Visual Analytics of Patterns of Gene Expression in the Developing Mammalian Brains

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Qihang Li, B.S., M.S.

Graduate Program in Department of Computer Science and Engineering

The Ohio State University

2017

Dissertation Committee:

Dr. Raghu Machiraju, Advisor

Dr. Kun Huang, co-Advisior

Dr. Christopher Bartlett

Abstract

Understanding the inherent genomic characteristic in the developing brain is a critical topic in developing neuroscience and translational bioinformatics. Since the brain is one of the longest and most complex developing structure, the exploration of the comprehensive patterns that describe the changes during growth poses significant challenges. With advances in data generation technologies, the Allen Developing Mouse Brain Atlas (ADMBA) and Allen Developing Human Brain Atlas (ADHBA) projects became an invaluable resource for neuroscientists and developmental biologists for exploring interesting spatial and temporal patterns of gene expression. However, given the extremely large amounts of data, it is desirable to apply visualization techniques to their access, analysis, and interpretation.

This dissertation proposes an extensible visual analytics framework for the spatiotemporal pattern exploration in the developing mouse and human brain. Targeting on the ADMBA and ADHBA data, I developed three visual analytics components: the spatial pattern exploration, the region-based temporal pattern exploration, and the integrative gene gradient-based spatiotemporal pattern exploration. The spatial pattern exploration component, HOS-Tree system, uses a tree-layout visualization to present the structural hierarchy and uses colors to indicate the developing orientations. Also, the region-based temporal pattern exploration component uses data-mining approaches to provide interactive pattern

presentation among genes and structures. In addition, we use the gradient of gene expression to define the spatiotemporal genomic characteristic, and also design a 3-D visualization component to provide the exploration of the spatiotemporal patterns. For each visualization component, I investigate the visual analytics result by seeking the biological interpretation of the explored patterns. The investigation shows that several explorations are well-interpreted by development ground truth. These explored patterns could lead to future studies potentially.

Finally, the proposed visual analytics framework and the containing approaches can be extended to generalized tools and applications in which exploration and integration of spatiotemporal data are needed. This dissertation also provides high-level design considerations for future researchers and practitioners about the conceptual methodologies in integrative visual analytics in spatiotemporal pattern exploration.

This document is dedicated to my family.

Acknowledgments

First of all, I would like to express my sincere gratitude to the committee chair, Prof. Raghu Machiraju. He has served as my primary advisor from the very beginning of my Ph.D. program at The Ohio Stage University, and expertly guided me through my entire graduate education. His expertise, understanding, patience, and support made it possible for me to focus on doing research. I am forever grateful to him for spending days and nights on my works and papers, and, I owe you.

I am hugely indebted to my co-adviser, Prof. Kun Huang. I am not sure many graduate students have two highly respected academics as mentors during the graduate program. But for me, this the best gift from God. Thank you, for providing me philosophical guidance, for giving the much-needed support, for everything.

I am also very grateful to my committee member Dr. Christopher Bartlett and graduate faculty representative Dr. Ralf Bundschuh. I have greatly benefited from their intellectual guidances and valuable comments. I will always be thankful for people in the Allen Institute for Brain Science: Dr. Carol Thompson and Dr. David Feng. It would not have been possible to finish this dissertation without their valuable suggestion and persistent help.

My appreciation also entends to my laboratory fellows: Dr. Okan Irfanoglu, Dr. Chao Wang, Dr. Hao Ding, Dr. Nan Meng, Dr. Zhi Han, Brian Arand, Arunima Srivastava.

Most importantly, I would like to thank my wife, Xingdan. Her support, patience, and encouragement were indisputably the bedrock for the past six years of my life. I also thank

my parents and relatives for their kindness and endless support. I would not complete my education without their support.

Vita

2007	B.S. Information and Computing Science, Hangzhou Dianzi University, China
2010	M.S. Information Systems Management, Morehead State University, USA
2015	M.S. Computer Science and Engineering, The Ohio State University, USA
2011-present	Graduate Research Assistant, Computer Science and Engineering, Biomedical Informatics, The Ohio State University, USA

Publications

Research Publications

Qihang Li, Gabriel Zachmann, David Feng, Kun Huang, and Raghu Machiraju, "2013 IEEE Scientific Visualization Contest Winner: Observing Genomics and Phenotypical Patterns in the Developing Mouse Brain". *IEEE Computer Graphics and Application*, 34(5):88-97, 2014.

