

4.2.4. Quality assessment

Bioanalyzer analysis was performed to check RNA quality. All RNA samples used for the time series achieved high RNA Integrity (RIN) scores above 9.0; while LCM samples scored from 6.2 to 7.5. All samples were sent to RIKEN Omics Center at Yokohama, Japan, as part of Functional Annotation of the Mammalian Genome 5 (FANTOM5) collaboration.

4.2.5. Transcriptome library generation by HeliScopeCAGE

Cap analysis gene expression (CAGE) is a technique that uses binding of 5' cap of RNA molecules to generate a genome-wide expression profile based on sequences from the 5' end of the mRNA. In CAGE, the first 27 bp from the 5' end of RNAs were extracted and reverse-transcribed to DNA. The short DNA fragments are then systematically sequenced using the Helicos next-gen sequencing platform[151]. Each sequenced tag is then mapped to the reference genome to identify the transcription start site (TSS) of the mRNA. A concentration-based expression level of a transcript is calculated as the number of a transcript's 5' tag sequence in a million tag count. Thus, a transcript with an expression level of 10 tpm indicates that this transcript is present, on average, 10 times in every million transcripts in the tissue that is queried.

4.2.6. Bioinformatics analysis

Our collaborators from the FANTOM Consortium (RIKEN, Japan) designed an analytical approach at whole genome and time-course levels to infer TFs with regulatory roles linked to gene expression changes in the time-course data[151]. Given that CAGE data gives a quantitative measure of the expression of active promoters, we determined mRNAs that are differentially expressed between successive time-points. An analysis of differential expression for each time-point was performed by applying the R package - edgeR to each pair of successive time-points. Student's t-test and false positive discovery rate analysis were performed in R programming packages. For Student's t-test, we considered differential expression with $p < 0.01$ as significant and subsequently used for Gene Ontology analysis at Database for Annotation, Visualization and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>). For false discovery rate analysis, we considered differential expression with $q < 0.05$ as significant. Furthermore, the most significant candidates from the false discovery rate analysis were subsequently selected for *in silico* and experimental validation.

4.2.7. Prediction of motif activity with Motif Activity Response Analysis (MARA)

Motif Activity Response Analysis (MARA) is an informatic tool that models genome-wide expression based on computationally predicted regulatory sites of transcription factors (TFs). In MARA, TF binding motifs are predicted for ~200 TFs (aka, the regulator gene set) in promoter regions using a comparative genomic Bayesian methodology[152]. Gene expression, as measured in tpm, from laser-

dissected EGL CAGE data was used as input. A target gene set, downstream from the regulator gene set, was established by identifying the genes sharing the same TFBSs as predicted above. The linear MARA model is used to explain the expression correlation between the regulator gene and its corresponding target genes. As output, MARA provides the motif activity profiles of all motifs across the samples sorted by a Z-score, which summarizes the significance of the motif in explaining the expression variation across time points (refer to [153] for detailed methods on MARA and Z-score). We used a threshold Z-score > 1.70 to identify active TFs across our samples.

4.2.8. In silico validation of gene expression with three online databases

We examined the expression of highly significant temporally-regulated and enriched genes in the granule cells with established online *in situ* resources: Genepaint (<http://genepaint.org>) and Allen Brain Atlas (<http://www.brain-map.org>). To further validate the expression of TFs that have external granular layer expression, we compared their expression in the Atoh1 mutant cerebellum to wild-type mice in CbGRiTS database (<http://www.cbgrits.org> [61]); the log scale comparison has been transformed into a percent reduction/increase in gene expression in Atoh1 mutant relative to the wild-type. The Atoh1 mutation eliminates granule cell progenitors in the cerebellum and, as such, can be used as a negative control.

4.2.5. Quantitative real-time PCR

A subset of genes that have not previously been reported in the analysis of granule cell development were chosen for quantitative real-time PCR (qRT-PCR) to validate our transcriptome-wide analysis. Cerebella from mice of the same strain (C57BL/6J) were used for HeliScopeCAGE analysis to generate cDNA for qRT-PCR evaluation. Two groups of samples were collected – dissected whole cerebella to validate our cerebellum results, and EGL cells scraped from glass slides to validate our LCM results. All procedures for the scraping were identical to the original LCM experiments, except that a micro-injection needle was used to scrape the EGL cells in place of a laser source. cDNA were produced with random hexamers using the High Capacity cDNA Archive kit (Applied Biosystems). cDNA products were diluted to 100 ng total RNA input. Sequences of the transcript of interest were loaded into Primer Express® software (Applied Biosystems). Amplicon lengths were between 80 and 120 bp. The qPCR was performed with the FAST SYBR Green PCR Master Mix (Applied Biosystems) on an ABI StepOne Plus Sequence Detection System (Applied Biosystems). All runs were normalized to the control gene, Gapdh. Three biological replicates were prepared for each gene target and three technical replicates were performed for each biological replicate. Gene expression was represented as relative quantity against the negative control which used water as the template (noted as “Relative Quantity vs. H2O” in figures).

The results of qRT-PCR were analyzed and graphed by ABI StepOne Plus Sequence Detection System (Applied Biosystems). This expression data were compared with the HeliscopeCAGE data.

4.3. Results

4.3.1. Overview of the dataset

HeliScopeCAGE data were obtained from both the LCM granule cells and the whole cerebella at E13, E15 and E18. For the LCM time series, there are a total of 3,874,054 tag reads (99,939 for E13, 604,536 for E15, 3,169,579 for E18). These tags represent 19,875 unique transcription start sites or 15,482 unique genes, as one gene could have multiple transcription start sites (alternative first exon). For the whole cerebellum time series, there are a total of 50,421,484 tag reads (16,210,864 for E13, 18,441,905 for E15, 15,768,715 for E18). These tags represent 25,207 unique transcription start sites or 20,027 unique genes. We used the unique transcripts (i.e. N=19,875 for EGL and N=25,207 for the whole cerebellum) for all analyses as different promoters of a single gene may have different expression patterns leading to products that have distinct function.

4.3.2. Temporally regulated granule cell transcripts

We performed Students' t-test and false discovery rate analysis on the LCM granule cell dataset among the three different developmental time points (E13 vs. E15, E15 vs. E18 and E13 vs. E18). Using Students' t-test, we found that there are 2,100 unique transcripts that are differentially expressed during granule cell development. These 2,100 unique transcripts arise from 2,511 significant comparisons of differential expression (262 significant comparisons in E13 vs. E15, 430 significant comparisons in E15 vs. E18 and 1,819 significant comparisons in E13 vs. E18, see Table 4.1). There are 411 transcripts that were found to be significantly different in more than one of the comparisons. The false discovery rate

analysis reveals 1,149 unique transcripts that are differentially expressed during granule cell development. These 1,149 unique transcripts arise from 1,173 significant comparisons of differential expression (20 significant comparisons in E13 vs. E15, 19 significant comparisons in E15 vs. E18 and 1,134 significant comparisons in E13 vs. E18, see Table 4.1). There are 24 transcripts that were found to be significantly different in more than one of the comparisons. Out of these temporally regulated granule cell transcripts, 70 are differentially expressed transcripts of genes coding for transcription factors (TFs). These 70 unique TF transcripts arise from 85 significant comparisons of differential expression (11 significant comparisons in E13 vs. E15, 15 significant comparisons in E15 vs. E18 and 60 significant comparisons in E13 vs. E18, see Table 4.1). There are 15 TF transcripts that were found to be significantly different in more than one of the comparisons.

Table 4.1. External germinal layer (EGL) gene transcripts that show temporal regulation.

EGL data is compared between different time points by t-test (p-values) and false discovery rate (q-values) analyses. From the t-test analysis, the total number of significantly altered genes are given as well as the numbers of transcription factors shown in parentheses.

*unique transcripts remove the duplicated gene from multiple comparisons. For example, gene X might be significant for “E13 vs. E15” and “E15 vs E18” comparisons, it would count as two transcripts for “Total” genes in Column 5; but as one “Unique” transcript in Column 6.

| | E13 EGL vs. E15 EGL | E15 EGL vs. E18 EGL | E13 EGL vs. E18 EGL | Total | Unique* |
|--------|---------------------|---------------------|---------------------|-----------|-----------|
| p<0.01 | 262 (11) | 430 (14) | 1819 (60) | 2511 (85) | 2100 (70) |
| q<0.05 | 20 | 19 | 1134 | 1173 | 1149 |

4.3.3. Granule cell enriched transcripts

We compared the LCM time series with the whole cerebellum time series to focus on transcripts that are strongly expressed in the EGL that consists of granule cells. A transcript was considered to be “enriched” in the granule cells if the transcript is expressed at least two times higher in the LCM time series when compared with the whole cerebellum series. Using Students’ t-test, we found that there are 317 transcripts that are significantly enriched in the laser-captured material (largely consisting of EGL cells). These 317 unique transcripts arise from 348 significant expression-comparisons for enrichment (50 significant comparisons at E13, 109 significant comparisons at E15 and 189 significant comparisons at E18, see Table 4.2). There are 31 transcripts that were found to be significantly enriched at more than one time point. The false discovery rate analysis reveals 196 unique transcripts significantly enriched in the laser-captured material. These 196 unique transcripts arise from 208 significant expression-comparisons for enrichment (63 significant comparisons at E13, 132 significant comparisons at E15 and 13 significant comparisons at E18, see Table 4.2). There are 12 transcripts that were found to be significantly enriched at more than one time point. Out of laser-captured, enriched transcripts, 36 code for TFs. These 36 unique TF transcripts arise from 38 significant expression-comparisons for enrichment (3 significant

comparisons at E13, 14 significant comparisons at E15 and 21 significant comparisons at E18, see Table 4.2). There are 2 TF transcripts that were found to be significantly enriched at more than one time point.

Table 4.2. External germinal layer (EGL) cell enriched transcripts (>2x expressions).

EGL data is compared with whole cerebellar data at the same time point by t-test (p-values) and false discovery rate (q-values) analyses. From the t-test analysis, the total number of significantly altered genes are given as well as the numbers of transcription factors shown in parentheses.

| | E13 EGL vs. E13 CB | E15 EGL vs. E15 CB | E18 EGL vs. E18 CB | Total | Unique |
|--------|--------------------|--------------------|--------------------|----------|----------|
| p<0.01 | 50 (3) | 109 (14) | 189 (21) | 348 (38) | 317 (36) |
| q<0.05 | 63 | 132 | 13 | 208 | 196 |

4.3.4. Temporal and spatial confirmation of gene expression using in situ hybridization databases

We found a total of 100 TF-coding transcripts that are significantly differentially expressed (70 for temporal regulation and 36 for granule cell enrichment that includes 6 genes that overlap between temporal regulation and granule cell enrichment). These transcripts arise from 82 unique TFs. Using the Genepaint database (www.genepaint.org), which documents in situ expression data in mouse brain at E14.5, we find expression data for 71 out of the 82 TFs of interest (Table 4.3). We find that 39 out of the 71 TFs are expressed in the cerebellum and 26 out of these 39 are expressed in the granule cell precursors located in the EGL.

Eleven of these 26 appear to be exclusively expressed by the granule cells (Table 4.3). We also analyzed E15.5 and E13.5 expression of our 82 TFs of interest on Allen Brain Atlas (ABA, Table 4.3). When we compared the two expression databases, we found that 17 out of 26 (65%) TFs that showed granule cell expression on Genepaint database also showed EGL and/or RL expression on ABA. To further validate the EGL expression of these TFs, we compared their expression in wild-type mouse with expression in the *Atoh1* mutant which lacks granule cells in the CbGRiTS database (Table 4.4). Five of the 26 genes (*Barhl2*, *Gtf3a*, *Patz1*, *Tgif1* and *Zfp488*) did not have expression data in the database. Of the 21 (out of 26) genes with expression data, 2 genes – *Atoh1* and *Insm1* – had a ~4-fold reduction in gene expression; 6 genes (*Atf7*, *Brca1*, *Hes6*, *Plagl1*, *Tcf4* and *Zic1*) showed a decrease more than 50%; 6 genes (*Bbx*, *E2f1*, *Mafb*, *Neurod1*, *Rfx3* and *Tfdp2*) showed 20%-50% decrease in expression; and the remaining 7 of the 21 genes did not show a substantial difference in expression in the *Atoh1* mutant (Table 4.4).

We validated the gene expression of significant candidates in the granule cell precursors with *in situ* expression data from the Allen Brain Atlas (ABA) database (www.brain-map.org). Although the number of genes sampled across time in the ABA is limited, there were 3 of our genes (*Insm1*, *Irx1* and *Pax3*) with complete E13, E15 and E18 profiles in the ABA. The ABA *in situ* data confirm that these genes have robust expression that is enriched in the EGL when compared with nearby non-neuronal regions (Figure 4.2). This suggests that the LCM procedure for EGL isolation was successful and LCM CAGE data set were highly enriched with

transcripts from granule cell precursors when compared with the whole cerebellar CAGE set.

Table 4.3. External germinal layer (EGL) cell temporally regulated transcription factors (L-L) and EGL enriched transcription factors (L-C)

In silico validation for these genes using in situ database Genepaint (<http://www.genepaint.org>) and Allen Brain Atlas (<http://www.brain-map.org/>), and the number of publications in PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) found for each gene with regard to the cerebellum and the cerebellar granule cells .

Column 3: DE – Time point comparison with Differential Expression. L – EGL expression from LCM, C – Cerebellum expression. 13, 15, 18 is short for E13, E15 and E18 respectively (e.g. L13-L18 means temporal alteration in expression between E13 EGL cells vs. E18 EGL cells; L13-C13 means >2X enrichment in EGL cells at E13 vs. whole cerebellum at E13)

Column 4 & 5: EGL – external germinal layer, NE – neuroepithelium, NTZ – nuclear transitory zone, N/E – not expressed in the cerebellum, N/A – data not available

Column 6: Cb - # of Cerebellum Literature in PubMed

Column 7: GC - # of Granule Cell Literature in PubMed

| Gene | Full Name | Differential expression with p<.01 | GenePaint expression | Allen Brain Atlas expression | Cb | GC |
|--------|---|------------------------------------|----------------------|------------------------------|-----|----|
| ARID3A | AT rich interactive domain 3A | L13-L18 | N/E | N/A | 0 | 0 |
| ATF7 | activating transcription factor 7 | L13-L18 | EGL, widespread | N/A | 0 | 0 |
| ATOH1 | atonal homolog 1 | L13-L18 | EGL | EGL,RL | 102 | 69 |
| BARHL2 | BarH-like 2 | L13-L18 | EGL | RL, EGL | 2 | 1 |
| BBX | bobby sox homolog | L13-L15 | EGL, widespread | RL, NE | 1 | 0 |
| BRCA1 | breast cancer 1 | L13-L18 | EGL, NE | EGL | 5 | 2 |
| CIZ1 | CDKN1A interacting zinc finger protein 1 | L13-L18 | EGL | N/A | 3 | 0 |
| CREB5 | cAMP responsive element binding protein 5 | L13-L18 | N/E | N/A | 0 | 0 |
| E2F1 | E2F transcription factor 1 | L15-C15 | EGL, widespread | EGL | 27 | 22 |
| EBF1 | early B-cell factor 1 | L15-L18 | Interior, widespread | Interior, widespread | 4 | 1 |
| ETS1 | E26 avian leukemia oncogene 1 | L15-C15 | vasculature | vasculature | 2 | 2 |
| FOXC1 | forkhead box C1 | L18-C18 | vasculature | vasculature | 4 | 0 |
| FOXF2 | forkhead box F2 | L13-L18 | N/E | widespread | 0 | 0 |
| FOXQ1 | forkhead box Q1 | L13-L18, L18-C18 | N/E | widespread | 1 | 1 |

| Gene | Full Name | Differential expression with p<.01 | GenePaint expression | Allan Brain Atlas expression | Cb | GC |
|----------|---|------------------------------------|--------------------------|------------------------------|------|------|
| GBX2 | gastrulation brain homeobox 2 | L15-L18, L13-L18 | anterior NE | anterior NE | 38 | 4 |
| GMEB2 | glucocorticoid modulatory element binding protein 2 | L13-L15, L13-L18 | N/E | N/A | 0 | 0 |
| GTF2IRD2 | GTF2I repeat domain containing 2 | L15-L18, L13-L18 | widespread | N/A | 0 | 0 |
| GTF3A | general transcription factor III A | L13-L18 | RL, widespread | RL, NE | 0 | 0 |
| HES1 | hairy and enhancer of split 1 | L13-L18 | N/A | RL | 17 | 9 |
| HES6 | hairy and enhancer of split 6 | L15-L18, L15-C15 | EGL | EGL, NE | 0 | 0 |
| HOPX | HOP homeobox | L15-L18, L13-L18, L18-C18 | N/A | N/E | 0 | 0 |
| IKZF5 | IKAROS family zinc finger 5 | L13-L18 | N/A | N/E | 0 | 0 |
| INSM1 | insulinoma-associated 1 | L15-C15 | EGL | EGL | 3 | 2 |
| IRF3 | interferon regulatory factor 3 | L13-L18 | N/E | N/A | 0 | 4 |
| IRX1 | Iroquois related homeobox 1 | L13-L15 | NE, EGL, NTZ | NE, EGL, NTZ | 1 | 0 |
| JUN | jun proto-oncogene | L13-C13 | NE | NE | 6270 | 4751 |
| KCNIP3 | Kv channel interacting protein 3, calsenilin | L13-L18 | N/A | N/A | 1 | 0 |
| KLF3 | Kruppel-like Factor 3 | L13-C13 | widespread | N/A | 0 | 1 |
| KLF4 | Kruppel-like Factor 4 | L13-L18 | EGL | EGL | 0 | 0 |
| LHX1 | LIM homeobox protein 1 | L13-L18 | NE | Interior, widespread | 10 | 3 |
| LYL1 | lymphoblastic leukemia | L18-C18 | N/E | N/E | 0 | 0 |
| MAFB | v-maf musculoaponeurotic fibrosarcoma oncogene family, member 1 | L13-L18 | EGL, NTZ, widespread | N/E | 1 | 1 |
| MEF2D | myocyte enhancer factor 2D | L15-L18 | N/E | N/E | 9 | 10 |
| MEIS2 | myeloid ecotropic viral integration site-related gene 1 | L15-L18, L13-L18 | NTZ | NTZ | 1 | 0 |
| MLXIP | MLX interacting protein | L13-L18 | N/E | N/E | 0 | 0 |
| MNT | max binding protein | L13-L18 | N/E | widespread | 3 | 6 |
| MSC | Musculin | L13-L18 | N/E | N/E | 47 | 21 |
| NEUROD6 | neurogenic differentiation 6 | L13-L18 | EGL, NE, NTZ | NTZ | 4 | 5 |
| NFATC3 | nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3 | L15-L18, L15-C15 | N/E | N/E | 0 | 1 |
| NR1D1 | nuclear receptor subfamily 1, group D, member 1 | L13-L18 | widespread | N/E | 6 | 2 |
| NR2F2 | nuclear receptor subfamily 2, group F, member 2 | L15-L18 | EGL, NE, widespread | N/E | 2 | 4 |
| ONECUT2 | one cut domain, family member 1 | L13-L18 | NE | N/E | 0 | 0 |
| PATZ1 | POZ (BTB) and AT hook containing zinc finger 1 | L13-L15, L15-C15 | EGL, NE, widespread | N/E | 0 | 0 |
| PAX3 | paired box gene 3 | L13-L15, L13-L18 | NE | NE, EGL, RL | 6 | 2 |
| PBX4 | pre-B-cell leukemia transcription factor 4 | L13-L18 | N/A | N/E | 0 | 0 |
| PER2 | period homolog 2 | L13-L18 | N/A | N/E | 11 | 5 |
| PKNOX1 | Pbx/knotted 1 homeobox | L13-L18 | N/E | NE | 0 | 1 |
| PLAGL1 | pleiomorphic adenoma gene-like 1 | L15-C15, L13-C13 | EGL, NTZ | EGL | 6 | 3 |
| RFX3 | regulatory factor X, 3 | L13-L15, L13-L18 | EGL | N/A | 0 | 0 |
| RUNX1 | runt related transcription factor 1 | L15-C15 | widespread | N/E | 2 | 10 |
| SMAD5 | MAD homolog 5 | L13-L18 | NE, EGL, NTZ, widespread | N/E | 2 | 2 |
| SNAPC4 | small nuclear RNA activating complex, polypeptide 4 | L13-L18 | widespread | N/A | 0 | 0 |
| SOX7 | SRY-box containing gene 7 | L13-L18 | N/E | NTZ | 0 | 1 |
| SP4 | trans-acting transcription factor 4 | L13-L18 | N/E | N/A | 15 | 11 |
| TBX15 | T-box 15 | L13-L18, L18-C18 | EGL | EGL | 0 | 0 |
| TBX3 | T-box 3 | L18-C18 | N/E | EGL, widespread | 1 | 0 |

| Gene | Full Name | Differential expression with p<.01 | GenePaint expression | Allan Brain Atlas expression | Cb | GC |
|---------|--|------------------------------------|-------------------------|------------------------------|----|----|
| TCF4 | transcription factor 4 | L13-L15, L13-L18 | EGL, NE, NTZ | EGL, NE, NTZ | 7 | 4 |
| TCFAP4 | transcription factor AP4 | L13-L18 | EGL | N/E | 0 | 0 |
| TFDP2 | transcription factor Dp 2 | L13-L15, L15-C15 | EGL, NE | EGL, NE | 0 | 0 |
| TGIF1 | TG interacting factor | L15-C15 | EGL | EGL | 0 | 0 |
| THRB | thyroid hormone receptor beta | L13-L18 | N/E | N/E | 1 | 1 |
| YY1 | YY1 transcription factor | L13-L15, L13-L18 | N/E | EGL | 7 | 3 |
| ZBED4 | zinc finger, BED domain containing 4 | L13-L18 | N/E | N/A | 0 | 0 |
| ZBTB3 | zinc finger and BTB domain containing 3 | L15-L18, L13-L18 | N/A | N/A | 0 | 0 |
| ZFP148 | zinc finger protein 148 | L13-L18 | N/A | N/A | 0 | 0 |
| ZFP187 | zinc finger protein 187 | L13-L18 | anterior NE, widespread | N/A | 0 | 0 |
| ZFP239 | zinc finger protein 239 (Zfp239), transcript variant 1 | L13-L18 | N/A | N/A | 0 | 0 |
| ZFP354A | zinc finger protein 354A | L13-L18 | N/A | N/A | 0 | 0 |
| ZFP488 | zinc finger protein 488 | L13-L18 | EGL | EGL, NTZ | 0 | 0 |
| ZFX | zinc finger protein X-linked | L13-L18 | N/E | N/A | 0 | 0 |
| ZIC1 | zinc finger protein of the cerebellum 1 | L13-L15 | NE, EGL, NTZ | NE, EGL, RL, NTZ | 37 | 18 |

Table 4.4. Expression of EGL expressed TFs in wild-type mice compared with expression in the Atoh1 KO at CbGRiTS microarray database (<http://cbgrits.org>).

Expression level in Column 4 and 5 are shown after 2Z+7 normalization.

Column 3: EGL – external germinal layer, NE – neuroepithelium, NTZ – nuclear transitory zone, N/E – not expressed in the cerebellum, N/A – data not available at Genepaint

Column 4: Wild-type expression, log2 normalized microarray data

Column 5: Atoh1 KO (a knockout mouse strain without granule cells) expression, log2 normalized microarray data

Column 6: Fold change is calculated by comparing normalized microarray expression:

$$[1-2^{(\text{Normal expression} - \text{Atoh1 KO expression})}] \times 100\%$$

For example, fold change for Atoh1: $[1-2^{(9.65-7.27)}] \times 100\% = -421\%$. So it is a decrease of 421% in expression in the Atoh1 KO when compared with wild type.

| Gene | Name | Genepaint | WT | Atoh1 | Fold |
|--------|---|---------------------|------|-------|-------|
| ATF7 | activating transcription factor 7 | EGL, widespread | 5.05 | 4.23 | -77% |
| ATOH1 | atonal homolog 1 | EGL | 9.65 | 7.27 | -421% |
| BARHL2 | BarH-like 2 | EGL | N/A | N/A | N/A |
| BBX | bobby sox homolog | EGL, widespread | 7.89 | 7.63 | -20% |
| BRCA1 | breast cancer 1 | EGL, NE | 6.67 | 5.85 | -77% |
| CIZ1 | CDKN1A interacting zinc finger protein 1 | EGL | 9.17 | 9.27 | 7% |
| E2F1 | E2F transcription factor 1 | EGL, widespread | 6.6 | 6.09 | -42% |
| GTF3A | general transcription factor III A | RL, widespread | N/A | N/A | N/A |
| HES6 | hairy and enhancer of split 6 | EGL | 10.5 | 9.69 | -77% |
| INSM1 | insulinoma-associated 1 | EGL | 10 | 7.72 | -386% |
| IRX1 | Iroquois related homeobox 1 | NE, EGL, NTZ | 8.68 | 8.7 | 1% |
| KLF4 | Kruppel-like Factor 4 | EGL | 7.58 | 7.42 | -12% |
| MAFB | v-maf musculoaponeurotic fibrosarcoma oncogene | EGL, NTZ, | 2.48 | 2.11 | -29% |
| NEURO | neurogenic differentiation 6 | EGL, NE, NTZ | 8.54 | 8.28 | -20% |
| NR2F2 | nuclear receptor subfamily 2, group F, member 2 | EGL, NE, widespread | 7.33 | 7.5 | 11% |
| PATZ1 | POZ (BTB) and AT hook containing zinc finger 1 | EGL, NE, widespread | N/A | N/A | N/A |
| PLAGL1 | pleiomorphic adenoma gene-like 1 | EGL, NTZ | 7.58 | 6.86 | -65% |
| RFX3 | regulatory factor X, 3 | EGL | 9.98 | 9.5 | -39% |
| SMAD5 | MAD homolog 5 | NE, EGL, NTZ, | 7.72 | 7.57 | -11% |
| TBX15 | T-box 15 | EGL | 8.17 | 8.01 | -12% |
| TCF4 | transcription factor 4 | EGL, NE, NTZ | 9.94 | 9.33 | -53% |
| TCFAP4 | transcription factor AP4 | EGL | 9.17 | 9.14 | -2% |
| TFDP2 | transcription factor Dp 2 | EGL, NE | 11.6 | 11.28 | -25% |
| TGIF1 | TG interacting factor | EGL | N/A | N/A | N/A |
| ZFP488 | zinc finger protein 488 | EGL | N/A | N/A | N/A |
| ZIC1 | zinc finger protein of the cerebellum 1 | NE, EGL, NTZ | 12.8 | 12.18 | -60% |

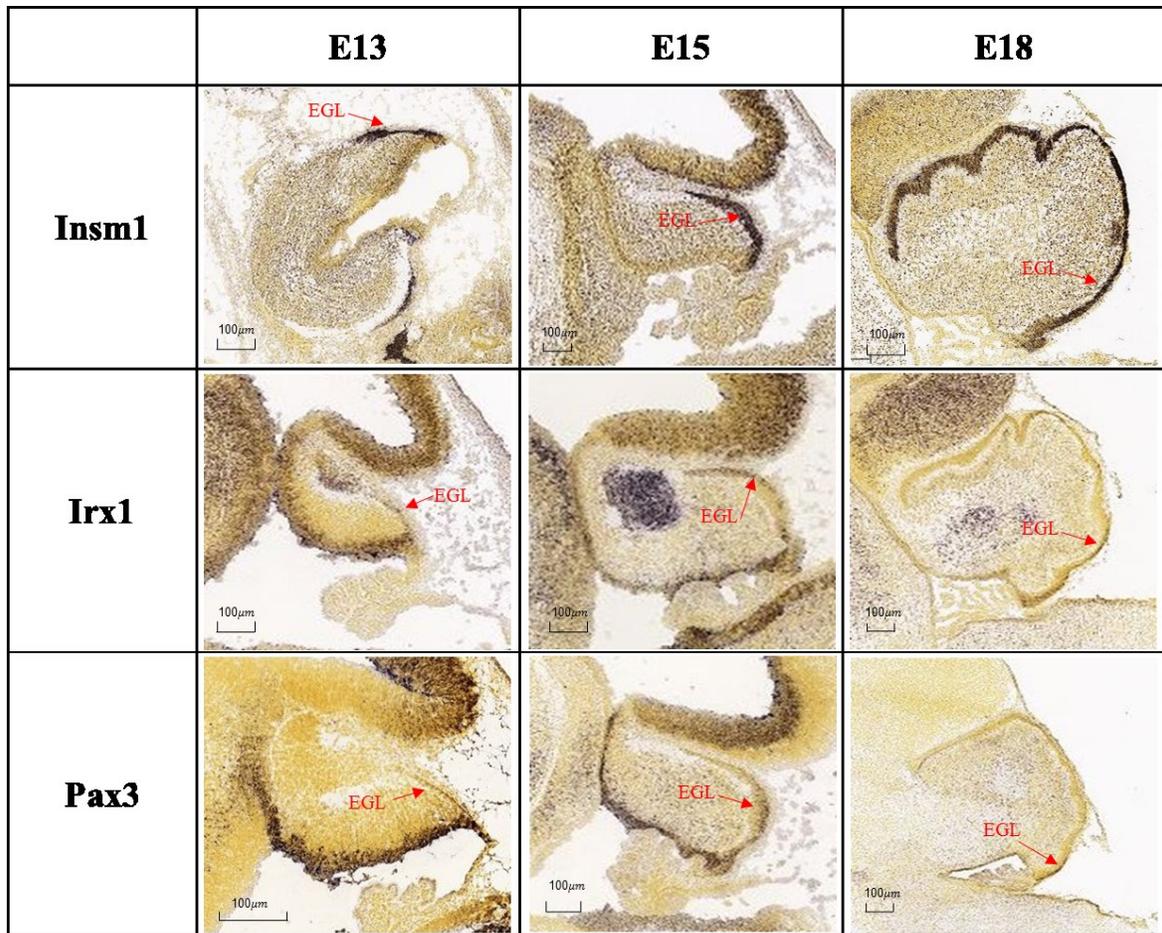


Figure 4.2. In situ hybridization expression pattern (from ABA) of three genes that were found to be significantly enriched in the EGL LCM material.

These images illustrate both the dynamic nature of express at E13, 15, and 18 and the expression of these genes in the EGL. (arrows point to EGL)

4.3.5. Gene Ontology analysis for granule cell enriched genes

To identify cellular processes and molecular pathways in the granule cells, we used Database for Annotation, Visualization and Integrated Discovery program (DAVID, <https://david.ncifcrf.gov/>) to examine the gene ontology of granule cell enriched genes. The top 10 GO terms at each of the ages of analysis are shown in Table 4.5. At E13, processes involving macromolecule assembly, such as actin, are activated in the granule cell precursors from EGL tissue. At E15, the enriched granule cell genes are mostly related to chromosome/histone function. At E18, signal transduction and extracellular matrix genes get an elevated expression in the granule cell precursors compared with the whole cerebellum (Table 4.5).

Table 4.5. Gene Ontology analysis of granule cell enriched gene

Total number of output (N) are shown in the top row; top 10 terms are shown in the table.

P-value for the associated term are shown in (parentheses)

| E13 LCM vs CB (N=50) | E15 LCM vs CB (N=109) | E18 LCM vs CB (N=189) |
|--|---|--|
| Macromolecular complex assembly(0.002) | Histone core (2.09E-18) | Extracellular matrix (2.02E-20) |
| Acetylation (0.002) | Chromosomal protein (3.78E-17) | Secreted (6.01E-15) |
| Structural molecule activity (0.004) | Cellular macromolecule complex assembly (8.52E-14)) | Signal (2.31E-12) |
| Actin binding (0.005) | Methylation (5.82E-9) | Glycoprotein (3.61E-12) |
| Actin cytoskeleton organization (0.01) | Acetylation (1.54E-7) | Disulfide bond (2.46E-10) |
| Actin filament-based process (0.01) | DNA binding (2.57E-5) | Cell adhesion (5.47E-6) |
| Focal adhesion (0.03) | Phosphoprotein (7.39E-5) | In utero embryonic development (2.60E-5) |
| Signal (0.04) | Nucleus (0.002) | Phosphoprotein (0.002) |
| Cytoskeleton (0.05) | Secreted (0.03) | Regulation of Cell Proliferation (0.02) |
| Non-membrane-bounded organelle (0.05) | Signal (0.06) | Calcium ion binding (0.02) |

4.3.6. Motif Activity Response Analysis

We used Motif Activity Response Analysis (MARA) to determine the key regulatory genes behind the expression variation of temporally regulated genes. We examined 196 motifs with known sequence matrices and 113 motifs showed

changes in motif activities in our time series. 26 of these 113 were statistically significant (Z-score > 1.70, Table 4.6). A binding motif can be shared by multiple TFs in the same TF family; for example, the E2F1..5 motif could be bound by 5 genes from the E2F family - E2F1, E2F2, E2F3, E2F4 and E2F5 (Table 4.6). Furthermore, a binding motif could be a target of a transcriptional complex consisting of multiple TFs; for example, the NFKB1_REL_RELTA motif could be bound by a transcriptional complex consisting of 3 genes - NFKB1, REL and RELTA (Table 4.6). Therefore, when accounting for motif sharing, the 26 motifs that are statistically significant from MARA analysis (Table 4.6, Col 1) could represent 46 TFs (Table 4.6, Col 2) which could drive the temporal variations in the cerebellar granule cell transcriptome during embryonic development. Figure 4.3 shows 9 TFs that have an alteration in motif activity which indicates that at one time point, their bioinformatically-predicted downstream targets are up-regulated; and at another time point, their downstream targets are down-regulated. For example, Jun and Foxp1 showed a shift in motif activity from E13 (negative motif activity) to E15 (positive motif activity); on the other hand, Tbp and Tead1 showed continuously decreasing motif activity over the three time points – positive motif activity at E13, around 0 at E15 and negative activity at E18 (Figure 4.3). Moreover, E2f1..5 and Rfx1 shifted their motif activity twice over the time series (expression suppression at E13, activation at E15, and back to suppression at E18) – indicating their dynamic regulatory roles during development of granule cell precursors.

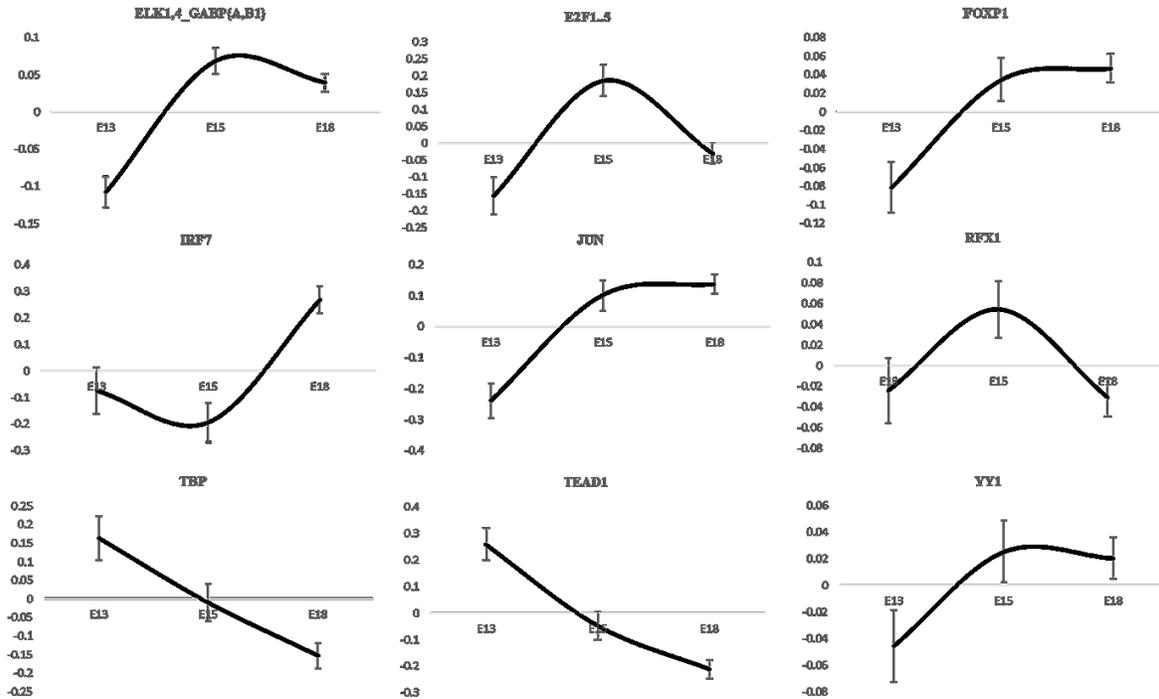


Figure 4.3. Nine genes with significant changes in motif activity during cerebellar development discovered with Motif Activity Response Analysis.

Motif activity (shown on the y-axis) are graphed at E13, E15 and E18 (shown on the x-axis). Positive activity indicates an activation of the transcription factor's downstream targets while negative activity indicates repression of its downstream targets.

Table 4.6. Twenty-six motifs that showed significant change in motif activity (shown in z-value >1.7) during cerebellar development discovered with MARA

| Motif | Transcription Factors | z-value | p-value |
|------------------|------------------------------|-------------|-------------|
| ELK1,4_GABP[121] | ELK1, ELK2, GABPA, GABPB1 | 3.005623716 | 0.001325184 |
| IRF7 | IRF7 | 2.513639314 | 0.005974629 |
| NRF1 | NRF1 | 2.506610635 | 0.006094745 |
| FOXP1 | FOXP1 | 2.486928373 | 0.006442567 |
| TBP | TBP | 2.43734896 | 0.007397697 |
| UFEwm | UFEwm | 2.434473691 | 0.007456732 |
| IRF1,2 | IRF1, IRF2 | 2.35314762 | 0.009307621 |
| TEAD1 | TEAD1 | 2.306134952 | 0.010551546 |
| MZF1 | MZF1 | 2.211013209 | 0.013517461 |
| NFY[154] | NFYA, NFYB, NFYC | 2.097906677 | 0.017956695 |
| RREB1 | RREB1 | 2.062720323 | 0.019569604 |
| RFX1 | RFX1 | 2.036539168 | 0.020848122 |
| JUN | JUN | 2.017613214 | 0.02181578 |
| STAT2,4,6 | STAT2, STAT4, STAT6 | 1.978749288 | 0.02392212 |
| ETS1,2 | ETS1, ETS2 | 1.977655696 | 0.023983781 |
| SPIB | SPIB | 1.929570221 | 0.026830056 |
| EGR1..3 | EGR1, EGR2, EGR3 | 1.920689887 | 0.027385408 |
| E2F1..5 | E2F1, E2F2, E2F3, E2F4, E2F5 | 1.903875641 | 0.028463191 |
| NFKB1_REL_REL | NFKB1_REL_REL | 1.90102444 | 0.028649406 |
| ZNF143 | ZNF143 | 1.775024063 | 0.037946957 |
| MYFfamily | MYFfamily | 1.75724137 | 0.039438338 |
| HIC1 | HIC1 | 1.74976657 | 0.040079301 |
| TFAP2B | TFAP2B | 1.729154293 | 0.041890742 |
| TOPORS | TOPORS | 1.724542156 | 0.042305008 |
| YY1 | YY1 | 1.721785034 | 0.042554233 |
| NANOG | NANOG | 1.703453823 | 0.044241586 |

4.3.7. Experimental validation

For quantitative validation, we performed qRT-PCR on 8 genes (Ciz1, Hes6, Insm1, Irx1, Klf3, Pax3, Rfx3 and Tcf4, Figure 4.4). These genes were chosen because they offer the greatest potential for novel findings on the role of TFs in cerebellar development; i.e., they had the highest q-Value scores, no prior knockout phenotype documented and no prior literature on their roles during cerebellar development. In Figure 4.4, we observed a similar expression pattern between our CAGE data (Figure 4.4a, c, e, g, i, k, m, o) and qRT-PCR data (Figure 4.4b, d, f, h, j, l, n, p). For example, Ciz1 showed activated expression at E15 and E18 in the granule cells from the CAGE data (Figure 4.4a) and the qRT-PCR showed the similar pattern (Figure 4.4b). Genes such as Hes6 and Insm1 had enriched EGL expression at E15 in CAGE data (Figure 4.4e and i) and showed similar enrichment in our qRT-PCR data with scraped EGL cells (Figure 4.4i and j).

For qRT-PCR data, the expression of genes are measured as relative quantity against H₂O as negative control (RQ, shown on the y-axis) at E13, E15 and E18 (shown on the x-axis). Red plots represent gene expression in the whole cerebellum while blue plots represent gene expression from the needle-scraped EGL tissue containing cerebellar granule cells.

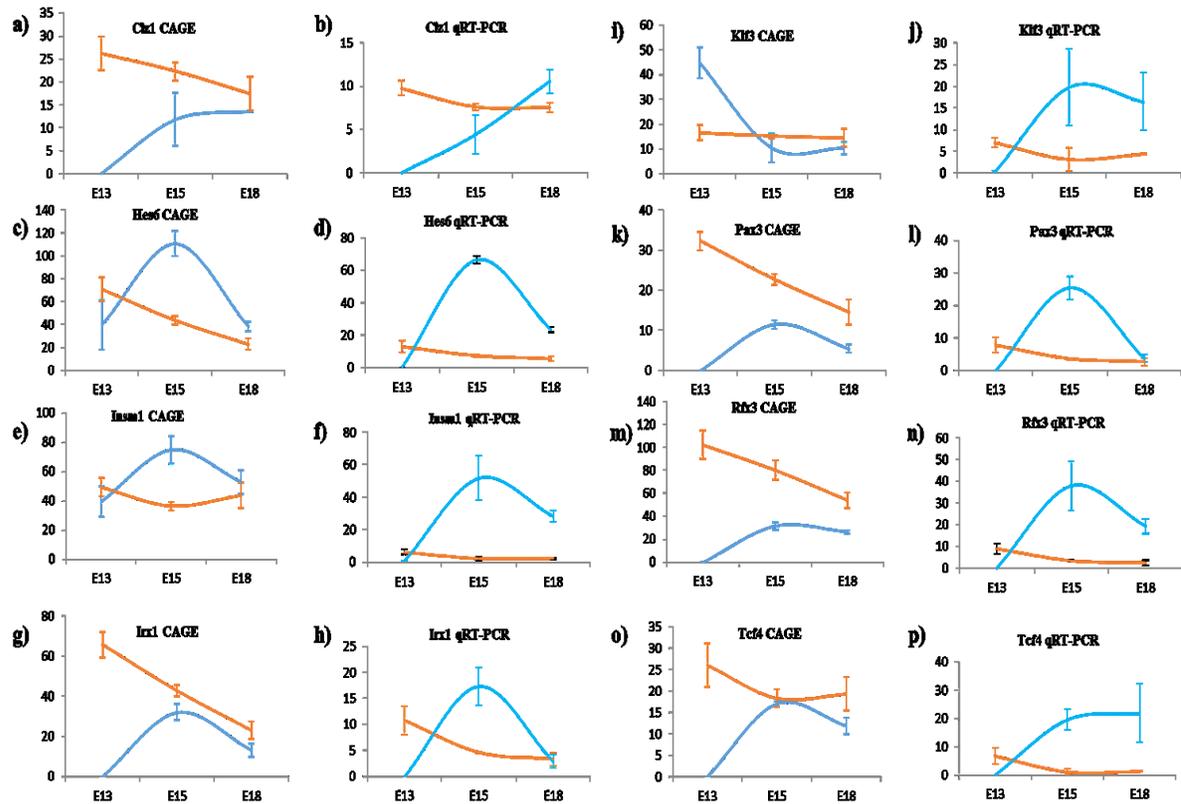


Figure 4.4. Eight genes significant in the differential expression analysis and their quantitative real-time PCR validation at E13, E15 and E18.

Each gene has its expression measured from HeliScopeCAGE transcriptome data on the left panel and qRT-PCR data on the panel to its right.

For HeliScopeCAGE data, the expression of genes is measured in tags-per-million (tpm, shown on the y-axis) and are shown at E13, E15 and E18 time points (displayed on the x-axis). Red plots represent gene expression in the whole cerebellum while blue plots represent gene expression in the LCM isolated EGL tissue containing cerebellar granule cells.

4.4. Discussion

4.4.1. The cerebellar granule cell precursor transcriptome

Cerebellar granule cells undergo complex processes such as cell specification, differentiation, proliferation and migration throughout development - each process requiring a set of genes to be regulated at the appropriate time. Mis-regulation of these cellular processes could lead to developmental defects – such as the medulloblastoma, which is thought to be associated with proliferation and apoptosis defects in cerebellar granule cells[155]. Transcription factors (TFs) are key regulators for these developmental processes since dysfunction of one TF could subsequently affect its downstream targets creating a cascade effect. Since the cerebellum consists of many types of neurons and glial cells, the study for granule cell-specific transcriptional regulators is inherently difficult. Therefore, discoveries of cerebellar TFs in the past usually utilized one of two (or a combination of the two) strategies: 1) Studies of single genes that are found to be expressed in the granule cells with granule cell-specific markers [156-158]; and 2) Studies of the whole cerebellar transcriptome with technologies such as microarray, followed by the generation of expression based databases; cerebellar development transcriptome database (CBT-DB, now known as BrainTx[60]) and Cerebellar Gene Regulation in Time and Space (CbGRiTS[61]) are examples of such studies. As an effort to retain the combined merits of these two traditional strategies, our study utilized LCM to isolate granule cell containing EGL, followed by HeliScopeCAGE to provide a holistic view of the granule cell transcriptome during early cerebellar development..

To look at the whole EGL transcriptome, we employed three bioinformatics analyses: differential expression, motif activity and gene ontology. Each analysis has its own strengths and weaknesses. However, when the three approaches are combined, we were able to identify potential key regulatory genes in granule cell development, which were further examined with experimental validation.

The differential expression analysis of the LCM time series, or between the LCM and whole cerebellar time series, identifies genes from all classes (membrane proteins, structure proteins, transcriptional factors, etc.) that are either temporally-regulated or granule cell-enriched. However, differential expression analysis does not focus on regulation of downstream targets since the results would be solely expression-based and lack information on TF binding motifs. Motif activity response analysis complements the weakness of differential expression analysis by identifying the activation or suppression of TF binding motifs, thus providing predictive data on the binding of TFs to their downstream targets. However, motif activity is limited by its requirement of established knowledge on motif matrices and its inability to look at genes from functional groups other than TFs. Finally, gene ontology analysis enables us to arrive at functional interpretations of our LCM time series; allowing us to validate bioinformatics predictions with previous knowledge and hypothesize the functional roles of key genes that are previously unknown. For example, at E15, the enriched granule cell genes belong mainly to chromosome/histone-related function. This makes biological sense as we know E15 is a time when the EGL is largely proliferative with cells engaged in the cell cycle and chromosomal structural genes are activated [24]. At E18, signal transduction and extracellular matrix genes

become activated, this is most likely due to the initiation of cell migration and the start of synapse development. Thus, the results of GO analysis allow us to explore the functions of our potentially important granule cell genes. However, gene ontology analysis is limited by its requirement of knowledge on gene structure, function and interaction; and its limitation to a small number of input genes (~3000[159]) generated by other analyses such as differential expression and MARA.

The differential and MARA analyses revealed 125 TFs (82 from differential expression analysis and 46 from motif activity analysis that includes 3 overlapping genes between the two sets) that may be important for the development of granule cells; and 71 of these TFs have expression data available on *in situ* database Genepaint (<http://www.genepaint.org>) or ABA (<http://www.genepaint.org>). These TFs fell into three groups: 1). Sixteen previously appreciated granule cell genes where at least 5 papers had been published on these genes in the cerebellum from the PubMed database. 2). Eighteen genes with limited knowledge that have between 1 to 5 publications associated with cerebellum; and 3). Thirty-seven genes were novel to cerebellar development where no previous publications were found to be associated with the cerebellum. The high proportion (>50%) of novel TFs highlights the productive nature of the LCM approach to discover novel genes that may play important roles in cerebellar development.

4.4.2. Confirmation of known genes in cerebellar granule cell development

From our study, 16 genes have at least 5 publications associated with the cerebellum in PubMed (www.ncbi.nlm.nih.gov/pubmed, Table 4.3). These genes have been reported to play critical roles during granule cell development. For example, *Atoh1* (also known as *Math1*, up-regulated at E13) is a bHLH transcription factor that is required for the specification of granule cells at the rhombic lip [21, 22]. Another gene that showed activation at E13 in our data is *Zic1* (zinc finger protein of the cerebellum) which is up-regulated at (E13) and is an important TF for granule cell proliferation and cerebellar foliation. Lastly, *Gbx2* (gastrulation brain homeobox 2) plays a key role in forebrain and hindbrain development and a conditional gain-of-function transgene leads to deletion of the cerebellum in mice[160].

There are two genes in this group with overlapping significance in the MARA and differential expression analysis: *E2f1* (27 publications) and *Yy1* (7 publications). *E2f1* (E2F transcription factor 1) is enriched in the granule cells at E15 in our analysis. It is involved in a caspase 3-independent apoptosis pathway in the granule cells [20]. *E2f1* plays opposing roles in mice (promoting apoptosis of granule cells) and rats (antagonizing apoptosis of granule cells)[161] hinting at its recent functional alteration in rodents. *E2f1* expression is increased in medulloblastoma suggesting that it may be involved in the up-regulated *Shh* proliferation pathway [24]. *Yy1* (Yin Yang 1) is down-regulated at E15 and E18 in the granule cells in our analysis. *Yy1* has been previously found to be a significant downstream target of stress-induced granule cell apoptosis[162]. *Yy1* has been identified to be part of a DNA complex that inhibits granule cell apoptosis. This inhibition effect can be disrupted by cytotoxic insults on the granule cells that cause the degradation of *Yy1* DNA

complexes and results in an increased rate of cell death [163]. However, the effect of down-regulation of Yy1 on granule cell survival during normal cerebellar development remains largely unknown.

4.4.3. Genes with limited information pertaining to cerebellar development

Our analyses revealed 18 genes that have 1-4 publications associated with the cerebellum. Thus, while these genes may have some information about their expression or function in the cerebellum, this information is limited. Our study would then support emergent roles for these genes in the cerebellum as well as providing novel temporal and spatial information for future studies. Here, we focus on two aspects of these genes – genes that showed significance in multiple analyses and genes that may play important roles in granule cell proliferation. Cell proliferation is active throughout granule cell development from E13 to early postnatal stages[164]. This proliferation of granule cells involves multiple molecular/signalling pathways[56, 147, 165] that are responsible for making cerebellar granule cells the most numerous neuronal type in the brain[166].

In addition to E2f1 and Yy1 (described in the previous section), the third of the three genes that are significant for both the differential expression and MARA analysis is Ets1 (E26 avian leukemia oncogene 1). While Ets1 has 2 previous publications associated with the cerebellum, it has been associated with cerebellar disorder - spinocerebellar ataxia type 2, as a direct activator of the mutation causing gene – Atxn2[25].

We identified two TFs, *Irx1* and *Ciz1*, that have high expression levels in the EGL at E15 and E18 (compared to E13). Although their functions are unknown in the cerebellum, we speculated that these TFs might play a role in granule cell proliferation based on their expression in the highly proliferative cells of the EGL. *Irx1* is a member of the Iroquois homeobox gene family. Members of this family play multiple roles during pattern formation in embryos as well as the development and patterning of lungs, limbs, heart, eyes, and nervous system[167-171]. Interestingly, *Irx1* regulates the Shh pathway in the retina through transcription activation of *Irx2*[172]. Since *Irx1* expression is activated in the granule progenitors during cerebellar development at E15, it could regulate granule cell proliferation through the well-known Shh pathway[56]. The second gene candidate is Cip1-interacting zinc finger protein (*Ciz1*), a zinc finger DNA binding TF that interacts with CIP1 (p21 / CDKN1A). It has been identified as an oncogene for gallbladder[173], colorectal[174] and lung[175] cancer that plays an important role as part of a cyclin E complex during cell cycle. Our data showed an activation of *Ciz1* at E18 suggesting that it could function as part of the cell proliferation through a cyclin-related pathway during late embryonic cerebellar development.

Insm1 (Insulinoma-associated protein 1) is one gene of interest in this group due to its ~4 fold reduction in the *Atoh1*-null cerebellum. While it has only 3 publications associated with the cerebellum, it is highly enriched in the granule cells at E15 from our differential expression analysis and its expression is abolished in the *Atoh1* mutant that lacks granule cells. *Insm1* has no introns, and it encodes a transcription factor with a zinc finger DNA-binding domain and a putative prohormone domain.

Previously, it has been found that *Insm1* is highly expressed during the neuroendocrine differentiation in human lung cancers[176]. Its functional role in the cerebellar granule cells is yet to be investigated.