Qihang Li, Kun Huang, and Raghu Machiraju, "Spatiotemporal Visualization of Gene Expression in the Developing Mouse Brain". *Eurographics Conference on Visualization (EuroVis) 2017*.

Fields of Study

Major Field: Computer Science and Engineering

Table of Contents

Page

Abstract ii

Dedication iv

Acknowledgments v

Vita vii

List of Tables xi

List of Figures xv

1. Introduction 1

    1.1 Motivation 8
    1.2 Challenges and aims 10
    1.3 Thesis statement 12
    1.4 Outline of Solutions 13
    1.5 Organization of this dissertation 15

2. Related Works 16

    2.1 Allen Brain Neuroscience Datasets 16
        2.1.1 The Allen Developing Mouse Brain Atlas 17
        2.1.2 The Allen Developing Human Brain Atlas 18
    2.2 Visualization Approaches for Spatial, Temporal, and Hierarchical Patterns 20
        2.2.1 Pattern Exploration for the Allen Brain Atlas 23
        2.2.2 Released Tools for AIBS Data Portal 25
        2.2.3 Existing Challenges in Capturing Spatiotemporal Gene Gradients 26
    2.3 Summary 27

3. Exploration of the Spatial Pattern of Brain Development 28

3.1 Background and Motivation 29
3.2 Data and Tasks 30
3.2.1 Data Collection 31
3.2.2 Tasks 31
3.3 Visual Analytics Design 33
3.3.1 Using Tree-layouts to Present Structures 35
3.3.2 Coloring the HOS-Tree Using Developing Orientation 35
3.3.3 Interactive Functions in HOS-Tree 39
3.4 Observations and Results 40
3.4.1 Exploring Global Spatial Pattern of Structures using HOS-Tree . 40
3.4.2 Exploring Gene Profiles at Various Stages 43
3.5 Summary 47

4. Exploration of the Temporal Pattern of Brain Development 48

4.1 Background and Movitation 49
4.2 Data and Tasks 53
4.2.1 Data Collection 53
4.2.2 Tasks 54
4.3 Visual Analytics Design 55
4.3.1 Data Processing 55
4.3.2 Learning Temporal Patterns 58
4.3.3 The Bi-clustered Gene Expression Flow Matrix (BGEFM) 59
4.4 Implementation and Result for the Developing Mouse Brain Dataset . . . 62
4.4.1 Temporal Patterns and BGEFM 62
4.4.2 Observations on Exploration of the ADMBA Dataset 66
4.5 Implementation and Result for the Developing Human Brain Dataset . . 72
4.5.1 Temporal Patterns and BGEFM 73
4.5.2 Observations on Exploration of the ADHBA Dataset 77
4.6 Summary 85

5. Exploration of Spatiotemporal Patterns of Brain Development 87

5.1 Movitation 88
5.2 Visual Analytics Design 91
5.2.1 Exploring Fold Changes of Gene Expression 91
5.2.2 Generating the Integrative Gene Gradient 93
5.2.3 Measuring the Spatiotemporal Anisotropy of the Gene Gradient . 94
5.2.4 Visualizing the Spatiotemporal Anisotropy 95
5.2.5 Role of Validation 96

5.3 Result 98
5.3.1 3D gradient and anisotropy provided a visualization of directionality of gene expression. 98
5.3.3 Genes with specific gradient and anisotropy properties enriched in specific biological functions. 104
5.4 Summary 109

6. Contributions and Future Work 111

Bibliography 118

List of Tables

Table **Page**

4.1 **Enrichment results of the genes in different gene clusters in the ADMBA data.** The p values are in the $-log_{10}$ space. 69

4.2 **Significant neurodevelopment events of the human in various stages.** The neuron-related regions and functions started to develop at very early stages and stopped at the 4^{th} stage, and only the 1^{st} and 2^{nd} temporal patterns best matched this. The myelination functions have significant change in the 5^{th}, 6^{th}, and 7^{th} stages. The 3^{rd} and 6^{th} temporal patterns best matched these changes. 78

4.3 **Enrichment results of the genes in different gene clusters in the ADMBA data.** The p values are in the $-log_{10}$ space and used the Benjamini correction. The p=0.05 is shown using red-dot line. The genes in the 1^{st} and 6^{th} gene clusters are strongly enriched in synaptic signaling and transmission. The genes in the 3^{rd} gene cluster are strongly enriched in neural-related function and development. The genes in the 8^{th} gene cluster also show enriched functions in learning-behaviors. 82