4.4.4. Discovery of novel transcription factors that may be involved in granule cell development

Over half (37 out of 71) of the differentially expressed TFs in our analyses have no publications related to the cerebellum; and three of these genes (*Creb5*, *Gtf3a* and *Ikzf5*) have no functional annotation in any tissues. From the 34 genes with previous publications, we speculated that these are novel regulators of granule cell development based on functional roles of these genes in other tissues. *Hes6*, *Rfx3* and *Onecut2* are examples of three genes that have been shown to be involved in neurogenesis.

Hes6 is a member of the hairy enhancer of split family of TFs that is expressed in the developing cerebral cortex [177]. It has been found to be an inhibitor of cortical astrocyte differentiation and promote neurogenesis[178]. Our findings indicate that it could also be associated with neurogenesis of the granule cells where it is activated and enriched in the EGL in the cerebellum at E15. The granule cell-specific transcription factor *Rfx3* is a member of the regulatory factor X gene family which encodes TFs that contain a highly-conserved winged helix DNA binding domain. It is involved in the development of cilia[179] and the corpus callosum[180]. *Rfx3*, which is abundantly expressed in neuronal cells, down-regulates the *Map1a* pathway resulting in the repression of neuronal differentiation[181]. *Rfx3* expression is activated in the EGL at E15 and E18 which coincides with the neurogenesis of granule cell precursors. Lastly, *Onecut2* is a

member of Onecut transcription factor family that has been mostly studied for cell differentiation in intestine, pancreas and liver [182-184]. Interestingly, one recent report found that it is involved in neurogenesis of the horizontal cells of the retina[185]. In addition, the transcription factor Pax6 has been identified as Onecut2's upstream regulator and Ptf1a as one of its downstream targets[185]. Since both Pax6 and Ptf1a have been found to play important roles during cerebellar development [24, 54, 158], further study on Onecut2 in cerebellar granule neurons could reveal the Pax6/Ptf1a pathway in cerebellar development.

4.4.5. Summary

In this study, we utilized laser capture microdissection to isolate the EGL containing cerebellar EGL cells. We identified 125 transcription factors as potential key regulators for cerebellar granule cell development. From this gene set, we further identified 37 transcription factors that had no previous knowledge about their roles in cerebellar development. The results from genome-wide analyses were validated with existing online databases and qRT-PCR. This study provides an initial insight into the transcription factors of cerebellar granule cells that might be important for development and provide valuable information for further functional studies on these transcriptional regulators.

Chapter 5 : Kruppel-like factor 4 regulates granule cell Pax6 expression and cell proliferation in early cerebellar development

5.1. Introduction

The cerebellum, which represents about 10% volume of the brain, consists of more than half of all neurons in the brain. The numerous cerebellar neurons belong to only a few neuronal groups and are arranged in a simple and well-defined cytoarchitectural organization[1]. However, development of the seemingly “simple” cerebellum requires a precise spatial and temporal regulation of different cellular processes. The cerebellar neurons can be grouped by their neurotransmitters: the excitatory neurons utilizing glutamate and the inhibitory neurons utilizing GABA. The glutamatergic granule cells (GCs) and other excitatory neurons and the GABAergic Purkinje cells (PCs) and inhibitory interneurons are born from two distinct neurogenic regions at different developmental stages; through tightly regulated developmental processes, these neurons proliferate, differentiate, migrate and interact to produce a mature and functional cerebellum. Thus, the cerebellum serves as an excellent model for neurodevelopmental research because it is structurally simple, yet the cerebellar neurons undergo all the major developmental events, such as cell specification, differentiation, proliferation and migration, that are critical and common to development of the central nervous system, in general.

The glutamatergic GC precursors are generated from a germinal epithelium known as the rhombic lip at around embryonic day 10 (E10). The rhombic lip progenitors, located lateral and caudal to the NE, are specified by the bHLH

transcription factor, Math1 (also known as Atoh1)[35, 42]. The GC precursors, which give rise to differentiated GCs, most actively proliferate from E15 to post-natal day 8 (P8)[35, 37]. At E12, the GCs begin their first migration to cover the dorsal surface of the cerebellum forming the external granular layer (EGL). It is not until the time of birth that the GCs begin their second migration into the cerebellar parenchyma to their final destination where they form the IGL. The EGL ceases to exist at around P21[1]. The paired-box transcription factor, Pax6, is strongly expressed in the GC throughout the course of development [54]. Pax6 is important for GC development[58]; however its regulatory pathway is not well understood; thus, we are interested in identifying genes that regulate, or are regulated by, Pax6 because these genes might play important roles during cerebellar development through interactions with Pax6.

Klf4 was bioinformatically identified as an important genetic regulator in Chapter 2 and 4 as well as a transcriptional regulator for Pax6 by the bioinformatic program - Enhancer Element Locator (EEL)[186]. Klf4 is one of the four genes necessary to create an induced pluripotent stem cell and has been extensively studied for its role in cell proliferation, differentiation and survival in multiple cell types[187] and its association with Pax6 has been documented in corneal development[87, 88]. However, the role of Klf4 in cerebellar development remains unknown. We hypothesized that Klf4 is a key transcription factor for cerebellar development as a regulator of Pax6.

Klf4 belongs to the Kruppel-like factor family, which contains three C-terminal C2H2-type zinc fingers that bind DNA. The name “Kruppel-like” comes from its

strong homology with the *Drosophila* gene product Kruppel, an important gene in segmentation of the developing embryo. Klf4 has been studied for its roles in stem cell maintenance, oncogenesis and embryonic development. Klf4 is one of the four genes necessary to create an induced pluripotent stem cell; although the mechanism of Klf4 in the self-renewal of the stem cell remains unclear, it is speculated that it might function to maintain cell proliferation[89] or inhibit apoptosis[188]. Klf4 also plays important roles in tumorigenesis – depending on tissue and environment, it can function as an oncogene as the over-expression of Klf4 could repress expression of p53 through the Ras/P21 pathway[189], which would prevent cell apoptosis. On the other hand, Klf4 can also function as a tumor suppressor as it can antagonize the Wnt pathway resulting in the inhibition of cell proliferation. Lastly, Klf4 is an important transcription factor for homeostasis of multiple tissue types. It is essential for the differentiation of goblet cells in the colon as knocking out Klf4 resulted in the absence of these cells [190]. Klf4 is also critical for development of the granular layer of the skin[191]; the Klf4^{-/-} mouse dies several hours after birth due to the defective body barrier which causes extensive loss of body fluid[191]. The function of Klf4 in brain development has been studied through myc-activated overexpression where cell proliferation and differentiation are inhibited along with defects in cilia genesis that lead to hydrocephalus[91].

Klf4 has been identified as a tumor suppressor gene that is frequently inactivated in medulloblastoma [192] – a tumor that oftentimes originates from cerebellar granule neurons. However, the role of Klf4 in normal cerebellar development has not been studied. Here we report our findings of roles of Klf4

during cerebellum development. We find that Klf4 is important for granule cell proliferation through E13.5 and E15.5; we also find that Pax6 expression is lowered in the Klf4^{-/-} cerebellum and we find that Klf4 acts as an upstream regulator of Pax6.

5.2. Materials & methods

5.2.1. Expression analysis and transcription factor binding site prediction

Two databases were used to access quantitative expression of genes of interest in the cerebellum: 1) CbGRiTS is a time-course, microarray database constructed from transcriptomes of mouse cerebellum from E12 to P9 [61]. 2) FANTOM5 contains 5'cap sequencing, time-course data of mouse cerebellar transcriptome from E12 to P9) [62]. To examine the spatial expression of genes of interest in the developing mouse cerebellum, histological data of various cerebellar transcription factors in wild-type mouse were obtained from the online Genepaint (www.genepaint.org) and Allen Brain Atlas (www.brain-map.org) databases.

To determine if there were any transcription binding sites in genes with strong granule-cell expression, Genbank (<http://www.ncbi.nlm.nih.gov/genbank>) sequence data from 200 Kbp upstream of the transcript start site were analyzed with Enhancer Element Locator (EEL) software [186]. The detailed methods for EEL have been described elsewhere [85]; the values for parameters Lambda, Xi, Nu, Mu and Nucleotides Per Rotation were set at the default setting for the program: 2.0, 200.0, 200.0, 0.5 and 10.4, respectively. Two statistical cut-offs were used for Enhancer Element Locator analysis: a p-value less than 0.001, which represents the significance that a binding site is over-represented upstream of Pax6 when compared with whole genome background; and a confidence level of 92%, which measures the number and conservation of binding sites found upstream of Pax6[186].

5.2.2. Klf4 colony maintenance and breeding

Canadian Council on Animal Care approved this research of ethical approval (approval number - A12-0190). The research was conducted in accordance with these policies and all efforts were made to minimize suffering. Klf4-null mice (Klf4^{-/-}) were a gift from Elaine Fuchs' Lab at Rockefeller University; they were housed at University of Chicago accredited by Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) with approval by Animal Care and Use Committee (approval number - NHGRI ACUC 08-0059). The knock-out of Klf4 was achieved by a substitution of the entire exon 2 and 3 as well as part of exon 1 of the Klf4 gene with a neomycin sequence (neo) on a C57BL/6 background; this results in the elimination of expression of Klf4 transcripts and protein products[191]. We recovered heterozygous Klf4 animals with in-utero transfer of frozen embryos. The colony was maintained in a heterozygous state that showed normal cerebellar and behavioral development. Klf4 knockout mice are perinatal lethal due to defect in skin development which causes excess loss of body fluid shortly after birth[191]. The Klf4^{-/-} embryos were generated with time pregnancy and collected at embryonic stages (E13.5-E18.5) and 3-7 embryos were used for immunohistological study at each age.

5.2.3. Histological methods and analysis

Trans-cardiac perfusion with PBS and 4% paraformaldehyde was used to prepare animal tissues after deep Avertin anesthesia. Tissues were post-fixed *in situ* for 2 hrs in 4% PFA. The brains were dissected and stored at 4 °C in PBS (0.1 M, containing 0.02% Na Azide) until processing. Tissues were cryoprotected in 30% sucrose in PBS overnight or until they sunk to the bottom of the solution and serially cryosectioned at 12–16 µm. Serial sections from brains were stained with cresyl violet for histology. Single-label immunofluorescence staining of the tissue was carried out as previously described: anti-Klf4 (AF3158, goat, R&D Systems, 1:200), to detect Klf4 expression in the wild-type animals; anti-Pax6 (PRB-278P, rabbit, Covance, 1:200), to highlight wild-type granule cells and granule cell precursors; anti-Calbindin D-28k (AB1778, rabbit, Chemicon, 1:200), to detect Purkinje cell soma and dendrites; anti-Pax2 (71-6000, rabbit, Invitrogen, 1:200), to identify interneuron precursors; and anti-Gfap(sc-51908, mouse, Santa Cruz, 1:200), anti-Glast(MABN794, goat, Millipore, 1:200), to identify cerebellar glial cells. Anti-Klf2(bs-2772R, rabbit, Bioss, 1:200), anti-Klf5(bs-2385R, rabbit, Bioss, 1:200) and anti-β catenin(ab6302, rabbit, abcam, 1:200) were used for further investigation of the Klf4 pathway. The secondary antibodies used for these analyses were donkey anti-goat (95382, Jackson ImmunoResearch, 1:200), goat anti-Rabbit whole IgG Alexa 594 conjugate (A11012, 1:200) and goat anti-mouse F(ab')₂ Alexa 594 conjugate (A11020, 1:200) (Molecular Probes/Invitrogen). Analysis and photomicroscopy of brightfield histochemistry was performed with a Zeiss Axiophot microscope with the Axiocam/Axiovision hardware-software components (Carl Zeiss).

Two quantitative analyses were conducted on histological sections: the number of Pax6+ cells in EGL and the number of BrdU+ in three regions of the cerebellum (the rhombic lip, the EGL, and the neuroepithelium). Cell counts were started at the first appearance of the external granular layer in the cerebellum, and at every 10th sagittal section that followed, throughout both sides of the cerebellum. The Klf4-null cerebellum was compared against the wild-type cerebellum using one-tailed Students' T-test. $p < 0.05$ was considered a significant difference in cell number between the two groups.

5.2.4. Assays for cell proliferation and cell death

To examine cell proliferation, mice were injected with 50 mg BrdU/kg 60 minute before to perfusion with a 3:1 70% EtOH:acetic acid. For anti-BrdU immunohistochemistry, brains were embedded in paraffin, sectioned on a microtome at 16 μ m, and mounted on glass slides. Sections were deparaffinized, rehydrated, and treated with 1M HCl for 30 minutes at 37 °C. Then, the slides were incubated with mouse anti-BrdU monoclonal antibody (1:200 dilution; BD Biosciences, Mississauga, ON, Canada) overnight followed by incubating in biotinylated horse anti-mouse immunoglobulin (1:200 dilution; Vector Laboratories, Burlingame, CA, USA) for 1 hour on the next day. Lastly, the slides were stained using the VECTASTAIN Elite ABC kit (Vector Laboratories) and 3, 3'-diaminobenzidine (SigmaAldrich). For assessing cell death, the mice were processed for perfusion and the brains for sectioning as described for the cell proliferation work, above. The

slides were immuno-stained with ApopTag® Plus Fluorescein In Situ Apoptosis Detection Kit (S7111, ApopTag FITC-direct, Chemicon).

5.2.5. Real-time PCR

cDNA from Klf4-null and wild-type littermates were produced with random hexamers using the High Capacity cDNA Archive kit (Applied Biosystems). cDNA products were diluted to 100 ng total RNA input. Sequences of the transcript of interest were loaded into Primer Express® software (Applied Biosystems). Amplicon lengths were between 75 and 125 bp. The qPCR was performed with the FAST SYBR Green PCR Master Mix (Applied Biosystems) on an ABI StepOne Plus Sequence Detection System (Applied Biosystems). All runs were normalised to 18s RNA. Three biological replicates were prepared for each gene target and three technical replicates were performed for each biological replicate. Gene expression was represented as relative quantity against the negative control which used water as the template (noted as “Relative Quantity vs. H₂O” in figures).

The results of Real-Time PCR were analyzed and graphed by ABI StepOne Plus Sequence Detection System (Applied Biosystems). The expression data was statistically analyzed using a one-tailed Students' T-test with $p < 0.05$ as significant.

5.3. Results

5.3.1. Discovery of Klf4 as a regulator of Pax6 in the cerebellum

Kruppel-like factor 4 (Klf4) from three independent analyses: 1) In Chapter 2, its motif activity shifts from positive regulation to negative regulation in our whole cerebellum time series; 2) In Chapter 4, it is up-regulated in the EGL cells in our LCM time series at E13; and 3) its binding site has been confirmed in the regulatory region of Pax6 by Enhancer Element Locator (EEL), a program that searches over-represented transcription factor binding sites in the regulatory regions of a target gene [85]. EEL has been used to successfully identify transcription factor binding sites for c-Myc and N-Myc developmental signaling pathways [186]. In this study, with Pax6 set as the target gene, the EEL analysis reveals that Klf4 is among the most significant transcription factors. Klf4's binding sites were in the top list of several other granule cell expressed genes such as Nfia, Cacna1a and Wnt7b. Thus, it was selected for experimental validation for its role in Pax6 regulation and cerebellum development.

5.3.2. Characterization of Klf4 expression

To investigate Klf4 expression in the developing cerebellum, cerebellar tissues were collected from E13.5, E15.5 and E18.5 wild-type and Klf4^{-/-} animals. RT-PCR and immunohistochemistry were performed on E13.5 and E15.5 cerebellar tissue to examine Klf4's expression at RNA and protein levels, respectively. Klf4 is expressed in the wild-type cerebellum in E13.5, E15.5 and P0 (5.S1 Figure).

Immunohistochemistry with anti-Klf4 was performed on E13.5 and E15.5 embryos. The staining in the cerebellum was much weaker than higher expressing regions such as the skin and the posterior rhombic lip (Figure 5.1b). Within the cerebellum, the main expression is in the cells of the EGL (Figure 5.1a). By E18.5, there was no detectible immune-positive staining for Klf4 (Figure 5.1c).

The expression of Klf4 in early developing EGL is co-incident with Pax6 expression[54]. To explore this possibility, we co-stained tissue with antibodies to Pax6 and Klf4. Almost all EGL cells were positive for Pax6 and Klf4 (Figure 5.1d and e). However, other cell populations that were Pax6-positive in deeper regions of the cerebellum were not Klf4-positive (Figure 5.1f).

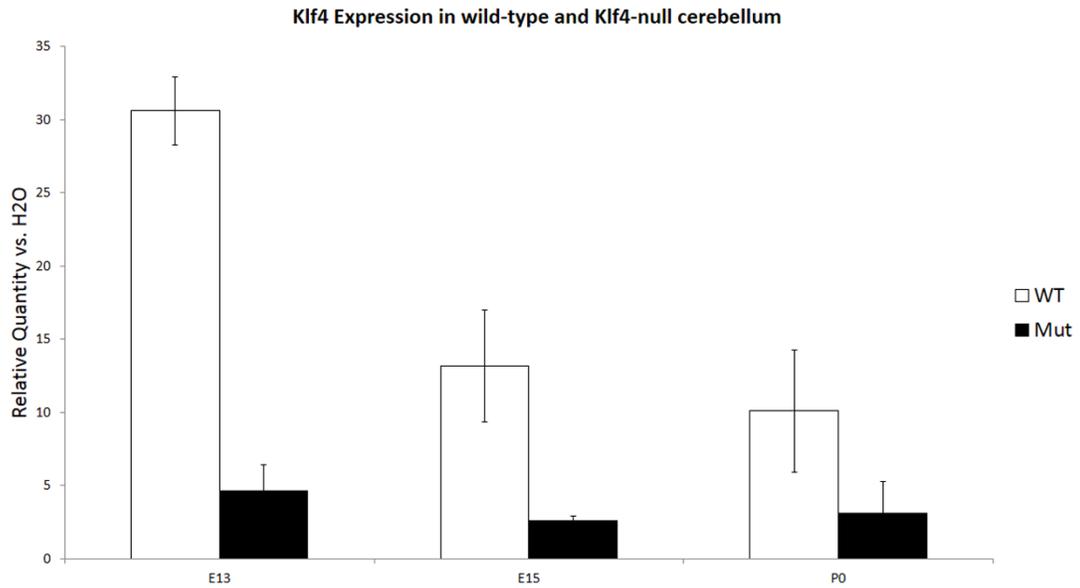


Figure 5.S1 Klf4 expression in wild-type and Klf4-null cerebellum.

Klf4 is expressed in wild-type cerebellum at E13.5, E15.5 and P0. Its expression is greatly abolished in the Klf4-null.

Y-axis: Relative Quantity vs H2O – target gene expression of the sample compared against with a negative control where H2O were used as template

X-axis: WT- wild-type, Mut – Klf4-null

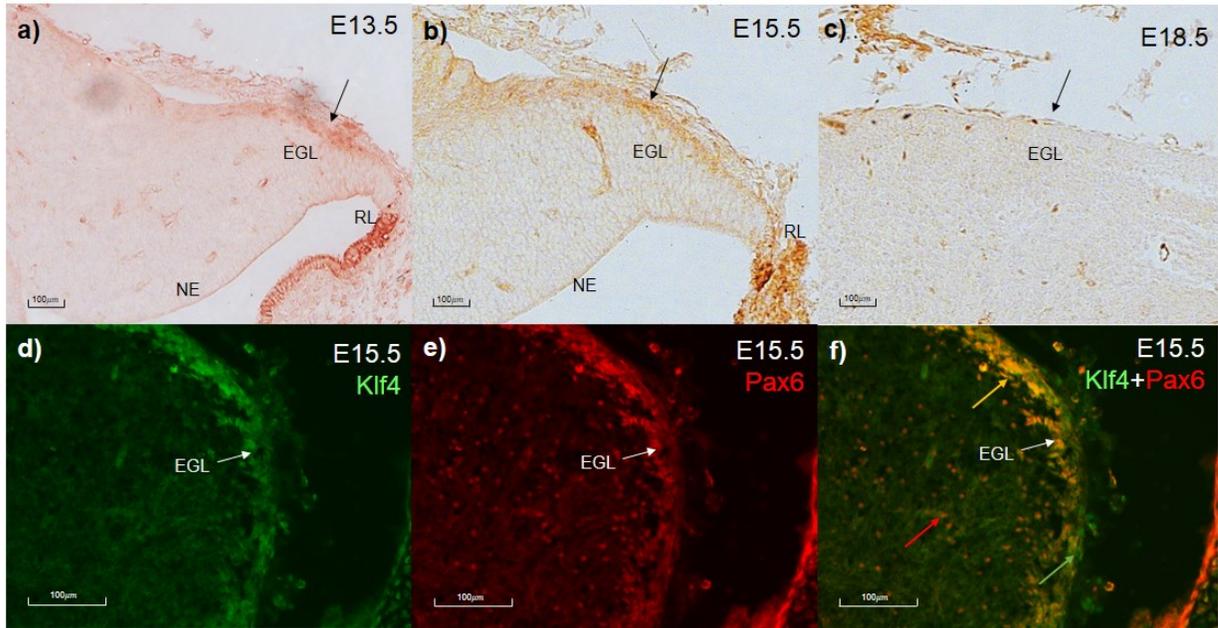


Figure 5.1. Klf4 expression in the cerebellum and its co-expression with Pax6
a-c) Klf4 expression at: a) E13.5, b)E15.5, and c)E18.5. Immunohistochemistry of Klf4 in the developing cerebellum. Klf4 is expressed in the EGL of the cerebellum at E13.5 and E15.5 but virtually no expression is seen at E18.5 (black arrows).
d-f) Co-expression of Klf4 and Pax6 in the EGL at E15.5. Immunofluorescence staining of Klf4 (green, d), Pax6 (red, e) and merged picture (f) in the developing cerebellum. Klf4 and Pax6 are co-expressed in the EGL (white arrows) of the cerebellum at E15.5. In (f), green arrow indicates EGL cells that express only Klf4, red arrow indicates cells in the cerebellar core that express only Pax6, and yellow arrow indicates EGL cells that co-express Klf4 and Pax6.
 EGL- external granular layer, NE – neuroepithelium, RL- Rhombic lip

5.3.3. Pax6 in the developing cerebellum following the elimination of Klf4 expression

In addition to the identification of Klf4 binding sites upstream of Pax6, the immunocytochemical data suggest, an interaction between Klf4 and Pax6. To test this hypothesis more directly, we used a KO of the KLF4 gene to more mechanistically study a possible interplay between genes. The Klf4 knockout is a perinatal lethal and pups die after 4-6 hours after birth as previously reported [191]. Thus, we were limited in the examination of the developing cerebellum to prenatal and very early postnatal times. We observe a marked reduction in Pax6 immunocytochemistry in Klf4^{-/-} EGL cells at E13.5 (Figure 5.2a-b); and an almost complete elimination of staining in the E15.5 Klf4-null when compared to the wild type cerebellum (Figure 5.2c-d). These observations suggest that Klf4 positively regulates Pax6 expression.

To examine if the reduction of Pax6 expression is due to a smaller number of cells expressing similar levels of Pax6 or the same number of cells expressing Pax6 at a lower level; we quantified the number of Pax6⁺ cells. Indeed, fewer Pax6-positive cells are found in the Klf4^{-/-} cerebellum at E13.5 (Figure 5.3a, $p < 0.05$). A similar reduction of Pax6⁺ cells is found in the rhombic lip (Figure 5.3a, $p < 0.001$) but not the proliferative neuroepithelium above the 4th ventricle. The quantitative assessment of Pax6 expression showed the similar reduction as seen in the sectioned and stained material at E13.5 (Figure 5.3b, $p < 0.01$) and 15.5 (Figure 5.3b, $p < 0.01$). Interestingly, at E16.5, there is a return of Pax6-immunopositivity in the EGL and RL (Figure 5.2e and 2f). The return of Pax6-positive staining appears almost complete by E18.5 (Figure 5.2g and 2h). The return of Pax6 expression at

E18.5 is confirmed with real-time PCR (Figure 5.3b). Our observations indicate that *Klf4* normally, positively regulates *Pax6* during early granule cell development, prior to E16.5.

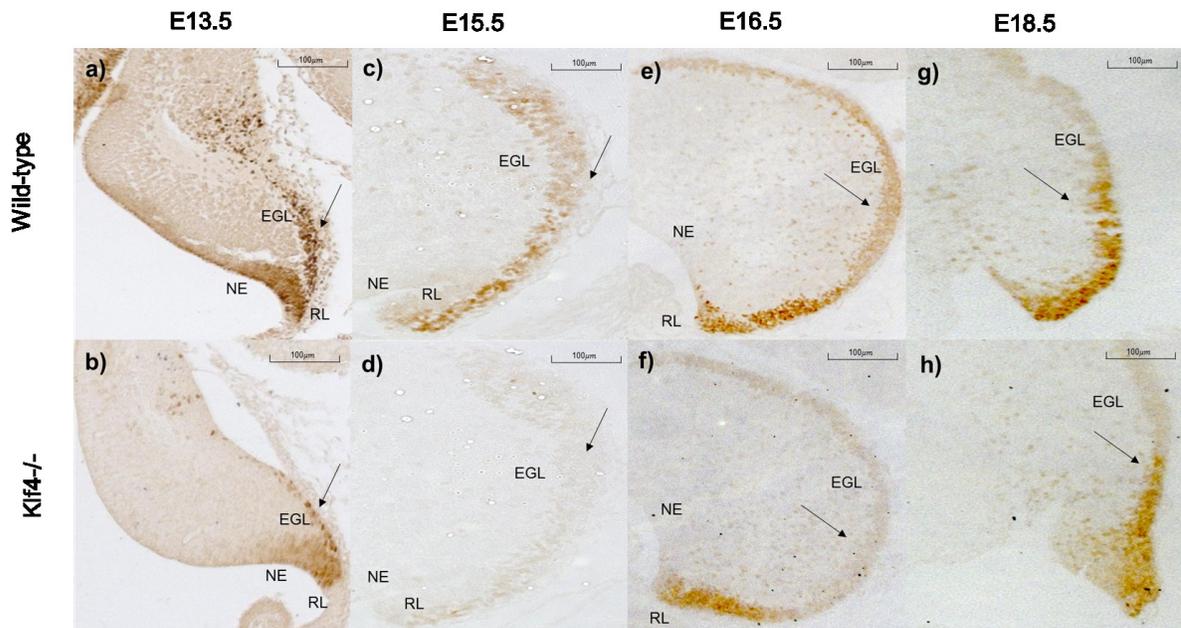


Figure 5.2. *Pax6* is down-regulated in *Klf4*-null cerebellum at E13.5 and E15.5
Immunohistochemical demonstration of *Pax6* expression during development in *Klf4*-wildtype (a,c,e,g) and *-null* (b,d,f,h) cerebellum. *Pax6* immunocytochemistry is similar in the developing EGL of the wildtype cerebellum from E13.5 to E18.5. *Pax6*'s expression is greatly reduced at E13.5 and E15.5 in the *Klf4*-null cerebellum but rebounds at E16.5 and E18.5.

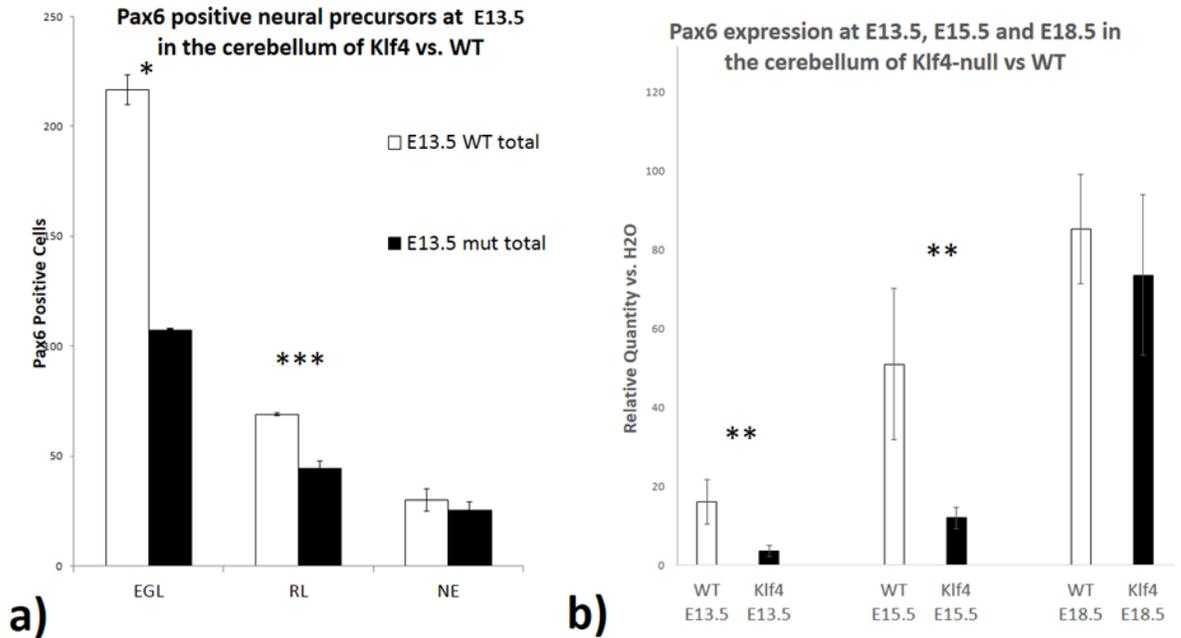


Figure 5.3. Quantification of Pax6 cell number and expression down-regulation in Klf4-null cerebellum.

- a) E13.5 Pax6 positive granule cell count in Klf4^{-/-} compared with wild-type in the EGL ($p < 0.05$), RL ($p < 0.001$), and NE. One-tail students' T-test was used and results were represented with $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***) .
- b) Real-time PCR showing the expression of Pax6 in the wild-type and Klf4-null at E13.5, E15.5 and E18.5 in the whole cerebellum. The expression of Pax6 is ~23% of the wild-type expression level in the Klf4-null in the E13.5 ($p < 0.01$) and 15.5 ($p < 0.01$). One-tail students' T-test was used and results were represented with $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***) .

Y-axis: Relative Quantity vs H2O – target gene expression of the sample compared against with a negative control where H2O were used as template.

X-axis: EGL - external granular layer, NE – neuroepithelium, RL- Rhombic lip, WT- wild-type, Mut – Klf4-null.

5.3.4. Klf4's roles in the developing cerebellum

To investigate possible biological functions of Klf4 in the developing cerebellum, we studied important developmental events, such as cell differentiation, cell death and cell proliferation, in the Klf4^{-/-} cerebella using immunohistochemistry with cell-specific markers: Gfap and Glast for glial cells, Calbindin for Purkinje cells, Pax2 for interneurons and Pax6 for granule cells. We did not observe any differences in the differentiation of Purkinje cells or cerebellar interneurons in the Klf4-null cerebellum. We also did not observe any differences in glial cell development in the Klf4^{-/-} cerebellum (data not shown).

When we examined the wild-type and Klf4-null in Nissl stained material for gross cerebellar morphology, the size and general structure of the E13.5 and 15.5 mutant cerebellum are comparable to the wild-type. However, we observed more heterochromatic GCs in the Klf4^{-/-} compared to its wild-type litter-mates suggesting a role of Klf4 in cell death and/or cell proliferation (Figure 5.4a). TUNEL and anti-Casp-3 immunostaining were performed to assess cell apoptosis in the Klf4^{-/-} cerebellum. Few cells were undergoing apoptosis in either the wild-type or Klf4^{-/-} cerebellum during early development; and no differences in apoptosis with TUNEL and Casp3 immunostaining were seen (data not shown).

To look at cell proliferation, we used a short term (1 hour) BrdU exposure to assay cell proliferation in Klf4-null embryos. In the E13.5 Klf4^{-/-} null cerebellum, we found a lower number of proliferating cells in the EGL ($p < 0.01$) and RL ($p < 0.05$) compared to wild-type litter-mates (Figure 5.4b and 4c). Furthermore, the EGL appears to be thinner and less extended in the Klf4-null (Figure 4b; see also Figure

2a-b). The reduced granule cell proliferation at E13 in the Klf4 $-/-$ suggests a positive regulatory role of Klf4 on early granule cell proliferation. However, this proliferative effect of Klf4 in the EGL is reversed at E15.5 when more proliferating cells are found in the Klf4-null EGL and RL (Figure 5.5a and b, $p < 0.01$). In addition, we also observed a decreased number of proliferating cells in the neuroepithelium (NE) where cerebellar interneurons are born at E15.5 (Figure 5.5a and b, $p < 0.01$). This opposite effect on cell proliferation in the Klf4-null hints at a differential (either direct or indirect) regulation of Klf4 in the two cerebellar neurogenic regions.

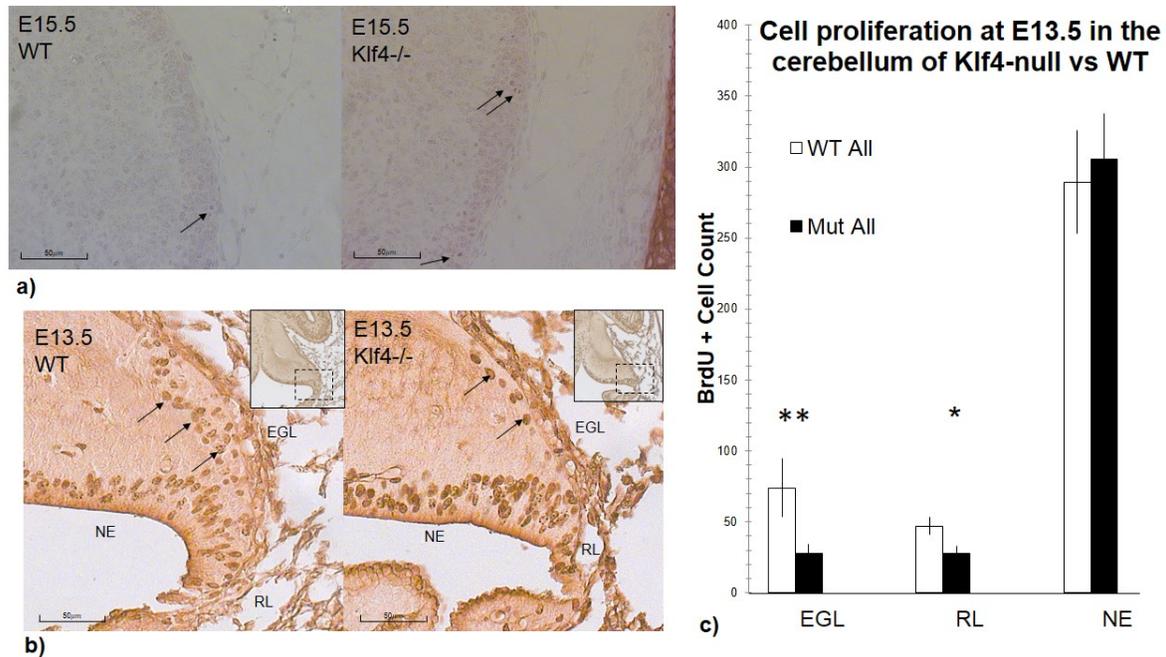


Figure 5.4. Effects of Klf4-knockout on cell death and/or cell proliferation in the developing cerebellum

a) Cresyl-violet staining of wild-type and Klf4-null cerebellum

The appearance of heterochromatic cells is a hallmark of the Klf4-null EGL compared to the wildtype at E15.5 (black arrows). The proliferating cells were identified as having condensed heterochromatin in one of the phases of mitosis..

b) and c) BrdU-staining demonstrates a reduced proliferation of EGL and RL cells in the Klf4-null at E13.5

b) Immunolabeling of BrdU in the cerebellum and c) counting of BrdU+ cells at E13.5. Proliferative cells incorporate BrdU into newly synthesized DNA and become BrdU+. There is a decreased number of BrdU+ cells in the EGL ($p < 0.01$) and RL ($p < 0.05$) of the Klf4-null cerebellum. BrdU+ cells were identified as a dark brown staining after histochemical reaction with DAB. Number of BrdU+ cells were compared with one-tail students' T-test and results were represented with $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***).

X-axis: EGL – external granular layer, RL-Rhombic lip, NE- neuroepithelium

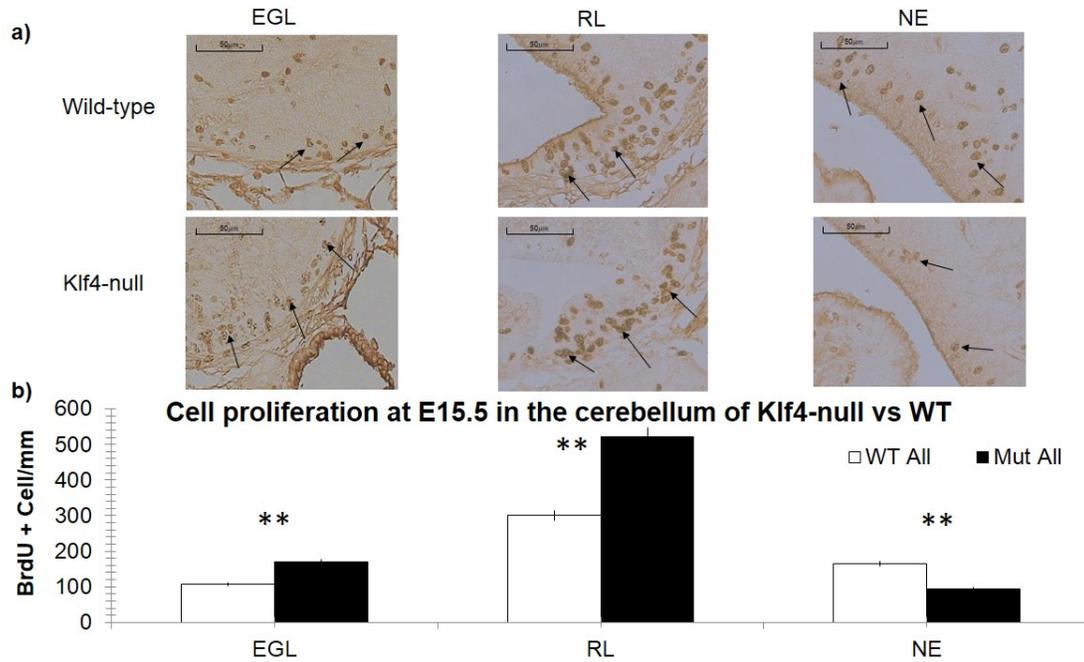


Figure 5.5. Klf4 has dual effects on the proliferation of epithelial cells in the cerebellum at E15.5

a) Immunolabeling of BrdU in the cerebellum and b) counting of BrdU+ cells at E15.5. There is an increased number of proliferating cells in the EGL ($p < 0.01$) and RL ($p < 0.01$), but a decreased number of proliferating cells in the NE ($p < 0.01$) in the Klf4-null cerebellum, indicated by a one-tailed Students' T-test $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***)..

X-axis: EGL – external granular layer, RL-Rhombic lip, NE- neuroepithelium

5.3.5. Investigation on functional redundancy in the Klf family

While the findings on Pax6 expression and granule cell proliferation were robust in the Klf4-null at E13.5, the effects of Klf4 knockout were diminished at E18.5. One explanation for this result could be that the dynamic expression of Klf4 in the cerebellum – it is expressed highest in the cerebellum at E13.5 and lowest at E18.5. Thus, the Klf4-null phenotypes may be due to expression level differences over developmental time. Another possibility for the observed temporal differences in the Klf4-null is that the proliferative roles of Klf4 in the cerebellum could be replaced by other genes and pathways by E18.5. Other members of the Kruppel-like factor family are candidates for complete or partial functional redundancy since they are structurally similar. Therefore, we investigated potential functional redundant or functional complementary genes to Klf4. Expression level of Klf2 and Klf5, two other Klf transcription factors that have overlapping functions with Klf4 in the iPS cells[193], were not altered in the Klf4^{-/-} at E13.5, E15.5 and P0 (Figure 5.6a and 5.6c). Immunohistochemistry staining at E13.5 and E15.5 also showed similar expression pattern for Klf2 (Figure 5.6b) and Klf5 (Figure 5.6d). The summarized expression data of all 17 Klf family members are shown in Table 5.1. Previous studies have shown that granule cell proliferation is regulated by at least two other molecular pathways: Zic and Wnt[147]. To examine these alternative granule cell proliferation pathways, the expression of the transcription factor Zic1 and β -catenin were measured with RT-PCR. We find that the expression level of Zic1 was normal in the Klf4^{-/-} at E13.5 and E15.5 (Figure 5.7a). Gene expression of β -catenin at E13.5, determined by RT-PCR, showed a suggestive increase in the Klf4^{-/-} when compared with wild-type, however, this increase was not significant (Figure 5.7b).

Further validation with Anti- β -catenin immunohistochemistry showed that β -catenin is enhanced at E13.5 at the rhombic lip and EGL in the mutant compared to the wildtype (Figure 5.7c). At E15.5, there is a significant increase in β -catenin expression in the Klf4-null cerebellum (Figure 5.7b, $p < 0.05$) determined by RT-PCR; however, this increased expression is not obvious with immunohistochemical staining of β -catenin at E15.5. At this time, the immunostaining of β -catenin in both Klf4^{-/-} and wild-type is much weaker compared with staining at E13.5 (Figure 5.7c).

Table 5.1. The expression of Klf family members in mouse cerebellum

| <i>Gene Name</i> | In situ Cerebellar Expression Pattern (Genepaint) | Microarray expression level (CbGRiTS, normalized and averaged) | FANTOM HeliScopeCAGE expression level (tpm, averaged) |
|------------------|---|--|---|
| <i>Klf1</i> | Not expressed | 7.388083 | 0.32837 |
| <i>Klf2</i> | Not expressed | 8.130833 | 15.45146 |
| <i>Klf3</i> | Not available | 11.62117 | 25.8526 |
| <i>Klf4</i> | Granule cells | 7.47 | 2.814696 |
| <i>Klf5</i> | Granule cells | 6.819083 | 2.02686 |
| <i>Klf6</i> | Not expressed | 7.89725 | 15.89473 |
| <i>Klf7</i> | Widespread | 13.753 | 97.85526 |
| <i>Klf8</i> | Not available | 7.331333 | 3.614032 |
| <i>Klf9</i> | Not expressed | 10.59317 | 12.99584 |
| <i>Klf10</i> | Granule cells | 6.6785 | 15.35252 |
| <i>Klf11</i> | Not available | 6.5545 | 9.870974 |
| <i>Klf12</i> | Not available | 6.567083 | 0.473217 |
| <i>Klf13</i> | Not expressed | 10.45975 | 43.46916 |
| <i>Klf14</i> | Not expressed | 6.936 | 0.05531 |
| <i>Klf15</i> | Purkinje cells | 8.219917 | 8.684977 |
| <i>Klf16</i> | Not available | 8.642083 | 13.5992 |
| <i>Klf17</i> | Not available | 6.642333 | 0.031649 |

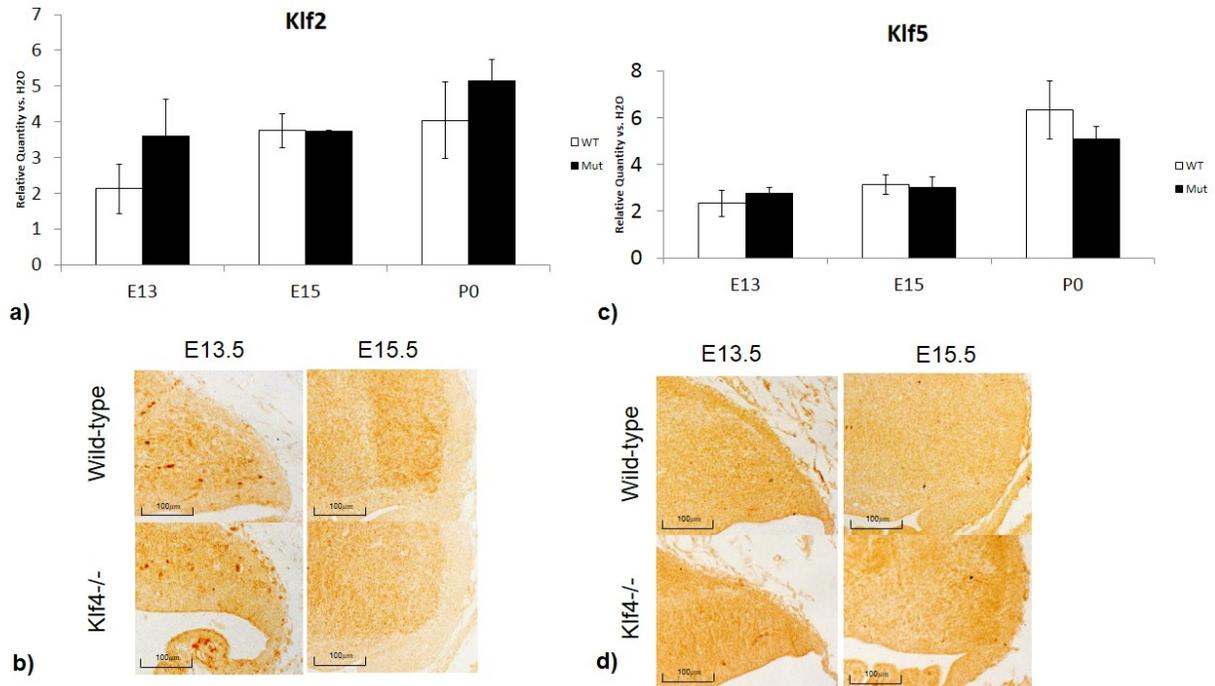


Figure 5.6. The expression levels of genes involved in complementary cell proliferative pathways in the Klf4-null with real-time PCR

a) RT-PCR and b) Immunohistochemistry showing Klf2 expression, a Kruppel-like factor belonging to the same gene family as Klf4, do not show expression changes in the Klf4-null.

c) RT-PCR and d) Immunohistochemistry showing Klf5 a Kruppel-like factor belonging to the same gene family as Klf4, do not show expression changes in the Klf4-null.

One-tail students' T-test was used for analysis and results were represented with $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***).

Y-axis: Relative Quantity vs H2O – target gene expression of the sample compared against with a negative control where H2O were used as template

X-axis: WT- wild-type, Mut – Klf4-null

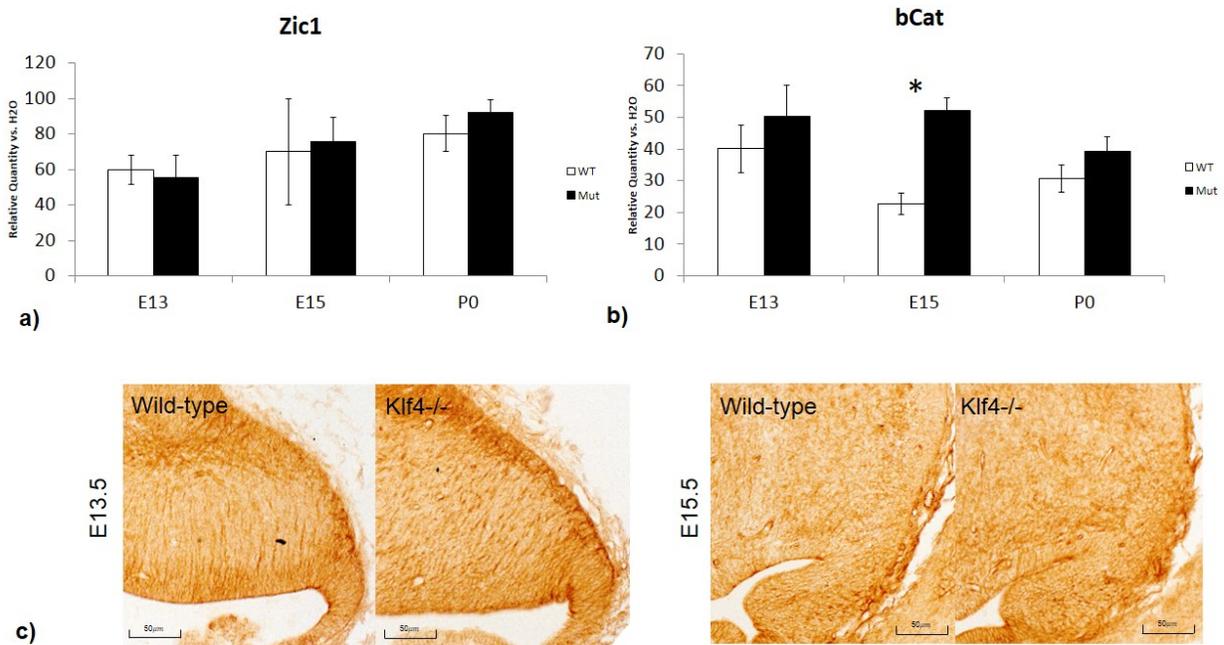


Figure 5.7. The expression levels of genes involved in alternative cell proliferative pathways in the Klf4-null with real-time PCR

a) Zic1, an early granule cell proliferation gene at E13.5, does not show expression changes in the Klf4-null.

b) β-catenin, a member of Wnt pathway, shows an activated expression in the Klf4-null ($p < 0.05$) at E15.5.

c) E13.5 and E15.5 immunohistochemistry against β-catenin, there is no difference observed at E15.5 as the stain is generally weaker than E13.5

One-tail students' T-test was used for analysis and results were represented with $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***)

Y-axis: Relative Quantity vs H2O – target gene expression of the sample compared against with a negative control where H2O were used as template

X-axis: WT- wild-type, Mut – Klf4-null

5.4. Discussion

Klf4 is an important gene in many physiological and pathological processes, such as stem cell maintenance, skin development, cellular specification in the brain, and axon outgrowth [91, 191, 194, 195]. The activation of Klf4 expression is found in immortalized kidney cells [196], laryngeal squamous cell carcinoma [196], ductal carcinoma of the breast [197] and skin carcinoma [198]. The activation of cell cycle by Klf4 could involve the repression of p53 pathway, which is a critical check point for cell cycle [199, 200].

Our current study shows that Klf4 regulates early granule cell proliferation and could positively regulate transcription factor Pax6. Similar to its positive role in promoting self-renewal of embryonic stem cells, Klf4 expression is important for granule cell proliferation at E13.5 in the cerebellum. This aligns with recent expression data using Cap-associated transcriptome sequencing that shows the highest expression of Klf4 in the cerebellum is found at embryonic day 13 [62].

Importantly, Klf4-null showed a decreased number of Pax6⁺ EGL cells as well as a decreased proliferation of these cells at E13.5 which likely resulted in the less extensive (in the caudal-to-rostral dimension) and thinner (in the dorsal to ventral dimension) EGL during early development. These data suggest that the expression of Klf4 is important to cerebellar development and generation of granule cells.

5.4.1. Klf4 regulates Pax6 expression

Our bioinformatic analysis showed that Klf4 is a potential upstream regulator of Pax6. The regulation of Pax6 by Klf4 has been demonstrated at the expression and phenotypic levels in eye development [87, 88]. The expression of Pax6 is lowered to

about half of its normal level in the cornea when Klf4 is conditionally knocked-out [9]. In addition, the defective corneal phenotype of the Klf4 knock-out resembles that of the heterozygous Pax6 knock-out [88]. Direct binding of Klf4 at Pax6 regulatory region has been found with genome-wide ChIP-seq analysis using mouse embryonic stem cells[201]. Therefore, we were interested to see if the expression of Pax6 is disrupted in the Klf4 knockout. Indeed, Pax6 expression is dysregulated in the EGL of the developing cerebellum in the Klf4-null. In the Klf4-null, the expression of Pax6 is greatly reduced at E13.5 and E15.5, indicating a positive regulation of Klf4 on Pax6. This observation is consistent with the role of Klf4 on Pax6 during corneal development where the Klf4-null showed about 50% Pax6 expression of a wild-type control[88].

However, our phenotypic data indicate that the Klf4-null granule cell is distinct from the Pax6-null granule cell; e.g., the Pax6-null phenotype is associated with deficits in neurite extension and cell migration, and a thickening of the EGL [54, 202]. None of these phenotypes are seen in the Klf4-null cerebellum. Two cerebellar phenotypes that we see in the Klf4-null, however, are not observed in the Pax6 mutant: reduced cell proliferation in the EGL [54, 87, 191] and the shorter, thinner EGL at E13.5. In summary, despite that expression of Pax6 is partially abolished in the Klf4-null, the Klf4-null phenotypes we observed were distinct from either Pax6-null or Pax6^{+/-} (which are phenotypically normal). This leads us to suggest that the phenotypes we observed in the Klf4-null cerebellum are independent of Pax6.

Finally, to examine the interplay between Klf4 and Pax6, we examined expression of Klf4 in the Pax6-null cerebellum in our CbGRiTS database [61]. We did not see a difference in Klf4 expression in the Pax6-null at E13.5, E15.5, or E18.5. This

suggests that Klf4 is upstream of Pax6 in terms of transcriptional activation and this regulation could be direct binding of Klf4 to the promoter region of Pax6 [201].

5.4.2. Klf4 as a regulator of cell proliferation

A key question in this study is what Klf4 is doing in the developing EGL during early cerebellar development. Previous studies demonstrated that Klf4 may serve either a role as a transcription activator or repressor depending on the gene targets and other co-factors; thus, it could either promote or inhibit cell proliferation under different cellular contexts. With BrdU labeling, we were able to show that in the Klf4-null, cell proliferation was up-regulated within the EGL and rhombic lip at E15.5. This suggests that Klf4 regulates a different set of gene targets at different developmental stages in the granule cells. Previous work has indicated the Wnt pathway as important to proliferation during granule cell development after E15, and Klf4 can inhibit Wnt signaling by directly interacting with β -catenin and TCF-4 [203]. Indeed, we observed an increased expression of β -catenin in the Klf4^{-/-} at E15.5 indicating an inhibitory role of Klf4 on granule cell proliferation through Wnt signaling at E15.5. The activation of the Wnt pathway by Klf4 at E15.5 could be indirect and serve as an internal “rescue” in the Klf4 null cerebellum to remedy the early loss of granule cell precursors at E13.5. The “rescue” is seen by a normal level of total and proliferating granule cells, in the E18.5 Klf4-null cerebellum. However, at E13.5 when we see an EGL proliferation deficit in the Klf4-null it is not likely to involve Wnt signaling as signaling is not activated until E15.5 [147]. A myriad of other molecular partners for Klf4 function have been identified through whole genome chromatin immune-precipitation work; identifying more than

1,800 loci in human embryonic stem cells that are directly bound by Klf4 – many of these genes, such as Oct4 and Nanog are important transcription factors in cell proliferation [201].

We also see a cell proliferation phenotype in the NE of the Klf4-null cerebellum. Interestingly, this phenotype is in the opposite direction of the EGL cell phenotype; that is the cells of the neuroepithelium (NE) demonstrated decreased cell proliferation in the E15.5 cerebellum. However, from our data, Klf4 is not measurably expressed in the NE. In any case, it is of interest that these two proliferative regions give rise to two distinct cell types based upon neurotransmitter phenotype [4], and these two cell classes have mutually exclusive neuronal markers throughout cerebellar development[27]. Thus, while the glutamatergic granule cells are generated from the rhombic lip, the GABAergic neurons of the cerebellum are generated from the NE, located at roof of the fourth ventricle, specified by the basic helix-loop-helix (bHLH) transcription factor, Ptf1a[24]. The Klf4-null showed a reduction in the number of proliferating cells in the NE at E15.5. At this time point, several types of interneurons, such as the Golgi cells and basket cells, are generated at the NE; these cells populate the molecular layer and provide an inhibitory input to PCs in the mature cerebellum[204]. The pro-proliferation effect in the NE and the anti-proliferation effect in the EGL of Klf4 observed at E15.5 could be the results of Klf4 acting in different neural progenitors; or that the proliferation of one germinal zone is secondary to the other. This cross-germinal-region effect on cell proliferation further suggests the importance and complexity of Klf4 during early cerebellar development.

In addition to cell proliferation, we also examined other important developmental processes in the Klf4-null. Klf4 has been previously identified as a key gene for cell differentiation, such as in the granular cells in the skin [191], goblet cells in the small intestine [190] and neurons in the cerebral cortex [86]. Klf4 has also been identified to affect apoptosis in various cancers, mostly due to its interaction with p53 [187]; however, the role of Klf4 in apoptosis was not reported in the developing skin, eye and intestine of the Klf4 knockout mouse [87, 190, 191]. Furthermore, Klf4 also regulates gliogenesis in the cerebral cortex by directly interacting with CBP/p300 [205]. We did not observe any changes in apoptosis and gliogenesis during cerebellar development in Klf4-null.

5.4.3. Temporal specificity of Klf-null phenotype

The phenotype of the Klf4-null cerebellum during early development were not observed at E18.5 or P0. We were intrigued to understand what might be responsible for the normal phenotype at later developmental stages in the knockout. Three possibilities emerged as candidates for altered phenotypic expression over time. First, this may be due to the temporal expression of Klf4 which is peaked at E13.5 but falls off later in embryonic development. Second, as discussed above, alternative molecular pathways could be responsible at different points in time. Third, other members of the Klf4 family could be substituting for the absence of Klf4 at later times but not earlier. Expression databases indicate that multiple members of the Klf family of transcription factors are frequently co-expressed and may have redundant functions. Functional redundancy among different Klf family members has been previously observed – e.g.,

Klf2 and Klf5 shared function with Klf4 in the stem cells so that cell cycle arrest only occurs when all three factors are knocked out[193]. We examined expression of Klf2 and Klf5 in the Klf4^{-/-}, and no significant changes in expression were found with these family members. However, there are 14 other Klf factors that we did not investigate and could share similar DNA binding sequences with Klf4. It will be interesting in future work to tease out the unique or overlapping functions of these Klfs in the granule cells during cerebellar development. Currently, we favour the second possibility as the most parsimonious explanation of the temporal specificity of the Klf4-null phenotype in cerebellar granule cells.

5.4.4. Comparison of the Klf4-null with the Pax6-null cerebellum

A key question in our analysis is whether Klf4 has a unique role, other than its interplay with Pax6 in cerebellar development. To address this question a comparison of cerebellar phenotypes between Klf4-null and Pax6-null is informative. While the structures of the Klf4-null and Pax6-null, cerebellum are apparently normal at E13.5, the Klf4-null is differentiated from the Pax6-null in that the EGL is thinner and less extended. In addition, later in development, the Klf4^{-/-} EGL shows an altered cell proliferation but normal migration and foliation, whereas Pax6^{-/-} was normal in this regard but showed aberrant differentiation, migration and foliation [202]. These differential phenotypes may be due to differences in Pax6 expression in the Klf4-null (30% of wild-type) compared to the Pax6-null (virtually 0%). In fact, the cerebellum of the Pax6^{+/-} (with 50% functional Pax6 molecules) appears normal [202]. This could suggest a dosage effects of Pax6 in cerebellar development. On the other hand, the

phenotypic differences could reflect Klf4 and Pax6's complex functions in transcriptional activation and/or suppression in the developing cerebellum.

5.4.5. Conclusions

Klf4 regulates early granule cell proliferation in a time-specific manner: at E13.5, Klf4 promotes granule cell proliferation through a pathway that is apparently independent of Zic1; whereas at E15.5, Klf4 showed an inhibitory role on granule cell proliferation, possibly through the suppression of the canonical Wnt pathway. Klf4 also positively regulates Pax6, this regulation might be direct as a Klf4 binding site has been found upstream of Pax6 in previous chip-seq studies. The next steps are to more fully elucidate the regulatory network involving Klf4 in developing cerebellar granule cells.

Chapter 6 : Discussion

6.1. Overview of objectives

In Chapter 2 of my thesis, CAGE transcriptome time courses were generated with the HeliScopeCAGE platform to enhance understanding of mammalian cerebellar development and its genetic regulatory mechanisms. Bioinformatics analyses of these time courses focused on all known transcription factors (TFs) and their potential differential expression during cerebellar development. In Chapter 3 of my thesis, utilizing the 5' end sequence information from the CAGE time course, we investigated the alternative usage of transcription start sites which could produce distinct mRNA and protein isoforms – another well-known form of regulation of gene expression and/or function, which had not previously been examined during cerebellar development. In Chapter 4 of my thesis, we have generated a granule cell-centric CAGE time course with laser capture microdissection to identify genes that are enriched in cerebellar granule neurons, and their potential function in granule cell development. In Chapter 5 of my thesis, we studied the Klf4 knock-out mouse model to explore the roles of a novel gene regulator in cerebellar development. In summary, my thesis provides a novel and comprehensive collection of potential transcriptional regulators important for the development of the cerebellum and cerebellar granule cells.

6.1.1. Utilization of cerebellar HeliScopeCAGE time series toward understanding neurological disorders: Rett Syndrome

Our cerebellar time series was incorporated into the FANTOM5 project, which currently consists of datasets of temporal transcriptomic datasets for 19 human and 14

mouse tissues[62]. We actively collaborated with the FANTOM consortium for data mining and biological validation of CAGE data. Our comprehensive cerebellar CAGE time course proved to be a valuable source of information in studies on neurological disorders, such as Rett Syndrome (RTT) as an example. RTT is a disorder caused by mutations in either Methyl CpG binding protein 2 (Mecp2), Forkhead box G1 (Foxg1) or Cyclin-dependent kinase-like 5 (Cdkl5)[206-208]. Data from FANTOM5 provided an unprecedented opportunity to identify the transcription start sites (TSSs) of these genes and study their expression profile in a wide range of mouse and human samples. The FANTOM5 data revealed the precise initiation sites for Mecp2, Foxg1 and Cdkl5 for the first time, and that each of these genes use the same TSS in most tissues throughout development. While a significant correlation between the expression levels of these three genes in the brain was not found, a genome-wide analysis uncovered common transcription factors that may coordinately regulate these three genes. The FANTOM5 CAGE dataset also allowed the location of putative enhancers regulating these three genes in humans (as described in[62]). Furthermore, using mouse ENCODE ChIP-seq data, genomic regions bearing promoter and enhancer marks were identified in the regulatory regions of these three genes. In parallel, by comparing expression profiles and chromosomal markers of disease-causing genes across different brain regions, the enrichment of the histone mark H3K27me3 was observed at the enhancer region of Foxg1[73]. Interestingly, active histone marks in the Foxg1 promoter region were absent in cerebellar tissue but present in the cerebral cortex[73]. This differential epigenetic marking with H3K27me3 (a modification known to inactivate expression when in the enhancer region of a gene) suggested a role of Polycomb Repressive Complex 2

(PRC2) (the only known complex to make the H3Kme3 modification[209]) in the silencing of Foxg1 in the cerebellum.

6.2. Construction of transcriptional network in cerebellar development

From the FANTOM5 database, we collected and examined a large temporal gene expression dataset spanning important stages of mouse cerebellar development. This was the first cerebellar developmental sequencing time series which utilized next generation sequencing technology, and thus significantly increased the depth of transcriptome analysis. The differential expression analyses in Chapter 2 revealed that development of the mouse cerebellum is programmed by thousands of different genes which are potentially regulated by more than two hundred transcription factors (TFs). However, it is difficult to construct a regulatory transcription network solely from a list of differentially expressed genes. Bioinformatic analyses are required to discover key regulators which drive the differential expression we observed through the binding of these regulators to the promoter and/or the enhancer of their downstream gene targets.

One such bioinformatics approach that could help to elucidate the regulatory roles of key TFs in cerebellar development is Motif Activity Response Analysis (MARA, FANTOM4, [210]). MARA has been used by the FANTOM consortium to search for over represented binding sequence (aka, motifs) of TFs that bind to the regulatory region of a group of differentially expressed genes with similar expression profiles. To appreciate MARA's application to the cerebellar time course dataset, we performed an initial analysis. We found 96 binding motifs with significantly altered binding activity over developmental time. In order to get a sense of the validity of these results we sought to identify TFs known to play important roles in cerebellar development among the

significantly activated binding motifs. We found a linear increase in motif activity of a Purkinje cell-specific TF, RAR-related orphan receptor alpha (Rora), that is correlated with the generation and maturation of cerebellar Purkinje cells[27, 211, 212]. As another example, Pax6, which is primarily expressed by cerebellar granule cells, showed positive motif binding activity (i.e., an upregulation of its downstream targets) during the embryonic developmental timeframe in which granule cell generation and proliferation take place[35, 58]. Interestingly, Pax6's motif activity became negative (ie, a down-regulation of Pax6's downstream targets) at post-natal time points when granule cells differentiate and mature[58, 213]. Since both the Rora and Pax6 results observed in our data matched previously-known changes in cellular processes of cerebellar neurons, our approach appeared valid. The TFs and their targets predicted from motif analysis extends our findings of differentially expressed genes and takes us a step further in building a transcriptional network in cerebellar development.

6.2.1. Role of promoter activity and enhancer activity in gene regulation during cerebellar development

As discussed above, HeliScopeCAGE datasets also reveal enhancer activities through the bidirectional expression of the eRNAs[76]. Enhancers play an important role in gene regulation as fully assembled enhancer complexes, or 'enhanceosomes,' can modify the local chromatin architecture and recruit the RNA polymerase II machinery to the promoter. The differential expression analyses we performed in Chapter 2 could not reveal the usage of promoters and/or the enhancers without further analyses.

One such bioinformatic analysis that could determine the role of promoter and enhancer activity is the Transfactivity tool[62]. To explore this analysis and determine if it could be a viable future direction, we used the Transfactivity tool to computationally predict motif binding activity in the promoter and enhancer regions in cerebellar development. Similar to MARA, the Transfactivity tool can search for over-represented TF binding motifs in the regulatory regions of differentially expressed genes with similar expression profiles over time; moreover, the Transfactivity tool utilizes an updated, higher coverage motif database (JASPAR[214]) and identifies activated motifs in the enhancer as well as the promoter regions (whereas MARA is restricted to promoter regions). We were able to identify 329 activated motifs in the promoter region and 252 activated motifs in the enhancer region of all genes. We found that Pax6 regulates its downstream targets by binding to the promoter region, and that Pax6 positively regulates its targets during embryonic development when the granule cells are generated and proliferative[55, 174]. In contrast, Pax6's negatively regulates its targets in post-natal time points when the granule cells differentiate and mature[60, 213]. Atoh1 (as known as Math1), known for its roles in granule cell specification[35, 37], was also shown to bind to the to the promoter regions of its targets, as Pax6 does. We observed an interesting transition from down-regulation of Math1-associated target expression before E15, to up-regulation from E16 to P6. Notably, this transition was the opposite pattern seen for targets of Pax6. Altogether, two granule cell specific transcription factors, Pax6 and Math1, both regulate their downstream targets by binding to the promoter region; however, their motif activity is inversely correlated throughout cerebellar development. The inverse relationship of Math1 and Pax6 suggests that they

may function in a coordinated gene regulatory pathway, where one serves as a functional suppressor of the other.

A post facto Transfactivity analysis revealed interesting motif activity binding patterns of the three genes that were used in RNAi knock-down validation experiments in Chapter 2. Activating transcription factor 4 (Atf4)'s motif activity pattern, along with its binding of the promoter region closely resemble those of Pax6. The regulatory factor x 3 (Rfx3) also shared a similar motif activity pattern to Pax6. However, unlike all genes mentioned above, Rfx3 regulates its down-stream targets through binding to their enhancer region rather than the promoter region. Lastly, scratch family transcriptional repressor 2 (Scrt2) showed dynamic motif activity at both the promoter and enhancer regions. Interestingly, the motif activity in the promoter region is the inverse of Pax6. Other than the results we present in Chapter 2, functional studies for these three genes in the cerebellum are currently lacking, and it will be important to fully elucidate the functional importance of these novel cerebellar regulators. In summary, in addition to identifying genes with the most dynamic motif activities, the Transfactivity analysis serves as an outstanding tool to discover novel TFs by searching for motif activity patterns that are correlated with known TFs important for cerebellar development (for example, Pax6 and Math1) at the promoter and/or the enhancer regions. Biological validation will be required to uncover the mechanism of how the identified genes may interact with known cerebellar TFs, such as Pax6 and Math1, predicted by the Transfactivity algorithm.

6.3. Extensive alternative promoter usage during cerebellar development

In this study, we used CAGE expression data to discover the temporal switching of TSS usage in mouse cerebellum. In the FANTOM project, CAGE has been shown to identify different TSSs and the corresponding promoters for a single gene[77-80]. The production of different isoforms due to the usage of alternative transcription start sites (TSSs), which was once considered as uncommon, has now been found in the majority of human genes[98, 215]. Previously, the majority of studies in alternative TSSs focused on their roles in tissue-specific regulation of transcription. As an example, Cyp19 is a human gene with tissue-specific expression governed by the usage of alternative promoters, and is involved in the conversion of C19 steroids to C18 estrogens[216]. Three distinct promoters drive the expression of Cyp19 in three tissue types – the placenta[217], the gonad[218], and the brain[219]. Notably, studies have suggested that the placenta-specific Cyp19 promoter is derived from a long terminal repeat of a primate endogenous retrovirus[220].

In this study, we have identified 9,767 crossover TSS switching events; with each event indicating an alteration in the dominant TSS over developmental time. Studies on the temporal usage of alternative TSSs are less common, compared with their tissue-specific counterpart, and often cover a short time span (e.g., over the period of one cell cycle, see below). For example, the alternative promoters of p18, a member of the p16–INK family of cyclin-dependent kinase inhibitors, are known to produce transcripts encoding the same protein, but differing by the presence of an additional 1.1 kb of sequence in the 5'UTR[221]. In undifferentiated myoblasts, all detectable p18 transcripts originate at the upstream promoter; whereas when differentiation begins, transcription completely shifts to the downstream promoter, resulting in a significantly shortened 5'

UTR and a 50-fold increase in the amount of p18 present in the cell. The increase in p18 in the cell is postulated to be involved in permanent arrest of the cell-cycle after terminal cell differentiation[221]. Unlike previous studies in alternative TSSs, our study focused on temporal usage of alternative TSSs in the mouse cerebellum over a comprehensive developmental time course spanning specification of the cerebellum at E11 to postnatal adulthood. Furthermore, rather than studying alternative TSSs of a single gene, our study surveyed the entire transcriptome in order to discover all alternative TSSs that may be functionally important for cerebellar development.

There are shortcomings of this approach due to technical limitations. First, a recent study reported that some genes experience transcription initiation but do not show detectable full-length transcripts[222]. This finding may be explained by post-transcriptional regulation such as rapid degradation of full-length transcripts through nonsense mediated decay, deadenylation mediated decay, or AU-rich element mediated decay processes[223]. Therefore, a switch in the dominant TSS may not always result in a change in function predicted *in silico* by bioinformatic analyses; such a conclusion must be supported with functional protein-based biological validation, such as immunohistochemistry with isoform-specific antibodies. Second, due to heterogeneity of the cerebellar tissues, it is bioinformatically difficult to account for changes of cellular composition over a developmental time course. For example, it is difficult to identify if a dominant TSS at E18 is due to increased usage of said TSS at E18, or if that observation is a reflection of a granule cell specific TSS reflecting a rapid cell proliferation of the granule cells. To firmly conclude whether a switch of TSSs is temporal or cell-specific, a cell-purification method, such as laser capture

microdissection, must be used to isolate a particular type of neuron over its developmental time.

From our cerebellar development time course, we have identified ~21% of all genes that exhibit differential TSS usage during cerebellar development. The common occurrence of alternative promoter usage in the cerebellum raises the question of how alternative promoters evolved and how they are regulated. Several pathways potentially leading to the creation of an alternative promoter include a promoter being formed *de novo* from accumulated mutations over time. As many predicted binding sites for transcription factors are relatively short, such an occurrence is not unlikely. Another possibility is a recombination event leading to duplication of an entire promoter region, followed by subsequent mutations which alter its affinity to trans-regulators or its tissue/temporal specificities. In this case, the alternative promoter regions should share some sequence similarity – the extent of which would depend on the age of the duplication. Finally, the insertion of a transposable element upstream of a gene could create a new promoter – numerous examples of this phenomenon are reviewed by Josette-Renée Landry et al. [113].

Alternative TSSs, similar to alternative splicing, can produce multiple mRNA isoforms of a single gene. It is possible in many cases that the alternative TSSs occurring during cerebellar development are the result of aberrant transcription and do not have any functional roles. There are several methods that one can use to investigate the functional importance of the alternative TSSs in cerebellar development:

- 1) Functional mRNA isoforms should be depleted of premature termination codons. Indeed, up to 35% of human mRNA isoforms contain a premature termination codon

that render them non-functional[224]. 2) Functional mRNA isoforms should be evolutionarily conserved across species. Recent meta-analysis reported that the conservation of alternative mRNA isoforms between human and mouse can be as low as 11%[225]. In non-crossover events, a lower conservation level is observed for the non-dominant isoform when compared with its dominant counter-part[226], suggesting it is more likely to be non-functional; and 3) Functional mRNA isoforms should have at least one functional domain. In human, about half of all mRNA isoforms contain a known functional component [227] and may be fully or partially functional. Overall, previous studies, as discussed above, suggest that less than half of mRNA isoforms are functional. However, not only could the non-sense mRNA transcripts play important regulatory roles, as discussed previously; they could also provide an important genetic reservoir so that functional mRNAs could arise *de novo* through the accumulation of mutations over evolutionary time.

In conclusion, the alternative usage of TSSs we discovered is the first report on the importance of alternative TSS usage during cerebellar development, and we have shown that these switching events occur at a high prevalence in the cerebellum and have potential roles in transcriptional, post-transcriptional and functional regulation in cerebellar transcriptome.

6.4. Temporally regulated and granule cell enriched transcription factors

The normal development of granule cells is the result of a precise regulation through a set of TFs or “driver” genes. Our objective was to identify which of these TFs are important for granule cell development, and to understand how these factors function during normal granule cell natural history. We used laser capture

microdissection (LCM), a technique that can isolate specific cell types of interest from discrete regions of tissue[82-84], to obtain pure populations of granule cells from 3 distinct early-stages (E13, E15 and E18) of mouse cerebellar development. This approach was used in an attempt to mitigate the expression noise from the several other cell types that reside in the cerebellum.

6.4.1. The advantage of LCM to obtain enriched EGL samples

The most important advantages of LCM are its preservation of cellular state, precision of targeted dissection, and speed of sample collection. Unlike other purification methods such as cell panning, sorting and culturing, laser captured granule cells have their morphology well preserved when directly taken directly from surround tissues with minimal environmental disruption. This ensures that the resulting transcriptome data obtained from subsequent assays closely reflect the GC's native developmental status. In addition, during early development, the progenitors of cerebellar granule cells are located in EGL - a continuous, densely populated and homogenous layer on the surface of the cerebellum[1]. LCM is very precise when coupled with cellular staining that highlights cells such as cresyl violet. Last, since thousands of cells can be collected with a single laser cut and capture, the efficient LCM process reduces the chance of DNA, RNA or protein degradation and serves as a reliable source of samples for subsequent transcriptome-wide assays. Banks et al compared the antigenicity of HSP-60 and β 2-microglobulin with Western blot and found no gross changes in protein profiles between LCM-collected and conventionally

collected tissue[228], thus supporting the excellent performance and tissue preservation capacity of LCM.

6.4.2. Limitations of LCM

Despite the advantages of LCM discussed above, there are several limitations to this approach. The limitations of LCM are its low yield, vulnerability to contamination and reduced accuracy when capturing smaller cells. First, the reduction of RNA yield from laser-captured granule cells is substantial – resulting samples often measure in the range of 20-50ng/ul, as compared to a yield of 3-5ug/ul from the unprocessed cerebellum. This reduction in yield mainly results from the isolation of a small population of cerebellar granule cell progenitors, along with degradation of RNA from freezing and thawing, staining of tissue, and the actual process of microdissection. We attempted to compensate for the low RNA yield of the LCM method by applying a revised, shorter staining process and collecting samples from a larger pool of cerebellum (~3-fold greater tissue volume than a single cerebellum). Second, the micro-dissected tissue sections were not cover-slipped to allow physical access to the tissue surface. Without a coverslip, dust or other contaminants may contact and attach to the surface of the tissue slide and adhesive capturing cap, introducing unwanted RNase or foreign DNA/RNA. To prevent contamination, the tissue slides were kept within a closed, RNAase-free container for transportation. Also, cell staining, laser capture and cell lysis were performed in rapid successive sessions to reduce the time tissue slides being vulnerable to contamination. Third, the minimum laser spot size of 7.5 μm is about two times the diameter of a granule cell. Thus, it is difficult to isolate the

tiny granule cells while excluding contaminating fragments of adjacent cells. To achieve higher EGL purity during LCM, we focused on the center of the EGL with the LCM laser so that any adjacent cells would also be granule cells; and the boundary of the EGL was carefully avoided to reduce the likelihood that non-EGL cells would be collected. Overall, easy handling, high speed, high purity, and cellular integrity make LCM a powerful tool for rapid collection of EGL samples from developing cerebellum for transcriptome analyses, and the clear choice for our approach despite the limitations.

6.4.3. Application of LCM in transcriptome studies

Using LCM combined with CAGE, we identified 125 transcription factors as potential key regulators for cerebellar granule cell development. From this gene set, we further identified 37 transcription factors that had not been previously understood in the context of cerebellar development. The coupling of LCM and gene expression assays has been established as a powerful and robust experimental approach since the development of laser capture in late 1990s [82]. In 2002, Sluka et al., used RT-PCR in LCM-purified seminiferous tubules to study transition protein-1 (TP-1) gene expression and found it to be involved in compaction of the spermatid nucleus during elongation[229]. In 2005, Trogan and Fisher used LCM to isolate foam cells from atherosclerotic arteries; the purified RNA was used for molecular analysis by real time quantitative polymerase chain reaction[230]. The power of combining LCM and transcriptome-wide assay was further demonstrated in 1999 by Luo et al., where they reported reproducible differences in gene expression between large and small neurons isolated from rat dorsal root ganglia. Because microarray studies require large amounts of high quality genomic

material, RNA needs to be amplified with T7 RNA polymerase to obtain sufficient starting material[231]. RNA amplification was not required for our study due to a large reduction of starting material required by HeliScopeCAGE compared to microarray. Elimination of the amplification step removed the artefact in RNA level caused by differential amplification efficiency. Another example of the combination of LCM and next-generation sequencing is that of Cañas et al., who recently developed a protocol coupling LCM and 454 pyrosequencing; this approach was used to analyze genetic networks among different tissues of conifers[232].

In conclusion, the combination of LCM and HeliScopeCAGE allowed us to provide an unprecedented insight into the transcription factors of cerebellar granule cells that may be important for development, and provided valuable information for further functional studies on these transcriptional regulators.

6.5. Klf4 as an important transcriptional regulator during cerebellar development

From the Klf4 mouse knock out, we found that KLF4 plays an important role in cell proliferation in the early development of the cerebellum. This role of KLF4 does not extend to late embryonic and perinatal cerebellar development, as we did not observe significant differences between the wild-type and Klf4 KO cerebellum at E18.5, consistent with the earlier report that the CNS of the Klf4 null mice appeared to develop normally[191]. Thus, Klf4 could function as an integral part of multiple molecular pathways that maintain a fine balance in cerebellar granule cell population through regulation of cell proliferation at different developmental time points. Although multiple developmental abnormalities are associated with Klf4-null in the skin[191]and ocular system[87], the human KLF4 locus has not yet been associated with any genetic

disorders. This may reflect the critical role of Klf4 in pluripotent stem cells, such that spontaneous human Klf4 mutations are lethal.

Previously, genome-wide assays have been used to elucidate Klf4 target gene profiles. In one study, Chen et al. identified Klf4 binding sites in the regulatory region of 1,840 genes with chromatin precipitation on chip assays in human ES cells[201]. Two sequence fragments were mapped to the upstream region of Pax6 when Klf4 was used as bait in ChIP, indicating Klf4's direct regulation on Pax6 by physically binding to its promoter region[201]. In another study, Swamynathan SK et al. utilized the Klf4-LoxP and Le-Cre transgene system to create a mouse with conditional Klf4 KO during ocular development [88]. They performed microarray analysis of the whole eye in the conditional Klf4 KO and wild-type control. In this work, they identified 1,269 genes with significant changes in expression. Pax6 expression was reduced to half in the conditional Klf4 KO indicating a positive regulation of Klf4 on Pax6 during ocular development[88]. Consistent with these earlier reports, our results showed that Klf4 positively regulates Pax6 in cerebellar granule cell progenitors at E13.5 and E15.5 during cerebellar development.

Our results, taken together with other reports, demonstrate that the proliferation of cerebellar granule neurons is governed by a complex set of transcription factors, such as Zic1[53, 165], Wnt1[15, 20, 55], Shh[56] and Pax6[158]. The regulation of Klf4 on granule cell proliferation during early cerebellar development could be direct (aka. through a Klf4 pathway yet to fully revealed) or indirect (aka. through regulation of other proliferation pathways). The present study shows that at E13.5 and E15.5, Klf4 activates the granule cell expression of Pax6, which has been found to play critical roles

in proliferation, migration and cell death of multiple cerebellar neurons[63, 64, 158]. Thus, the regulation of Klf4 on granule cell proliferation could be an indirect effect through pathways involving Pax6. Furthermore, a recovery effect was observed in the EGL after a reduction of GC population due to tissue damage through an activation of neurogenetic processes[233]. It is also conceivable that while the down-regulation of GC proliferation at E13.5 is a direct consequence of the absence of Klf4 in the EGL; the up-regulation of GC proliferation at E15.5 may arise as a secondary effect of recovery mechanism from low number of proliferating GCs at E13.5.

In conclusion, in Klf4 knock-out mice, we found that Klf4 regulates early granule cell proliferation in a temporally-specific manner: at E13.5, Klf4 promotes granule cell proliferation apparently through a pathway independent of Zic1, that is yet to be fully revealed; whereas at E15.5, Klf4 showed an inhibitory role on granule cell proliferation, possibly as an indirect regulatory effect through the suppression of the canonical Wnt pathway. Klf4 also positively regulates Pax6, this regulation might be direct as a Klf4 binding site has been found upstream of Pax6 in previous chip-seq studies.

6.6. Future directions in the study of cerebellar development utilizing our HeliScopeCAGE database

6.6.1. Comparative bioinformatic analyses of the FANTOM5 (Zenbu) and CbGRiTS databases

In Chapter 2, we made a first pass at comparing the differentially expressed (DE) genes ($p < .05$ and at least a 2-fold difference in expression) in cerebellar development common to the FANTOM5 and CbGRiTS datasets. When we examined one of the most dynamic developmental periods, embryonic day 12 versus 13 (E12 vs E13), we observed a large overlap in the two datasets. In particular, of the 262 shared DE genes between datasets, all of them have changed in the same direction, suggesting a high fidelity between datasets. This initial analysis showed a cross-validation between HeliScopeCAGE and microarray data. Further analyses could focus on day-to-day comparisons, as well as comparisons between time periods (for example, E11-13 as early embryonic periods, E15-E18 as late embryonic periods, P0-P9 as postnatal periods) of DE genes that identify common and dataset-specific DE genes throughout cerebellar development (E11 – P9). A high concurrence of DE genes between the two datasets will highlight their reliability for future transcriptional and functional studies in cerebellar development, such as enhancer RNA and protein interaction analyses.

6.6.2. Discovery of novel genetic elements and their regulatory interactions

HeliScopeCAGE dataset allows us to study genetic elements that were impossible to identify with the traditional expression-microarray method. At the genomic level, we could investigate the changes in epigenetic markers, such as histone modification.

Unlike the microarray platform used by CbGRiTS – which targets a random region of mRNAs, the CAGE assay used by FANTOM5 specifically captures the 5' end of mRNA sequences, which allow one to identify the promoters of mRNA as well as their associated cis-regulatory elements. When coupled with a genome-wide methylation assay[234], we could explore the temporal epigenetic changes during cerebellar development. FANTOM5 HeliScopeCAGE data also enables the identification of another cis-regulatory element – the enhancer - that is often associated with CpG-poor mRNA promoters. The enhancer is identified as a peak of bidirectional, exosome-sensitive, relatively short un-spliced RNAs (eRNA). With our cerebellar time series, we could generate an eRNA atlas that identifies temporally regulated eRNAs and proximal promoters that the enhancers activate.

At the transcriptional level, HeliScopeCAGE data can be used to identify non-coding RNAs (ncRNAs) which have a regulatory role on other mRNAs rather than being translated. A subclass of ncRNA transcripts, the long non-coding RNA (lncRNA), is encoded by highly regulated, multi-exon, transcriptional units, that are processed like typical protein-coding mRNAs and could play important roles during development[235]. We could identify candidate functional lncRNAs during cerebellar development by using a systematic computational filtering approach that parse out protein-coding mRNAs. Similarly, HeliScopeCAGE data also allows us to identify microRNAs – a type of small non-coding RNA – that inhibits the expression of their target genes. Although the general mechanisms of microRNA-based gene silencing have been revealed, little is known about their regulatory roles in the cerebellum. It is possible for us to identify

active microRNAs, their downstream regulatory targets, and their functional importance during cerebellar development

Finally, at the post-transcriptional level, we can explore genetic interactions (including regulatory ncRNA-mRNA, and protein-protein interaction) during cerebellar development with the HeliScopeCAGE data. Many bioinformatic analyses for the discovery of genetic interactions utilize parametric functions such as continuous linear, switch-like, sigmoidal and copula functions[236]. While parametric functions offer more accuracy and precision if a correct biological parameter (such as temporal or spatial information) is used[237], choosing the right parameter can be difficult due to biological complexity and variability in our HeliScopeCAGE data. Therefore, we are collaborating with the Song lab from New Mexico on the development of a nonparametric functional dependency (NPfD) method. The NPfD methodology imposes no assumptions on functional parameters[238]; therefore, it is highly desirable for its un-biased interaction prediction during cerebellar development in which a previous genetic “interactome” is largely lacking. Altogether, our main objectives with the NPfD analysis are to reveal novel and unbiased (by previous knowledge) genetic interactions, including TF-target, ncRNA-mRNA and protein-protein interactions, in cerebellar development.

6.7. Concluding remarks

There are four major pieces of work in my thesis: Chapter 2 is a study of cerebellar transcriptome using next generation HeliScopeCAGE technology; Chapter 3 is the identification of differential usage of alternative promoters in cerebellar development; Chapter 4 details the discovery of granule cell progenitor enriched genes and Chapter 5 investigates the functional importance of the transcription factor – Klf4.

Together, these four projects represent a balance between the application of genome-wide assays (Ch 2, 3 and 4) and single gene study (Ch 5); a balance between development of heterogenous tissue (Ch 2 and 3) and a single cell type (Ch 4 and 5); as well as a balance between the use of bioinformatics (Ch 2, 3 and 4) and molecular genetics (Ch 4 and 5). We have made important discoveries in the analysis of cerebellar and granule cell transcriptome including identifying key transcriptional regulators and their promoter usages. We hope this work will spark interest from developmental neuroscientists around the world and provide them with fundamental knowledge and powerful tools for their research in the CNS.

References

1. Goldowitz, D. and K. Hamre, *The cells and molecules that make a cerebellum*. Trends in Neurosciences, 1998. **21**(9): p. 375-382.
2. Akshoomoff, N.A. and E. Courchesne, *A new role for the cerebellum in cognitive operations*. Behavioral neuroscience, 1992. **106**(5): p. 731.
3. Courchesne, E., et al., *Impairment in shifting attention in autistic and cerebellar patients*. Behavioral neuroscience, 1994. **108**(5): p. 848.
4. Fiez, J.A., *Cerebellar contributions to cognition*. Neuron, 1996. **16**(1): p. 13-15.
5. DOW, R.S., *The evolution and anatomy of the cerebellum*. Biological reviews, 1942. **17**(3): p. 179-220.
6. Joubert, M., et al., *Familial agenesis of the cerebellar vermis A syndrome of episodic hyperpnea, abnormal eye movements, ataxia, and retardation*. Neurology, 1969. **19**(9): p. 813-813.
7. Klein, O., et al., *Dandy-Walker malformation: prenatal diagnosis and prognosis*. Child's Nervous System, 2003. **19**(7-8): p. 484-489.
8. Gleeson, J.G., et al., *Molar tooth sign of the midbrain–hindbrain junction: occurrence in multiple distinct syndromes*. American journal of medical genetics Part A, 2004. **125**(2): p. 125-134.
9. Barth, A., *The infrared absorption of amino acid side chains*. Progress in biophysics and molecular biology, 2000. **74**(3): p. 141-173.
10. His, W., *Histogenese und Zusammenhang der Nervelemente*. 1890: publisher not identified.
11. Shepherd, G.M., *Foundations of the neuron doctrine*. 1991: Oxford Univ Press.
12. Hamburger, V., *Historical landmarks in neurogenesis*. Trends in Neurosciences, 1981. **4**: p. 151-155.
13. Gramsbergen, A., *Motor Development*, in *Current Issues in Developmental Psychology*. 1999, Springer. p. 75-106.
14. Wingate, R. and M.E. Hatten, *The role of the rhombic lip in avian cerebellum development*. Development, 1999. **126**(20): p. 4395-4404.
15. McMahon, A.P. and A. Bradley, *The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain*. Cell, 1990. **62**(6): p. 1073-1085.
16. Chi, E.Y., et al., *Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation*. Pharmaceutical research, 2003. **20**(9): p. 1325-1336.
17. Crossley, P.H., S. Martinez, and G.R. Martin, *Midbrain development induced by FGF8 in the chick embryo*. 1996.
18. Martinez, A., et al., *Involvement of striate and extrastriate visual cortical areas in spatial attention*. Nature neuroscience, 1999. **2**(4): p. 364-369.
19. Joyner, A.L., W.C. Skarnes, and J. Rossant, *Production of a mutation in mouse En-2 gene by homologous recombination in embryonic stem cells*. 1989.
20. Joyner, A.L., *Engrailed, Wnt and Pax genes regulate midbrain-hindbrain development*. Trends in Genetics, 1996. **12**(1): p. 15-20.
21. Diño, M.R., A.A. Perachio, and E. Mugnaini, *Cerebellar unipolar brush cells are targets of primary vestibular afferents: an experimental study in the gerbil*. Experimental brain research, 2001. **140**(2): p. 162-170.
22. Lainé, J. and H. Axelrad, *Extending the cerebellar Lugaro cell class*. Neuroscience, 2002. **115**(2): p. 363-374.
23. Chan-Palay, V., S. Palay, and S. Billings-Gagliardi, *Meynert cells in the primate visual cortex*. Journal of neurocytology, 1974. **3**(5): p. 631-658.
24. Hoshino, M., et al., *Ptf1a, a bHLH transcriptional gene, defines GABAergic neuronal fates in cerebellum*. Neuron, 2005. **47**(2): p. 201-213.
25. Pascual, M., et al., *Cerebellar GABAergic progenitors adopt an external granule cell-like phenotype in the absence of Ptf1a transcription factor expression*. Proceedings of the National Academy of Sciences, 2007. **104**(12): p. 5193-5198.
26. Aldinger, K.A. and G.E. Elsen, *Ptf1a is a molecular determinant for both glutamatergic and GABAergic neurons in the hindbrain*. The Journal of Neuroscience, 2008. **28**(2): p. 338-339.
27. Seto, Y., et al., *Temporal identity transition from Purkinje cell progenitors to GABAergic interneuron progenitors in the cerebellum*. Nature communications, 2014. **5**.
28. Morales, D. and M.E. Hatten, *Molecular markers of neuronal progenitors in the embryonic cerebellar anlage*.

- The Journal of neuroscience, 2006. **26**(47): p. 12226-12236.
29. Alcaraz, W.A., et al., *Zfp423 controls proliferation and differentiation of neural precursors in cerebellar vermis formation*. Proceedings of the National Academy of Sciences, 2006. **103**(51): p. 19424-19429.
 30. Koirala, S. and G. Corfas, *Identification of novel glial genes by single-cell transcriptional profiling of Bergmann glial cells from mouse cerebellum*. PloS one, 2010. **5**(2): p. e9198.
 31. Yamada, K. and M. Watanabe, *Cytodifferentiation of Bergmann glia and its relationship with Purkinje cells*. Anatomical science international, 2002. **77**(2): p. 94-108.
 32. Rakic, P., *Principles of neural cell migration*. Experientia, 1990. **46**(9): p. 882-891.
 33. Yamada, K., et al., *Dynamic transformation of Bergmann glial fibers proceeds in correlation with dendritic outgrowth and synapse formation of cerebellar Purkinje cells*. Journal of Comparative Neurology, 2000. **418**(1): p. 106-120.
 34. Iino, M., et al., *Glia-synapse interaction through Ca²⁺-permeable AMPA receptors in Bergmann glia*. Science, 2001. **292**(5518): p. 926-929.
 35. Ben-Arie, N., et al., *Math1 is essential for genesis of cerebellar granule neurons*. Nature, 1997. **390**(6656): p. 169-171.
 36. Englund, C., et al., *Unipolar brush cells of the cerebellum are produced in the rhombic lip and migrate through developing white matter*. The Journal of neuroscience, 2006. **26**(36): p. 9184-9195.
 37. Wang, V.Y., M.F. Rose, and H.Y. Zoghbi, *Math1 expression redefines the rhombic lip derivatives and reveals novel lineages within the brainstem and cerebellum*. Neuron, 2005. **48**(1): p. 31-43.
 38. Alder, J., N.K. Cho, and M.E. Hatten, *Embryonic precursor cells from the rhombic lip are specified to a cerebellar granule neuron identity*. Neuron, 1996. **17**(3): p. 389-399.
 39. Klein, C., et al., *Cerebellum-and forebrain-derived stem cells possess intrinsic regional character*. Development, 2005. **132**(20): p. 4497-4508.
 40. Aruga, J., *The role of Zic genes in neural development*. Molecular and Cellular Neuroscience, 2004. **26**(2): p. 205-221.
 41. Dymecki, S.M. and H. Tomasiwicz, *Using Flp-Recombinase to Characterize Expansion of Wnt1-Expressing Neural Progenitors in the Mouse*. Developmental biology, 1998. **201**(1): p. 57-65.
 42. Machold, R. and G. Fishell, *Math1 is expressed in temporally discrete pools of cerebellar rhombic-lip neural progenitors*. Neuron, 2005. **48**(1): p. 17-24.
 43. Wingate, R.J., *The rhombic lip and early cerebellar development*. Current opinion in neurobiology, 2001. **11**(1): p. 82-88.
 44. Sekerkova, G., E. Ilijic, and E. Mugnaini, *Time of origin of unipolar brush cells in the rat cerebellum as observed by prenatal bromodeoxyuridine labeling*. Neuroscience, 2004. **127**(4): p. 845-858.
 45. Zhang, L. and J.E. Goldman, *Generation of cerebellar interneurons from dividing progenitors in white matter*. Neuron, 1996. **16**(1): p. 47-54.
 46. Maricich, S.M. and K. Herrup, *Pax-2 expression defines a subset of GABAergic interneurons and their precursors in the developing murine cerebellum*. Journal of neurobiology, 1999. **41**(2): p. 281-294.
 47. Morin, F., M. Diño, and E. Mugnaini, *Postnatal differentiation of unipolar brush cells and mossy fiber-unipolar brush cell synapses in rat cerebellum*. Neuroscience, 2001. **104**(4): p. 1127-1139.
 48. Kalinichenko, S. and V. Okhotin, *Unipolar brush cells—a new type of excitatory interneuron in the cerebellar cortex and cochlear nuclei of the brainstem*. Neuroscience and behavioral physiology, 2005. **35**(1): p. 21-36.
 49. Marzban, H., et al., *Cellular commitment in the developing cerebellum*. Frontiers in cellular neuroscience, 2015. **8**: p. 450.
 50. Miale, I.L. and R.L. Sidman, *An autoradiographic analysis of histogenesis in the mouse cerebellum*. Experimental neurology, 1961. **4**(4): p. 277-296.
 51. Alder, J., et al., *Generation of cerebellar granule neurons in vivo by transplantation of BMP-treated neural progenitor cells*. Nature neuroscience, 1999. **2**(6): p. 535-540.
 52. Knoepfler, P.S., P.F. Cheng, and R.N. Eisenman, *N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation*. Genes & development, 2002. **16**(20): p. 2699-2712.
 53. Ebert, P.J., et al., *Zic1 represses Math1 expression via interactions with the Math1 enhancer and modulation of Math1 autoregulation*. Development, 2003. **130**(9): p. 1949-1959.
 54. Engelkamp, D., et al., *Role of Pax6 in development of the cerebellar system*. Development, 1999. **126**(16): p.

- 3585.
55. Pei, Y., et al., *WNT signaling increases proliferation and impairs differentiation of stem cells in the developing cerebellum*. *Development*, 2012. **139**(10): p. 1724-1733.
 56. Wechsler-Reya, R.J. and M.P. Scott, *Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog*. *Neuron*, 1999. **22**(1): p. 103-114.
 57. Komuro, H., et al., *Mode and tempo of tangential cell migration in the cerebellar external granular layer*. *The Journal of neuroscience*, 2001. **21**(2): p. 527-540.
 58. Yamasaki, T., et al., *Pax6 regulates granule cell polarization during parallel fiber formation in the developing cerebellum*. *Development*, 2001. **128**(16): p. 3133-3144.
 59. Roussel, M.F. and M.E. Hatten, *Cerebellum: development and medulloblastoma*. *Current topics in developmental biology*, 2011. **94**: p. 235.
 60. Sato, A., et al., *Cerebellar development transcriptome database (CDT-DB): profiling of spatio-temporal gene expression during the postnatal development of mouse cerebellum*. *Neural Networks*, 2008. **21**(8): p. 1056-1069.
 61. Ha, T., et al., *CbGRiTS: Cerebellar gene regulation in time and space*. *Developmental biology*, 2015. **397**(1): p. 18-30.
 62. Arner, E., et al., *Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells*. *Science*, 2015. **347**(6225): p. 1010-1014.
 63. Yeung, J., et al., *Wls provides a new compartmental view of the rhombic lip in mouse cerebellar development*. *The Journal of Neuroscience*, 2014. **34**(37): p. 12527-12537.
 64. Yeung, J., et al., *A Novel and Multivalent Role of Pax6 in Cerebellar Development*. *Journal of Neuroscience*, 2016. **36**(35): p. 9057-9069.
 65. Grunstein, M. and D.S. Hogness, *Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene*. *Proceedings of the National Academy of Sciences*, 1975. **72**(10): p. 3961-3965.
 66. Bollen, A., *BioConductor: Microarray versus Next-Generation Sequencing toolsets*. 2014.
 67. Ronaghi, M., M. Uhlén, and P. Nyren, *A sequencing method based on real-time pyrophosphate*. *Science*, 1998. **281**(5375): p. 363.
 68. Drmanac, R., et al., *Novel nucleic acid sequences obtained from various cDNA libraries*. 2001, Google Patents.
 69. Valouev, A., et al., *A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning*. *Genome research*, 2008. **18**(7): p. 1051-1063.
 70. Thompson, J.F. and K.E. Steinmann, *Single molecule sequencing with a HeliScope genetic analysis system*. *Current Protocols in Molecular Biology*, 2010: p. 7.10. 1-7.10. 14.
 71. Consortium, T.F., *A promoter-level mammalian expression atlas*. *Nature*, 2014. **507**(7493): p. 462-470.
 72. Carninci, P., et al., *High-efficiency full-length cDNA cloning by biotinylated CAP trapper*. *Genomics*, 1996. **37**(3): p. 327-336.
 73. Vitezic, M., et al., *CAGE-defined promoter regions of the genes implicated in Rett Syndrome*. *BMC genomics*, 2014. **15**(1): p. 1177.
 74. Kawaji, H., et al., *The FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation*. *Genome biology*, 2009. **10**(4): p. 1.
 75. Li, H., et al., *Identification of gene expression patterns using planned linear contrasts*. *BMC bioinformatics*, 2006. **7**(1): p. 245.
 76. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues*. *Nature*, 2014. **507**(7493): p. 455-461.
 77. Birney, E., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project*. *Nature*, 2007. **447**(7146): p. 799-816.
 78. Carninci, P., et al., *The transcriptional landscape of the mammalian genome*. *Science*, 2005. **309**(5740): p. 1559-1563.
 79. Tsuchihara, K., et al., *Massive transcriptional start site analysis of human genes in hypoxia cells*. *Nucleic acids research*, 2009. **37**(7): p. 2249-2263.
 80. Okazaki, Y., et al., *Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs*. *Nature*, 2002. **420**(6915): p. 563-573.
 81. Dimont, E., et al., *CAGExploreR: an R package for the analysis and visualization of promoter dynamics across multiple experiments*. *Bioinformatics*, 2014: p. btu125.

82. Emmert-Buck, M.R., et al., *Laser capture microdissection*. *Science*, 1996. **274**(5289): p. 998.
83. Espina, V., et al., *Laser-capture microdissection*. *Nature protocols*, 2006. **1**(2): p. 586-603.
84. Bonner, R.F., et al., *Laser capture microdissection: molecular analysis of tissue*. *Science (New York, NY)*, 1997. **278**(5342): p. 1481.
85. Palin, K., J. Taipale, and E. Ukkonen, *Locating potential enhancer elements by comparative genomics using the EEL software*. *NATURE PROTOCOLS-ELECTRONIC EDITION-*, 2006. **1**(1): p. 368.
86. Qin, S. and C.-L. Zhang, *Role of Krüppel-like factor 4 in neurogenesis and radial neuronal migration in the developing cerebral cortex*. *Molecular and cellular biology*, 2012. **32**(21): p. 4297-4305.
87. Swamynathan, S.K., et al., *Conditional deletion of the mouse Klf4 gene results in corneal epithelial fragility, stromal edema, and loss of conjunctival goblet cells*. *Molecular and cellular biology*, 2007. **27**(1): p. 182.
88. Swamynathan, S.K., J. Davis, and J. Piatigorsky, *Identification of candidate Klf4 target genes reveals the molecular basis of the diverse regulatory roles of Klf4 in the mouse cornea*. *Investigative ophthalmology & visual science*, 2008. **49**(8): p. 3360-3370.
89. Yamanaka, S., *Induction of pluripotent stem cells from mouse fibroblasts by four transcription factors*. *Cell proliferation*, 2008. **41**: p. 51-56.
90. Evans, P.M., et al., *KLF4 Interacts with {beta}-Catenin/TCF4 and Blocks p300/CBP Recruitment by {beta}-Catenin*. *Molecular and cellular biology*. **30**(2): p. 372.
91. Qin, S., et al., *Dysregulation of Kruppel-like factor 4 during brain development leads to hydrocephalus in mice*. *Proceedings of the National Academy of Sciences*, 2011. **108**(52): p. 21117-21121.
92. Nakahara, Y., et al., *Genetic and epigenetic inactivation of Kruppel-like factor 4 in medulloblastoma*. *Neoplasia (New York, NY)*, 2010. **12**(1): p. 20.
93. Rio, D.C., et al., *Purification of RNA using TRIzol (TRI reagent)*. *Cold Spring Harbor Protocols*, 2010. **2010**(6): p. pdb. prot5439.
94. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. *Bioinformatics*, 2010. **26**(1): p. 139-140.
95. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool*. *BMC bioinformatics*, 2013. **14**(1): p. 128.
96. Tong, Y., et al., *Spatial and temporal requirements for huntingtin (Htt) in neuronal migration and survival during brain development*. *The Journal of Neuroscience*, 2011. **31**(41): p. 14794-14799.
97. Schmutz, J., et al., *Quality assessment of the human genome sequence*. *Nature*, 2004. **429**(6990): p. 365-368.
98. Davuluri, R.V., et al., *The functional consequences of alternative promoter use in mammalian genomes*. *Trends in Genetics*, 2008. **24**(4): p. 167-177.
99. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome*. *Nature*, 2012. **489**(7414): p. 57-74.
100. Cheng, C., et al., *Understanding transcriptional regulation by integrative analysis of transcription factor binding data*. *Genome research*, 2012. **22**(9): p. 1658-1667.
101. Rojas-Duran, M.F. and W.V. Gilbert, *Alternative transcription start site selection leads to large differences in translation activity in yeast*. *Rna*, 2012. **18**(12): p. 2299-2305.
102. Cooper, S.J., et al., *Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome*. *Genome research*, 2006. **16**(1): p. 1-10.
103. Shiraki, T., et al., *Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage*. *Proceedings of the National Academy of Sciences*, 2003. **100**(26): p. 15776-15781.
104. Silvapulle, M.J., *Tests against qualitative interaction: Exact critical values and robust tests*. *Biometrics*, 2001: p. 1157-1165.
105. Visel, A., C. Thaller, and G. Eichele, *GenePaint. org: an atlas of gene expression patterns in the mouse embryo*. *Nucleic acids research*, 2004. **32**(suppl 1): p. D552-D556.
106. Hawrylycz, M.J., et al., *An anatomically comprehensive atlas of the adult human brain transcriptome*. *Nature*, 2012. **489**(7416): p. 391-399.
107. Thierry-Mieg, D. and J. Thierry-Mieg, *AceView: a comprehensive cDNA-supported gene and transcripts annotation*. *Genome biology*, 2006. **7**(1): p. 1.
108. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations*. *Nucleic acids research*,

2015. **43**(D1): p. D512-D520.
109. Huang, D.W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nature protocols, 2009. **4**(1): p. 44-57.
110. Baek, D., et al., *Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters*. Genome research, 2007. **17**(2): p. 145-155.
111. Sun, H., et al., *MPromDb: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-chip experimental data*. Nucleic acids research, 2006. **34**(suppl 1): p. D98-D103.
112. Takeda, J.-i., et al., *H-DBAS: alternative splicing database of completely sequenced and manually annotated full-length cDNAs based on H-Invitational*. Nucleic acids research, 2007. **35**(suppl 1): p. D104-D109.
113. Landry, J.-R., D.L. Mager, and B.T. Wilhelm, *Complex controls: the role of alternative promoters in mammalian genomes*. TRENDS in Genetics, 2003. **19**(11): p. 640-648.
114. Kimura, K., et al., *Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes*. Genome research, 2006. **16**(1): p. 55-65.
115. Tan, J.S., N. Mohandas, and J.G. Conboy, *High frequency of alternative first exons in erythroid genes suggests a critical role in regulating gene function*. Blood, 2006. **107**(6): p. 2557-2561.
116. Kim, T.H., et al., *A high-resolution map of active promoters in the human genome*. Nature, 2005. **436**(7052): p. 876-880.
117. Murray-Zmijewski, F., D. Lane, and J. Bourdon, *p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress*. Cell Death & Differentiation, 2006. **13**(6): p. 962-972.
118. Hu, J.-m., et al., *Functional analyses of albumin expression in a series of hepatocyte cell lines and in primary hepatocytes*. Cell growth and differentiation, 1992. **3**: p. 577-577.
119. Simarro, M., et al., *Fas-activated serine/threonine phosphoprotein (FAST) is a regulator of alternative splicing*. Proceedings of the National Academy of Sciences, 2007. **104**(27): p. 11370-11375.
120. Schaefer, A.W., N. Kabir, and P. Forscher, *Filopodia and actin arcs guide the assembly and transport of two populations of microtubules with unique dynamic parameters in neuronal growth cones*. The Journal of cell biology, 2002. **158**(1): p. 139-152.
121. Katyal, S. and R. Godbout, *Alternative splicing modulates Disabled-1 (Dab1) function in the developing chick retina*. The EMBO journal, 2004. **23**(8): p. 1878-1888.
122. Rice, D.S. and T. Curran, *Role of the reelin signaling pathway in central nervous system development*. Annual review of neuroscience, 2001. **24**(1): p. 1005-1039.
123. Qiao, S., et al., *Dab2IP GTPase activating protein regulates dendrite development and synapse number in cerebellum*. PloS one, 2013. **8**(1): p. e53635.
124. Arcondéguy, T., et al., *VEGF-A mRNA processing, stability and translation: a paradigm for intricate regulation of gene expression at the post-transcriptional level*. Nucleic acids research, 2013. **41**(17): p. 7997-8010.
125. Wang, S., et al., *The expression and distributions of ANP32A in the developing brain*. BioMed research international, 2015. **2015**.
126. Sánchez, I., et al., *A novel function of Ataxin-1 in the modulation of PP2A activity is dysregulated in the spinocerebellar ataxia type 1*. Human molecular genetics, 2013: p. ddt197.
127. Kadota, S. and K. Nagata, *pp32, an INHAT component, is a transcription machinery recruiter for maximal induction of IFN-stimulated genes*. J Cell Sci, 2011. **124**(6): p. 892-899.
128. Chen, S., et al., *PP2A 1 affects Tau phosphorylation via association with the catalytic subunit of protein phosphatase 2A*. Journal of Biological Chemistry, 2008. **283**(16): p. 10513-10521.
129. Veugelers, M., et al., *Glypican-6, a new member of the glypican family of cell surface heparan sulfate proteoglycans*. Journal of Biological Chemistry, 1999. **274**(38): p. 26968-26977.
130. Bassett, J., et al., *Thyroid hormone regulates heparan sulfate proteoglycan expression in the growth plate*. Endocrinology, 2006. **147**(1): p. 295-305.
131. Campos-Xavier, A.B., et al., *Mutations in the heparan-sulfate proteoglycan glypican 6 (GPC6) impair endochondral ossification and cause recessive omodysplasia*. The American Journal of Human Genetics, 2009. **84**(6): p. 760-770.
132. Allen, N.J., et al., *Astrocyte glypicans 4 and 6 promote formation of excitatory synapses via GluA1 AMPA receptors*. Nature, 2012. **486**(7403): p. 410-414.
133. Bergers, G., et al., *Alternative promoter usage of the Fos-responsive gene Fit-1 generates mRNA isoforms*

- coding for either secreted or membrane-bound proteins related to the IL-1 receptor.* The EMBO Journal, 1994. **13**(5): p. 1176.
134. Poliak, S., et al., *Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K+ channels.* Neuron, 1999. **24**(4): p. 1037-1047.
 135. Anczuków, O., et al., *Does the nonsense-mediated mRNA decay mechanism prevent the synthesis of truncated BRCA1, CHK2, and p53 proteins?* Human mutation, 2008. **29**(1): p. 65-73.
 136. Darieva, Z., et al., *A competitive transcription factor binding mechanism determines the timing of late cell cycle-dependent gene expression.* Molecular cell, 2010. **38**(1): p. 29-40.
 137. Alarcón, M., et al., *Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene.* The American Journal of Human Genetics, 2008. **82**(1): p. 150-159.
 138. Arking, D.E., et al., *A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism.* The American Journal of Human Genetics, 2008. **82**(1): p. 160-164.
 139. Bakkaloglu, B., et al., *Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders.* The American Journal of Human Genetics, 2008. **82**(1): p. 165-173.
 140. Poliak, S., et al., *Juxtaparanodal clustering of Shaker-like K+ channels in myelinated axons depends on Caspr2 and TAG-1.* The Journal of cell biology, 2003. **162**(6): p. 1149-1160.
 141. Ellegood, J., et al., *Clustering autism: using neuroanatomical differences in 26 mouse models to gain insight into the heterogeneity.* Molecular psychiatry, 2015. **20**(1): p. 118-125.
 142. Kloth, A.D., et al., *Cerebellar associative sensory learning defects in five mouse autism models.* Elife, 2015. **4**: p. e06085.
 143. Leiner, H.C., A.L. Leiner, and R.S. Dow, *Cognitive and language functions of the human cerebellum.* Trends in neurosciences, 1993. **16**(11): p. 444-447.
 144. Altman, J. and S. Bayer, *Development of the Cerebellar System in Relation to its Evolution.* Structure, and Functions. CRC, New York, 1997.
 145. Wechsler-Reya, R.J., *Analysis of gene expression in the normal and malignant cerebellum.* Recent progress in hormone research, 2003. **58**: p. 227-248.
 146. Alcantara, S., et al., *Netrin 1 acts as an attractive or as a repulsive cue for distinct migrating neurons during the development of the cerebellar system.* Development, 2000. **127**(7): p. 1359-1372.
 147. Patapoutian, A. and L.F. Reichardt, *Roles of Wnt proteins in neural development and maintenance.* Current opinion in neurobiology, 2000. **10**(3): p. 392-399.
 148. Fishell, G. and M. Hatten, *Astroctactin provides a receptor system for CNS neuronal migration.* Development, 1991. **113**(3): p. 755-765.
 149. Komuro, H. and P. Rakic, *Distinct modes of neuronal migration in different domains of developing cerebellar cortex.* The Journal of neuroscience, 1998. **18**(4): p. 1478-1490.
 150. PORTERFIELD, S.P. and C.E. HENDRICH, *The role of thyroid hormones in prenatal and neonatal neurological development—current perspectives.* Endocrine Reviews, 1993. **14**(1): p. 94-106.
 151. Kanamori-Katayama, M., et al., *Unamplified cap analysis of gene expression on a single-molecule sequencer.* Genome research, 2011. **21**(7): p. 1150-1159.
 152. Arnold, P., et al., *MotEvo: integrated Bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of DNA sequences.* Bioinformatics, 2012. **28**(4): p. 487-494.
 153. Arner, E., et al., *Adipose tissue microRNAs as regulators of CCL2 production in human obesity.* Diabetes, 2012. **61**(8): p. 1986-1993.
 154. Akazawa, C., et al., *A mammalian helix-loop-helix factor structurally related to the product of Drosophila proneural gene atonal is a positive transcriptional regulator expressed in the developing nervous system.* Journal of Biological Chemistry, 1995. **270**(15): p. 8730.
 155. Topka, S., et al., *The transcription factor Cux1 in cerebellar granule cell development and medulloblastoma pathogenesis.* The Cerebellum, 2014. **13**(6): p. 698-712.
 156. Frank, C.L., et al., *Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum.* Nature neuroscience, 2015.
 157. Green, M.J., et al., *Independently specified Atoh1 domains define novel developmental compartments in rhombomere 1.* Development, 2014. **141**(2): p. 389-398.
 158. Swanson, D.J. and D. Goldowitz, *Experimental Sey mouse chimeras reveal the developmental deficiencies of Pax6-null granule cells in the postnatal cerebellum.* Developmental biology, 2011. **351**(1): p. 1-12.

159. Huang, D.W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nature protocols, 2008. **4**(1): p. 44-57.
160. Sunmonu, N.A., L. Chen, and J.Y. Li, *Misexpression of Gbx2 throughout the mesencephalon by a conditional gain-of-function transgene leads to deletion of the midbrain and cerebellum in mice*. genesis, 2009. **47**(10): p. 667-673.
161. Yuan, Z., et al., *Opposing roles for E2F1 in survival and death of cerebellar granule neurons*. Neuroscience letters, 2011. **499**(3): p. 164-169.
162. Korhonen, P., et al., *Glutamate-induced changes in the DNA-binding complexes of transcription factor YY1 in cultured hippocampal and cerebellar granule cells*. Molecular brain research, 1997. **52**(2): p. 330-333.
163. Korhonen, P., et al., *Changes in DNA binding pattern of transcription factor YY1 in neuronal degeneration*. Neuroscience letters, 2005. **377**(2): p. 121-124.
164. Wang, V.Y. and H.Y. Zoghbi, *Genetic regulation of cerebellar development*. Nature Reviews Neuroscience, 2001. **2**(7): p. 484-491.
165. Aruga, J., et al., *Mouse Zic1 is involved in cerebellar development*. The Journal of neuroscience, 1998. **18**(1): p. 284-293.
166. Rubenstein, J. and P. Rakic, *Cellular Migration and Formation of Neuronal Connections: Comprehensive Developmental Neuroscience*. Vol. 2. 2013: Academic Press.
167. Cheng, C.W., et al., *Zebrafish homologue irx1a is required for the differentiation of serotonergic neurons*. Developmental Dynamics, 2007. **236**(9): p. 2661-2667.
168. Becker, M.-B., et al., *Ir1 and Ir2 expression in early lung development*. Mechanisms of development, 2001. **106**(1): p. 155-158.
169. Bosse, A., et al., *Identification of the vertebrate Iroquois homeobox gene family with overlapping expression during early development of the nervous system*. Mechanisms of development, 1997. **69**(1): p. 169-181.
170. Christoffels, V.M., et al., *Patterning the embryonic heart: identification of five mouse Iroquois homeobox genes in the developing heart*. Developmental biology, 2000. **224**(2): p. 263-274.
171. Díaz-Hernández, M.E., et al., *Ir1 and Ir2 Are Coordinately Expressed and Regulated by Retinoic Acid, TGFβ and FGF Signaling during Chick Hindlimb Development*. 2013.
172. Choy, S.W., et al., *A cascade of irx1a and irx2a controls shh expression during retinogenesis*. Developmental Dynamics, 2010. **239**(12): p. 3204-3214.
173. Zhang, D., et al., *CIZ1 promoted the growth and migration of gallbladder cancer cells*. Tumor Biology, 2015. **36**(4): p. 2583-2591.
174. Yin, J., et al., *CIZ1 regulates the proliferation, cycle distribution and colony formation of RKO human colorectal cancer cells*. Molecular medicine reports, 2013. **8**(6): p. 1630-1634.
175. Higgins, G., et al., *Variant Ciz1 is a circulating biomarker for early-stage lung cancer*. Proceedings of the National Academy of Sciences, 2012. **109**(45): p. E3128-E3135.
176. Lan, M.S. and M.B. Breslin, *Structure, expression, and biological function of INSM1 transcription factor in neuroendocrine differentiation*. The FASEB Journal, 2009. **23**(7): p. 2024-2033.
177. Bae, S., et al., *The bHLH gene Hes6, an inhibitor of Hes1, promotes neuronal differentiation*. Development, 2000. **127**(13): p. 2933-2943.
178. Jhas, S., et al., *Hes6 inhibits astrocyte differentiation and promotes neurogenesis through different mechanisms*. The Journal of neuroscience, 2006. **26**(43): p. 11061-11071.
179. El Zein, L., et al., *RFX3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies*. Journal of cell science, 2009. **122**(17): p. 3180-3189.
180. Benadiba, C., et al., *The ciliogenic transcription factor RFX3 regulates early midline distribution of guidepost neurons required for corpus callosum development*. 2012.
181. Nakayama, A., et al., *Role for RFX transcription factors in non-neuronal cell-specific inactivation of the microtubule-associated protein MAP1A promoter*. Journal of Biological Chemistry, 2003. **278**(1): p. 233-240.
182. Fang, R., L.C. Olds, and E. Sibley, *Spatio-temporal patterns of intestine-specific transcription factor expression during postnatal mouse gut development*. Gene expression patterns, 2006. **6**(4): p. 426-432.
183. Jacquemin, P., et al., *Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3*. Molecular and cellular biology, 2000. **20**(12): p. 4445-4454.
184. Dusing, M.R., et al., *Onecut-2 knockout mice fail to thrive during early postnatal period and have altered*

- patterns of gene expression in small intestine. *Physiological genomics*, 2010. **42**(1): p. 115-125.
185. Klimova, L., et al., *Onecut1 and Onecut2 transcription factors operate downstream of Pax6 to regulate horizontal cell development*. *Developmental biology*, 2015. **402**(1): p. 48-60.
186. Hallikas, O., et al., *Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity*. *Cell*, 2006. **124**(1): p. 47-59.
187. Evans, P.M. and C. Liu, *Roles of Kruppel-like factor 4 in normal homeostasis, cancer and stem cells*. *Acta biochimica et biophysica Sinica*, 2008. **40**(7): p. 554-564.
188. Zhang, W., et al., *Novel Cross Talk of Kruppel-Like Factor 4 and {beta}-Catenin Regulates Normal Intestinal Homeostasis and Tumor Repression*. *Molecular and cellular biology*, 2006. **26**(6): p. 2055.
189. Rowland, B.D. and D.S. Peeper, *KLF4, p21 and context-dependent opposing forces in cancer*. *Nature Reviews Cancer*, 2005. **6**(1): p. 11-23.
190. Katz, J.P., et al., *The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon*. *Development*, 2002. **129**(11): p. 2619-2628.
191. Segre, J.A., C. Bauer, and E. Fuchs, *Klf4 is a transcription factor required for establishing the barrier function of the skin*. *Nature genetics*, 1999. **22**(4): p. 356-360.
192. Nakahara, Y., et al., *Genetic and epigenetic inactivation of Kruppel-like factor 4 in medulloblastoma*. *Neoplasia (New York, NY)*. **12**(1): p. 20.
193. Jiang, J., et al., *A core Klf circuitry regulates self-renewal of embryonic stem cells*. *Nature cell biology*, 2008. **10**(3): p. 353-360.
194. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors*. *cell*, 2006. **126**(4): p. 663-676.
195. Moore, D.L., et al., *KLF family members regulate intrinsic axon regeneration ability*. *Science*, 2009. **326**(5950): p. 298-301.
196. Foster, K.W., et al., *Oncogene expression cloning by retroviral transduction of adenovirus E1A-immortalized rat kidney RK3E cells: transformation of a host with epithelial features by c-MYC and the zinc finger protein GSKF*. *Cell growth & differentiation: the molecular biology journal of the American Association for Cancer Research*, 1999. **10**(6): p. 423-434.
197. Pandya, A.Y., et al., *Nuclear localization of KLF4 is associated with an aggressive phenotype in early-stage breast cancer*. *Clinical Cancer Research*, 2004. **10**(8): p. 2709-2719.
198. Huang, C.C., et al., *Research Paper KLF4 and PCNA Identify Stages of Tumor Initiation in a Conditional Model of Cutaneous Squamous Epithelial Neoplasia*. *Cancer biology & therapy*, 2005. **4**(12): p. 1401-1408.
199. Rowland, B.D., R. Bernards, and D.S. Peeper, *The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene*. *Nature cell biology*, 2005. **7**(11): p. 1074-1082.
200. Yang, Y., et al., *Research Paper KLF4 and KLF5 Regulate Proliferation, Apoptosis and Invasion in Esophageal Cancer Cells*. *Cancer biology & therapy*, 2005. **4**(11): p. 1216-1221.
201. Chen, X., et al., *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells*. *Cell*, 2008. **133**(6): p. 1106-1117.
202. Swanson, D.J., Y. Tong, and D. Goldowitz, *Disruption of cerebellar granule cell development in the Pax6 mutant, *Sey* mouse*. *Developmental brain research*, 2005. **160**(2): p. 176-193.
203. Zhang, W., et al., *Novel cross talk of Krüppel-like factor 4 and β -catenin regulates normal intestinal homeostasis and tumor repression*. *Molecular and cellular biology*, 2006. **26**(6): p. 2055-2064.
204. Pierce, E.T., *Histogenesis of the deep cerebellar nuclei in the mouse: an autoradiographic study*. *Brain research*, 1975. **95**(2-3): p. 503-518.
205. Evans, P.M., et al., *Krüppel-like factor 4 is acetylated by p300 and regulates gene transcription via modulation of histone acetylation*. *Journal of biological chemistry*, 2007. **282**(47): p. 33994-34002.
206. Amir, R.E., et al., *Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2*. *Nature genetics*, 1999. **23**(2): p. 185-188.
207. Ariani, F., et al., *FOXP1 is responsible for the congenital variant of Rett syndrome*. *The American Journal of Human Genetics*, 2008. **83**(1): p. 89-93.
208. Weaving, L.S., et al., *Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation*. *The American Journal of Human Genetics*, 2004. **75**(6): p. 1079-1093.
209. Margueron, R. and D. Reinberg, *The Polycomb complex PRC2 and its mark in life*. *Nature*, 2011. **469**(7330): p. 343-349.

210. Suzuki, H., et al., *The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line*. *Nature genetics*, 2009. **41**(5): p. 553-562.
211. Altman, J., *Postnatal development of the cerebellar cortex in the rat. II. Phases in the maturation of Purkinje cells and of the molecular layer*. *The Journal of comparative neurology*, 1972. **145**(4): p. 399-463.
212. Mallet, J., R. Christen, and J.P. Changeux, *Immunological studies on the Purkinje cells from rat and mouse cerebella* 1: I. Evidence for antibodies characteristic of the Purkinje cells*. *Developmental biology*, 1979. **72**(2): p. 308-319.
213. Gallo, V., et al., *The role of depolarization in the survival and differentiation of cerebellar granule cells in culture*. *The Journal of neuroscience*, 1987. **7**(7): p. 2203-2213.
214. Sandelin, A., et al., *JASPAR: an open-access database for eukaryotic transcription factor binding profiles*. *Nucleic acids research*, 2004. **32**(suppl 1): p. D91-D94.
215. Consortium, E.P., *The ENCODE (ENCyclopedia of DNA elements) project*. *Science*, 2004. **306**(5696): p. 636-640.
216. Fisher, C.R., et al., *Characterization of mice deficient in aromatase (ArKO) because of targeted disruption of the cyp19 gene*. *Proceedings of the National Academy of Sciences*, 1998. **95**(12): p. 6965-6970.
217. Kamat, A., et al., *A 500-bp region, ~ 40 kb upstream of the human CYP19 (aromatase) gene, mediates placenta-specific expression in transgenic mice*. *Proceedings of the National Academy of Sciences*, 1999. **96**(8): p. 4575-4580.
218. Kamat, A., et al., *Mechanisms in tissue-specific regulation of estrogen biosynthesis in humans*. *Trends in Endocrinology & Metabolism*, 2002. **13**(3): p. 122-128.
219. Golovine, K., M. Schwerin, and J. Vanselow, *Three different promoters control expression of the aromatase cytochrome p450 gene (cyp19) in mouse gonads and brain*. *Biology of reproduction*, 2003. **68**(3): p. 978-984.
220. Dunn, C.A., P. Medstrand, and D.L. Mager, *An endogenous retroviral long terminal repeat is the dominant promoter for human β 1, 3-galactosyltransferase 5 in the colon*. *Proceedings of the National Academy of Sciences*, 2003. **100**(22): p. 12841-12846.
221. Phelps, D.E. and Y. Xiong, *Regulation of cyclin-dependent kinase 4 during adipogenesis involves switching of cyclin D subunits and concurrent binding of p18INK4c and p27Kip1*. *Cell growth & differentiation: the molecular biology journal of the American Association for Cancer Research*, 1998. **9**(8): p. 595-610.
222. Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells*. *Cell*, 2007. **130**(1): p. 77-88.
223. Garneau, N.L., J. Wilusz, and C.J. Wilusz, *The highways and byways of mRNA decay*. *Nature reviews Molecular cell biology*, 2007. **8**(2): p. 113-126.
224. Lewis, B.P., R.E. Green, and S.E. Brenner, *Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans*. *Proceedings of the National Academy of Sciences*, 2003. **100**(1): p. 189-192.
225. Thanaraj, T., F. Clark, and J. Muilu, *Conservation of human alternative splice events in mouse*. *Nucleic Acids Research*, 2003. **31**(10): p. 2544-2552.
226. Irimia, M., et al., *Quantitative regulation of alternative splicing in evolution and development*. *Bioessays*, 2009. **31**(1): p. 40-50.
227. Zdobnov, E.M. and R. Apweiler, *InterProScan—an integration platform for the signature-recognition methods in InterPro*. *Bioinformatics*, 2001. **17**(9): p. 847-848.
228. Banks, R.E., et al., *The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis—preliminary findings*. *Electrophoresis*, 1999. **20**(4-5): p. 689-700.
229. Sluka, P., L. O'Donnell, and P.G. Stanton, *Stage-specific expression of genes associated with rat spermatogenesis: characterization by laser-capture microdissection and real-time polymerase chain reaction*. *Biology of reproduction*, 2002. **67**(3): p. 820-828.
230. Trogan, E. and E.A. Fisher, *Laser capture microdissection for analysis of macrophage gene expression from atherosclerotic lesions*. *Laser Capture Microdissection: Methods and Protocols*, 2005: p. 221-232.
231. Luo, L., et al., *Gene expression profiles of laser-captured adjacent neuronal subtypes*. *Nature medicine*, 1999. **5**(1): p. 117-122.
232. Cañas, R.A., et al., *Transcriptome analysis in maritime pine using laser capture microdissection and 454 pyrosequencing*. *Tree physiology*, 2014. **34**(11): p. 1278-1288.
233. Pisu, M.B., et al., *Proliferation and migration of granule cells in the developing rat cerebellum: cisplatin*

- effects*. The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology, 2005. **287**(2): p. 1226-1235.
234. Bibikova, M. and J.-B. Fan, *GoldenGate® assay for DNA methylation profiling*. DNA Methylation: Methods and Protocols, 2009: p. 149-163.
235. Numata, K., et al., *Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection*. Genome research, 2003. **13**(6b): p. 1301-1306.
236. Kim, J.-M., et al., *A copula method for modeling directional dependence of genes*. BMC bioinformatics, 2008. **9**(1): p. 1.
237. Zhang, W., C. Song, and M. Ye, *Further studies on nonlinear oscillations and chaos of a symmetric cross-ply laminated thin plate under parametric excitation*. International Journal of Bifurcation and Chaos, 2006. **16**(02): p. 325-347.
238. Drira, W. and F. Ghorbel. *Non parametric feature discriminate analysis for high dimension*. in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*. 2012. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Supplementary Table 2.1. 213 Differentially expressed (DE) transcripts that are shared between 1832 DE CAGE tags and 469 DE CbGRiTS probes at E12 – E13.

| DE Refseq | FANTOM5 Annotation | T.test | Fold Change | CbGRiTS Annotation | T.test | Fold Change |
|--------------|--------------------|-------------|-------------|--------------------|-------------|-------------|
| N=213 | | p<0.05 | <0.5 or >2 | | p<0.05 | <0.5 or >2 |
| NM_024283 | 1500015O10Rik | 0.031053041 | 2.592571642 | 1500015O10Rik | 0.026308483 | 2.292624371 |
| NM_183287 | 2610318N02Rik | 0.017254869 | 0.301511961 | 2610318N02Rik | 0.017045119 | 0.45177197 |
| NM_026515 | 2810417H13Rik | 0.000938936 | 0.333348862 | 2810417H13Rik | 0.01112421 | 0.398136028 |
| NM_177854 | 6030405A18Rik | 0.02021341 | 3.877136705 | 6030405A18Rik | 0.001449467 | 3.258783104 |
| NM_176921 | 6030419C18Rik | 0.015400798 | 2.84513776 | 6030419C18Rik | 0.000404885 | 2.075798874 |
| NM_027519 | 6330406I15Rik | 0.028997107 | 6.681467135 | 6330406I15Rik | 0.001894506 | 2.277314169 |
| NM_007417 | Adra2a | 0.004261215 | 2.838706945 | Adra2a | 0.015866327 | 3.161625891 |
| NM_028390 | Anln | 0.032431761 | 0.350888126 | Anln | 0.014730811 | 0.386623169 |
| NM_175535 | Arhgap20 | 0.029272039 | 8.052752955 | Arhgap20 | 0.008206203 | 2.056702798 |
| NM_021493 | Arhgap23 | 0.016839943 | 2.634981058 | 4933428G20Rik | 0.019845589 | 2.284692494 |
| NM_008113 | Arhgdig | 0.007524434 | 5.546956834 | Arhgdig | 0.016841958 | 2.383365149 |
| NM_009791 | Aspm | 0.039309058 | 0.268569597 | Aspm | 0.00165743 | 0.486664687 |
| NM_001029856 | Atad5 | 0.038947926 | 0.462617466 | Atad5 | 0.0255601 | 0.405001737 |
| NM_015731 | Atp9a | 0.012166576 | 2.214724382 | Atp9a | 0.005681928 | 3.03914744 |
| NM_011497 | Aurka | 0.009370853 | 0.30077459 | Aurka | 0.031015155 | 0.342220953 |
| NM_178737 | AW551984 | 0.019135016 | 3.292010161 | AW551984 | 0.016733588 | 2.216576775 |
| NM_144935 | BC018242 | 0.005519839 | 2.493139229 | BC018242 | 0.01584381 | 2.609306948 |
| NM_009773 | Bub1b | 0.03512861 | 0.296049405 | Bub1b | 0.016345976 | 0.439824538 |
| NM_175524 | C130060K24Rik | 0.009778235 | 0.152896482 | C130060K24Rik | 0.015006511 | 0.465117352 |
| NM_026161 | C1qtnf4 | 0.001776646 | 4.640363482 | C1qtnf4 | 0.004918782 | 4.57149698 |
| NM_007583 | Cacng2 | 0.041103566 | 2.21761564 | Cacng2 | 0.006476341 | 2.472550522 |
| NM_007583 | Cacng2 | 0.041103566 | 2.21761564 | Cacng2 | 0.011181576 | 3.126756055 |
| NM_007586 | Calb2 | 0.037400496 | 12.29226044 | Calb2 | 0.004772223 | 3.947666184 |
| NM_007586 | Calb2 | 0.037400496 | 12.29226044 | Calb2 | 0.034443971 | 2.706321683 |
| NM_007595 | Camk2b | 0.043164257 | 4.818691039 | Camk2b | 0.006776609 | 4.954249984 |
| NM_028296 | Car10 | 0.043462017 | 5.132047724 | Car10 | 0.005884053 | 2.834314793 |
| NM_009800 | Car11 | 0.01986112 | 2.462023674 | Car11 | 0.022538379 | 2.087341521 |
| NM_009824 | Cbfa2t3 | 0.007578661 | up from 0 | Cbfa2t3h | 0.006943062 | 2.331082396 |
| NM_027411 | Ccdc99 | 0.021735814 | 0.314306839 | Ccdc99 | 0.020051991 | 0.457232545 |
| NM_172301 | Ccnb1 | 0.036659583 | 0.254621361 | Ccnb1 | 0.000487364 | 0.326842312 |
| NM_172301 | Ccnb1 | 0.036659583 | 0.254621361 | Ccnb1 | 0.01007709 | 0.358820279 |
| NM_007631 | Ccnd1 | 0.001137538 | 0.297882243 | Ccnd1 | 0.026245454 | 0.375269774 |
| NM_009829 | Ccnd2 | 0.03440036 | 0.199067622 | Ccnd2 | 0.004328801 | 0.490162946 |
| NM_009829 | Ccnd2 | 0.03440036 | 0.199067622 | Ccnd2 | 0.016288724 | 0.333709964 |
| NM_001037134 | Ccne2 | 0.034814494 | 0.314636519 | Ccne2 | 0.013234525 | 0.472591854 |
| NM_010818 | Cd200 | 0.033109844 | 2.683764525 | Cd200 | 0.015937393 | 2.747272467 |
| NM_010818 | Cd200 | 0.033109844 | 2.683764525 | Cd200 | 0.027038815 | 2.694467154 |

| DE Refseq | FANTOM5 Annotation | T.test | Fold Change | CbGRiTS Annotation | T.test | Fold Change |
|--------------|--------------------|-------------|-------------|--------------------|-------------|-------------|
| NM_023223 | Cdc20 | 0.030460145 | 0.257086483 | Cdc20 | 0.013413358 | 0.492433221 |
| NM_009860 | Cdc25c | 0.004668656 | 0.307610894 | Cdc25c | 0.024945689 | 0.454599079 |
| NM_009860 | Cdc25c | 0.004668656 | 0.307610894 | Cdc25c | 0.029475371 | 0.448651353 |
| NM_026772 | Cdc42ep2 | 0.025385232 | 2.345534563 | Cdc42ep2 | 0.012422166 | 2.521589327 |
| NM_001025779 | Cdc6 | 0.008004521 | 0.2831293 | Cdc6 | 0.022884895 | 0.483861705 |
| NM_175384 | Cdca2 | 0.008161333 | 0.284354098 | Cdca2 | 0.003535064 | 0.473903975 |
| NM_013538 | Cdca3 | 0.032767101 | 0.31821488 | Cdca3 | 0.02308864 | 0.399241427 |
| NM_026410 | Cdca5 | 0.003176217 | 0.290001449 | Cdca5 | 0.006655126 | 0.39896479 |
| NM_026560 | Cdca8 | 0.013291118 | 0.275551966 | Cdca8 | 0.028727584 | 0.446273486 |
| NM_130878 | Cdhr1 | 0.004987537 | 3.532462247 | Pcdh21 | 0.009025889 | 6.622304429 |
| NM_016756 | Cdk2 | 0.028059286 | 0.37227651 | Cdk2 | 0.005643516 | 0.425825229 |
| NM_016756 | Cdk2 | 0.028059286 | 0.37227651 | Cdk2 | 0.032185185 | 0.478304079 |
| NM_173762 | Cenpe | 0.000320551 | 0.286811814 | Cenpe | 0.004202131 | 0.370873733 |
| NM_021886 | Cenph | 0.003153866 | 0.342916931 | Cenph | 0.031493299 | 0.494371221 |
| NM_021790 | Cenpk | 0.039067746 | 0.442885118 | Cenpk | 0.038656322 | 0.461797841 |
| NM_028131 | Cenpn | 0.016996327 | 0.390888608 | Cenpn | 0.018107996 | 0.47554925 |
| NM_025495 | Cenpp | 0.012053087 | 0.366537977 | Cenpp | 0.015696039 | 0.495400307 |
| NM_028760 | Cep55 | 0.018591155 | 0.308693135 | Cep55 | 0.000504148 | 0.371903441 |
| NM_013733 | Chaf1a | 0.0317759 | 0.286086118 | Chaf1a | 0.006390244 | 0.496890547 |
| NM_001081376 | Chd5 | 0.03752828 | 7.723698804 | Chd5 | 0.001286196 | 5.269502153 |
| NM_181589 | Ckap2l | 0.0458131 | 0.250248314 | Ckap2l | 0.003163882 | 0.352655962 |
| NM_013805 | Cldn5 | 0.042815094 | 2.179433893 | Cldn5 | 0.006847744 | 2.525670902 |
| NM_175554 | Clspn | 0.013667632 | 0.261203039 | Clspn | 0.004362574 | 0.487565066 |
| NM_175554 | Clspn | 0.013667632 | 0.261203039 | Clspn | 0.011867865 | 0.377792683 |
| NM_130457 | Cntnap4 | 0.030252461 | 49.91243135 | Cntnap4 | 0.001607027 | 5.41889991 |
| NM_198300 | Cpeb3 | 0.031765901 | 2.084116823 | Cpeb3 | 0.032163536 | 2.119415821 |
| NM_026412 | D2ErtD750e | 0.012302791 | 0.27950813 | D2ErtD750e | 0.026961944 | 0.466516496 |
| NM_021532 | Dact1 | 0.003350712 | 3.462601632 | Dact1 | 0.005294892 | 3.619189545 |
| NM_013726 | Dbf4 | 0.042402732 | 0.348274933 | Dbf4 | 0.037970268 | 0.378929142 |
| NM_145217 | Diras1 | 0.02557951 | 2.005137548 | Diras1 | 0.002714342 | 2.925466437 |
| NM_001024474 | Diras2 | 0.017120636 | 2.183006106 | Diras2 | 0.001371884 | 2.783692784 |
| NM_144553 | Dlgap5 | 0.018270548 | 0.308211846 | Dlg7 | 0.001645432 | 0.42122608 |
| NM_207666 | Dlk2 | 0.011126196 | 5.838566365 | Dlk2 | 0.003993651 | 2.466844312 |
| NM_010065 | Dnm1 | 0.036376054 | 2.100844665 | Dnm1 | 0.00324879 | 2.146024679 |
| NM_008052 | Dtx1 | 0.005452502 | 4.30008881 | Dtx1 | 0.018755669 | 2.590085998 |
| NM_172442 | Dtx4 | 0.029398233 | 0.430788542 | Dtx4 | 0.030311202 | 0.486439852 |
| NM_010330 | Emb | 0.027253835 | 0.385282727 | Emb | 0.012562588 | 0.456071946 |
| NM_001003815 | Epb4.111 | 0.020200901 | 2.584940585 | Epb4.111 | 0.002778153 | 2.565073206 |
| NM_007939 | Epha8 | 0.003525624 | 7.080252701 | Epha8 | 0.004388796 | 2.231994285 |
| NM_011934 | Esrrb | 0.011497484 | 44.7746343 | Esrrb | 0.001234522 | 9.458862288 |
| NM_011935 | Esrrg | 0.036727395 | 12.7536294 | Esrrg | 0.023535812 | 2.765103214 |
| NM_173446 | Fam155a | 0.0041619 | 3.885608162 | Tmem28 | 0.015328982 | 3.036339968 |
| NM_207583 | Fam5b | 0.036826753 | 2.855266066 | 6430517E21Rik | 0.009485044 | 2.121865681 |

| DE Refseq | FANTOM5 Annotation | T.test | Fold Change | CbGRiTS Annotation | T.test | Fold Change |
|--------------|--------------------|-------------|-------------|--------------------|-------------|-------------|
| NM_144526 | Fam64a | 0.012008473 | 0.343395778 | 6720460F02Rik | 0.020644389 | 0.498615626 |
| NM_172930 | Fam70a | 0.012112908 | 3.607486816 | 6430550H21Rik | 0.002652523 | 3.08370908 |
| NM_007999 | Fen1 | 0.005696685 | 0.389685558 | Fen1 | 0.019828504 | 0.423470487 |
| NM_021891 | Figl1 | 0.027000031 | 0.339340866 | Figl1 | 0.044164426 | 0.402669098 |
| NM_001007580 | Fndc3c1 | 0.00611138 | 0.110169311 | Gm784 | 0.004199163 | 0.21668432 |
| NM_178856 | Gins2 | 0.024190193 | 0.392323648 | Gins2 | 0.008285019 | 0.486664687 |
| NM_010290 | Gjd2 | 0.00306062 | 3.248620845 | Gja9 | 0.011712243 | 3.288280416 |
| NM_183427 | Gira2 | 0.027796526 | 5.417804631 | Gira2 | 0.001452911 | 4.922302541 |
| NM_001035122 | Golm1 | 0.020844063 | 0.324260797 | Golm1 | 0.002709378 | 0.483191395 |
| NM_001004761 | Gpr158 | 0.042578241 | 2.853285019 | Gpr158 | 0.019717211 | 2.091686542 |
| NM_013533 | Gpr162 | 0.027047288 | 2.133067398 | Ms10h | 0.040825176 | 2.085413299 |
| NM_177383 | Gpr21 | 0.023159897 | 5.293500881 | Gpr21 | 0.003674267 | 2.384466752 |
| NM_015764 | Greb1 | 0.012802394 | 0.339333825 | Greb1 | 0.001365972 | 0.488241444 |
| NM_010353 | Gsg2 | 0.009782923 | 0.339897776 | Gsg2 | 0.020132196 | 0.456071946 |
| NM_021896 | Gucy1a3 | 0.049969244 | 2.296467351 | Gucy1a3 | 0.048077066 | 2.25219748 |
| NM_008216 | Has2 | 0.016676865 | 4.755452662 | Has2 | 0.000760597 | 2.229932437 |
| NM_173400 | Haus6 | 0.003123742 | 0.477776093 | 6230416J20Rik | 0.001534968 | 0.447305789 |
| NM_173400 | Haus6 | 0.003123742 | 0.477776093 | 6230416J20Rik | 0.004775085 | 0.461264659 |
| NM_008234 | Hells | 0.043211936 | 0.313800153 | Hells | 0.003479279 | 0.43951978 |
| NM_175659 | Hist1h2ah | 0.010638793 | 0.237936001 | Hist1h2ah | 0.027694099 | 0.463079993 |
| NM_016710 | Hmgn5 | 0.00304962 | 0.420610452 | Nsbp1 | 0.037370103 | 0.424155939 |
| NM_019455 | Hpgds | 0.011004549 | 0.28300076 | Ptgds2 | 0.002128648 | 0.493458273 |
| NM_001033354 | Iqsec3 | 0.049568151 | 2.481912365 | Iqsec3 | 0.003062941 | 2.454904099 |
| NM_133207 | Kcnh7 | 0.047705381 | 2.783832799 | Kcnh7 | 0.020431781 | 2.05955597 |
| NM_145588 | Kif22 | 0.005598823 | 0.302223563 | Kif22 | 0.002813125 | 0.426317446 |
| NM_024245 | Kif23 | 0.012469297 | 0.328524334 | Kif23 | 0.008675807 | 0.465009899 |
| NM_134471 | Kif2c | 0.000687214 | 0.325648594 | Kif2c | 0.001936657 | 0.361900917 |
| NM_008446 | Kif4 | 0.003580507 | 0.292035293 | Kif4 | 0.00202683 | 0.367801686 |
| NM_008446 | Kif4 | 0.003580507 | 0.292035293 | Kif4 | 0.034185956 | 0.421323415 |
| NM_026324 | Kirrel3 | 0.030568024 | 3.809849662 | Kirrel3 | 0.008991779 | 2.401606855 |
| NM_026324 | Kirrel3 | 0.030568024 | 3.809849662 | Kirrel3 | 0.021602517 | 2.348923942 |
| NM_010655 | Kpna2 | 0.026459846 | 0.484225764 | Kpna2 | 0.02755381 | 0.418219818 |
| NM_133815 | Lbr | 0.016318242 | 0.351978934 | Lbr | 0.00614217 | 0.330563651 |
| NM_025681 | Lix1 | 0.026492878 | 0.266571709 | Lix1 | 0.006973583 | 0.365683305 |
| NM_178714 | Lrfn5 | 0.027633004 | 3.735175233 | Lrfn5 | 0.032793249 | 2.344586219 |
| NM_016753 | Lxn | 0.037922598 | 2.561631097 | Lxn | 0.000693729 | 4.053964157 |
| NM_019499 | Mad2l1 | 0.015098463 | 0.33901826 | Mad2l1 | 0.032368037 | 0.484868914 |
| NM_001038609 | Mapt | 0.034959575 | 2.820323372 | Mapt | 0.01720518 | 3.487032958 |
| NM_001045533 | Mar4 | 0.041111478 | 4.170238927 | Mar4 | 0.000510172 | 3.532443944 |
| NM_008564 | Mcm2 | 0.028423225 | 0.298151313 | Mcm2 | 0.042514578 | 0.496202187 |
| NM_008563 | Mcm3 | 0.011155901 | 0.310840895 | Mcm3 | 0.016811036 | 0.456071946 |
| NM_008565 | Mcm4 | 0.021338024 | 0.418265279 | Mcm4 | 0.000476458 | 0.482076274 |
| NM_008566 | Mcm5 | 0.007512975 | 0.398831891 | Mcm5 | 0.010853237 | 0.416003239 |

| DE Refseq | FANTOM5 Annotation | T.test | Fold Change | CbGRiTS Annotation | T.test | Fold Change |
|--------------|--------------------|-------------|-------------|--------------------|-------------|-------------|
| NM_008568 | Mcm7 | 0.016353796 | 0.391326533 | Mcm7 | 0.001879211 | 0.474232574 |
| NM_207010 | Mdga2 | 0.045062206 | 2.579060581 | Mdga2 | 0.015242177 | 2.010656547 |
| NM_172578 | Mis18bp1 | 0.018536223 | 0.293134808 | C79407 | 0.001328813 | 0.360315678 |
| NM_010833 | Msn | 0.009042958 | 0.428869609 | Msn | 0.015656342 | 0.488918759 |
| NM_010836 | Msx3 | 0.033166764 | 0.202885397 | Msx3 | 0.004960799 | 0.488467111 |
| NM_013603 | Mt3 | 0.014563587 | 2.396694807 | Mt3 | 0.006746373 | 3.136886507 |
| NM_177369 | Myh8 | 0.042837452 | 12.52917634 | Myh8 | 0.007052858 | 12.53885175 |
| NM_001093775 | Myt1l | 0.043768664 | 2.337243234 | Myt1l | 0.012416298 | 2.36416927 |
| NM_001093778 | Myt1l | 0.043768664 | 2.337243234 | Myt1l | 0.000405941 | 2.682044796 |
| NM_001081475 | Nasp | 0.014858922 | 0.48574163 | Nasp | 0.01564181 | 0.471937156 |
| NM_144818 | Ncaph | 0.011994828 | 0.339692164 | Ncaph | 0.030306214 | 0.469652849 |
| NM_023294 | Ndc80 | 0.049564029 | 0.325484025 | Ndc80 | 0.039209048 | 0.491978327 |
| NM_023317 | Nde1 | 0.018125875 | 0.336671256 | Nde1 | 0.01565469 | 0.421615555 |
| NM_013864 | Ndrp2 | 0.037217852 | 2.839361472 | Ndrp2 | 0.000861629 | 4.003698494 |
| NM_010895 | Neurod2 | 0.019394379 | 2.416323519 | Neurod2 | 0.008833 | 4.202749105 |
| NM_001077403 | Nrp2 | 0.030637236 | 4.94779986 | Nrp2 | 0.021754049 | 2.593079911 |
| NM_001042652 | Nusap1 | 0.023745651 | 0.283500985 | Nusap1 | 0.004372374 | 0.380332554 |
| NM_183297 | Nxph4 | 0.012895163 | 3.546153101 | Nxph4 | 0.010630114 | 4.89734557 |
| NM_023209 | Pbk | 0.014486826 | 0.299478759 | Pbk | 0.000532136 | 0.422786144 |
| NM_001098170 | Pcdh10 | 0.046596237 | 2.881472006 | Pcdh10 | 0.000527882 | 2.733974904 |
| NM_001098170 | Pcdh10 | 0.046596237 | 2.881472006 | Pcdh10 | 0.040287121 | 2.177490865 |
| NM_001081377 | Pcdh9 | 0.002502439 | 6.395556664 | Pcdh9 | 0.0151043 | 2.012050711 |
| NM_001081377 | Pcdh9 | 0.002502439 | 6.395556664 | Pcdh9 | 0.029091488 | 2.550888783 |
| NM_011045 | Pcna | 0.017844358 | 0.420556034 | Pcna | 0.028674714 | 0.477089994 |
| NM_016861 | Pdlim1 | 0.020103838 | 0.288084215 | Pdlim1 | 0.00428959 | 0.496890547 |
| NM_172453 | Pif1 | 0.009102603 | 0.413608232 | Pif1 | 0.004874469 | 0.40007249 |
| NM_172453 | Pif1 | 0.009102603 | 0.413608232 | Pif1 | 0.014615826 | 0.477420802 |
| NM_011121 | Plk1 | 0.013973386 | 0.303398018 | Plk1 | 0.018684106 | 0.449377563 |
| NM_012040 | Pnck | 0.029009623 | 3.414076147 | Pnck | 0.013471916 | 2.206357644 |
| NM_011132 | Pole | 0.010173332 | 0.470254391 | Pole | 0.005720413 | 0.49425701 |
| NM_011625 | Ppp1r13b | 0.035369836 | 2.687613859 | Ppp1r13b | 0.025310818 | 2.716972569 |
| NM_027531 | Ppp2r2b | 0.015085927 | 3.557506055 | Ppp2r2b | 0.004846808 | 3.122424454 |
| NM_145150 | Prc1 | 0.014623729 | 0.291225389 | Prc1 | 0.002396914 | 0.425530172 |
| NM_008935 | Prom1 | 0.00470212 | 0.301683231 | Prom1 | 0.010931323 | 0.42533358 |
| NM_008935 | Prom1 | 0.00470212 | 0.301683231 | Prom1 | 0.01121146 | 0.424744351 |
| NM_001093750 | Ptchd1 | 0.009037371 | 7.98624984 | Ptchd1 | 0.013482338 | 2.806294985 |
| NM_013645 | Pvalb | 0.036023913 | 2.916684423 | Pvalb | 0.037371312 | 3.094414943 |
| NM_176971 | Rab9b | 0.021094585 | 2.110728952 | Rab9b | 0.003838773 | 2.069573281 |
| NM_133223 | Rac3 | 0.011596064 | 2.666294236 | Rac3 | 0.011782991 | 3.128201257 |
| NM_009013 | Rad51ap1 | 0.006605128 | 0.292496549 | Rad51ap1 | 0.003422431 | 0.360648835 |
| NM_001039556 | Rad54b | 0.015822171 | 0.404726051 | E130016E03Rik | 0.000178972 | 0.481519679 |
| NM_030690 | Rai14 | 0.007927375 | 0.479814695 | Rai14 | 0.009281247 | 0.47587899 |
| NM_011249 | Rbl1 | 0.027698817 | 0.384912275 | Rbl1 | 0.008818555 | 0.465439858 |

| DE Refseq | FANTOM5 Annotation | T.test | Fold Change | CbGRiTS Annotation | T.test | Fold Change |
|--------------|--------------------|-------------|-------------|--------------------|-------------|-------------|
| NM_153793 | Rell2 | 0.015287301 | 4.158204414 | Rell2 | 0.022887049 | 2.132186055 |
| NM_178779 | Rnf152 | 0.014323596 | 3.322790226 | Rnf152 | 0.007563178 | 2.29739671 |
| NM_175549 | Robo2 | 0.003518151 | 2.792960545 | Robo2 | 0.012809423 | 2.072444308 |
| NM_013646 | Rora | 0.030108338 | 4.617289857 | Rora | 0.022190747 | 3.138336392 |
| NM_011284 | Rpa2 | 0.023136883 | 0.438626862 | Rpa2 | 0.005032288 | 0.492319458 |
| NM_146244 | Rps6kl1 | 0.021414021 | 3.443608363 | Rps6kl1 | 0.01388721 | 2.462288827 |
| NM_009103 | Rrm1 | 0.019441251 | 0.406810847 | Rrm1 | 0.001029001 | 0.368993336 |
| NM_009104 | Rrm2 | 0.010276888 | 0.238981787 | Rrm2 | 0.025745887 | 0.321153146 |
| NM_009107 | Rxrg | 0.02759614 | 2.101756052 | Rxrg | 0.02918701 | 2.055752619 |
| NM_009129 | Scg2 | 0.041874163 | 3.266851357 | Scg2 | 0.009589226 | 2.663518559 |
| NM_053197 | Sfxn3 | 0.02160703 | 2.461040659 | Sfxn3 | 0.021939623 | 2.084449856 |
| NM_199007 | Sgol2 | 0.022159742 | 0.320372753 | Sgol2 | 0.005074608 | 0.440027827 |
| NM_009171 | Shmt1 | 0.027496194 | 0.35812413 | Shmt1 | 0.001680811 | 0.461477858 |
| NM_013787 | Skp2 | 0.019358463 | 0.402871214 | Skp2 | 0.002086231 | 0.353063601 |
| NM_009199 | Slc1a1 | 0.006495782 | 5.781559915 | Slc1a1 | 0.010538225 | 2.277840401 |
| NM_172479 | Slc38a5 | 0.000125112 | 8.376688747 | Slc38a5 | 0.000289388 | 2.930878848 |
| NM_008017 | Smc2 | 0.01189068 | 0.321605061 | Smc2 | 0.00302453 | 0.317171123 |
| NM_017407 | Spag5 | 0.015504526 | 0.312565224 | Spag5 | 0.00507416 | 0.437999158 |
| NM_025565 | Spc25 | 0.002898123 | 0.229345312 | Spc25 | 0.0099898 | 0.415618948 |
| NM_177774 | Srsf12 | 0.031706256 | 2.706601418 | Srrp | 0.048100948 | 2.091203315 |
| NM_009285 | Stc1 | 0.028500412 | 3.057908635 | Stc1 | 0.002655842 | 2.242332156 |
| NM_013873 | Sult4a1 | 0.017422727 | 2.239389298 | Sult4a1 | 0.007639557 | 2.120885398 |
| NM_153579 | Sv2b | 0.041180915 | 6.971811589 | Sv2b | 0.000725876 | 4.238831642 |
| NM_011522 | Syngn3 | 0.017880379 | 3.403376188 | Syngn3 | 0.008714543 | 2.588291309 |
| NM_172804 | Syt16 | 0.028781934 | 4.747358345 | Syt16 | 0.000124968 | 2.493202389 |
| NM_009387 | Tk1 | 0.01460037 | 0.434521497 | Tk1 | 0.035393569 | 0.457972645 |
| NM_177735 | Tmem130 | 0.024217587 | 2.652906884 | Tmem130 | 0.008201005 | 4.050219229 |
| NM_183311 | Tmem145 | 0.018615213 | 2.612926434 | Tmem145 | 0.017068272 | 2.115501927 |
| NM_001002267 | Tmem158 | 0.049613056 | 2.365856343 | Tmem158 | 0.034440673 | 2.754900093 |
| NM_178915 | Tmem179 | 0.000586949 | 2.408578656 | Tmem179 | 0.00100862 | 2.002311826 |
| NM_011607 | Tnc | 0.026341718 | 23.37374355 | Tnc | 0.005539685 | 3.139061585 |
| NM_011623 | Top2a | 0.006711827 | 0.254426401 | Top2a | 0.00320143 | 0.358323193 |
| NM_011623 | Top2a | 0.006711827 | 0.254426401 | Top2a | 0.010967954 | 0.399333682 |
| NM_009413 | Tpd52l1 | 0.010126614 | 3.410985052 | Tpd52l1 | 0.012782522 | 2.006016306 |
| NM_028109 | Tpx2 | 0.006766565 | 0.275778442 | Tpx2 | 0.029226124 | 0.436584656 |
| NM_028417 | Ttc9b | 0.034713183 | 2.062713094 | Ttc9b | 0.007314654 | 3.927650796 |
| NM_009448 | Tuba1c | 0.002186798 | 0.318106436 | Tuba1c | 0.030676118 | 0.421907898 |
| NM_134028 | Tubg2 | 0.032110039 | 3.086753642 | Tubg2 | 0.014417575 | 2.706321683 |
| NM_021288 | Tyms | 0.034719766 | 0.440876185 | Tyms | 0.024058168 | 0.483861705 |
| NM_010931 | Uhrf1 | 0.014233783 | 0.313608563 | Uhrf1 | 0.001875238 | 0.448236903 |
| NM_145967 | Vstm2a | 0.020603975 | 4.111164339 | Vstm2a | 0.001494602 | 3.866424388 |
| NM_145967 | Vstm2a | 0.020603975 | 4.111164339 | Vstm2a | 0.022330639 | 2.913998228 |
| NM_198627 | Vstm2l | 0.028898318 | 2.420530681 | Vstm2l | 0.023958267 | 2.603285128 |

| DE Refseq | FANTOM5 Annotation | T.test | Fold Change | CbGRiTS Annotation | T.test | Fold Change |
|------------------|---------------------------|---------------|--------------------|---------------------------|---------------|--------------------|
| NM_026940 | Ydjc | 0.03505929 | 2.397074905 | 1810015A11Rik | 0.014929493 | 2.177994031 |
| NM_153541 | Zbtb8b | 0.043651004 | 2.743957808 | Zbtb8 | 0.001114683 | 2.245442844 |
| NM_175494 | Zfp367 | 0.013147837 | 0.316211334 | Zfp367 | 0.01402851 | 0.428489968 |
| NM_027678 | Zranb3 | 0.000983845 | 0.283282269 | Zranb3 | 0.001667365 | 0.492888537 |