5.1 **Enrichment analysis of the top GOI with significant MGO properties.** GOI with top 100 largest MGO in $< +x >$ direction, and with top 100 largest MGO in $< -z, +z >$ direction were chosen for the analysis (the entire 1753 genes are used as enrichment background). The genes with the largest MGO in $< +x >$ direction have strong function enrichment in phosphorylation and metabolic process, while those with largest MGO in $< -z, +z >$ direction have strong function enrichment in the cell and nervous system development. 105

5.2 **GOI analysis based on the 50 top smallest GA values.** (a) The genomic enrichment of the GOI of top the 50 diffusional genes (with the smallest GA values), which enriched consistently in nervous and tube development. (b) Averaged 3D expression data of the sample genes with the smallest GA values. 106

5.3 **GOI analysis based on the 50 top largest GA values.** (a) The genomic enrichment of the GOI of the top 50 directional genes (with the largest GA values) which strongly enriched in structure development. (b) Averaged 3D expression data of the sample genes with the largest GA values. 107

A.1 Biological enrichment analysis of the 1^{st} and 2^{nd} gene clusters in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 130

A.2 Biological enrichment analysis of the 3^{rd} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 131

A.3 Biological enrichment analysis of the 4^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 132

A.4 Biological enrichment analysis of the 5^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 133

A.5 Biological enrichment analysis of the 6^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 134

A.6 Biological enrichment analysis of the 7^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 135

A.7 Biological enrichment analysis of the 8^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 136

A.8 Biological enrichment analysis of the 9^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 137

A.9 Biological enrichment analysis of the 10^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 138

A.10 List of the neurodevelopment events of mouse brain [6]. Sorted in the increasing PC days order. 139

A.11 List of the neurodevelopment events of mouse brain (continued) [6]. 140

A.12 List of the neurodevelopment events of mouse brain (continued) [6]. Sorted in the increasing PC days order. 141

A.13 List of the neurodevelopment events of mouse brain (continued) [6]. Sorted in the increasing PC days order. 142

B.1 Biological enrichment analysis of the 1^{st} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 144

B.2 Biological enrichment analysis of the 2^{nd} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 145

B.3 Biological enrichment analysis of the 3^{rd} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 146

B.4 Biological enrichment analysis of the 4^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 147

B.5 Biological enrichment analysis of the 5^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 148

B.6 Biological enrichment analysis of the 6^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 149

B.7 Biological enrichment analysis of the 7^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 150

B.8 Biological enrichment analysis of the 8^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 151

B.9 Biological enrichment analysis of the 9^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 152

B.10 Biological enrichment analysis of the 10^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage. 153

B.11 List of the neurodevelopment events of human brain [6]. Sorted in the increasing PC days order. 154

B.12 List of the neurodevelopment events of human brain (continued) [6]. Sorted in the increasing PC days order. 155

List of Figures

Figure **Page**

1.1 **Developmental progress of the neuronal system in the human brain.** (Reproduced from [1]). Starting from the Carnegie stage 11 (29 days) at the bottom-left corner, the neuronal system in the human brain develops in years until delivery (in the center). At the earliest stage, only two embryological structures can be defined: *mesonephric duct* and *sinus venosus*. In the following stages, neurons and connections are growing while more subdivisions appear to form the major regions of the brain. As less as a quarter of its adult size, the human brain is relatively immature at birth and it requires a significantly longer time to develop continually. Commonly, the development of the human brain continues until the age of 10. 4

1.2 **The developmental ontology of mouse brain in the first six levels.** (Reproduced from [2]). The developmental ontology of the mouse brain from Level 00 to Level 05. Starting from the neural plate (NP) at Level 00, neuroanatomical subdivisions are continually generated in an increasing complexity through development. Early proto-segments such as forebrain (F), midbrain (M), hindbrain (H), and spinal cord (SC) are clearly depicted in Level 01 and more subdivisions start to appear in the following levels. The entire developmental ontology of mouse brain contains 14 levels (00-13). 6

2.1 **The atlas of developing mouse brain at three embryonic and three postnatal stages [2]**. Example atlas slides in the sagittal plane chosen from ADMBA: 13.5 days after embryo (E13.5), 15.5 days after embryo (E15.5), 18.5 days after embryo (E18.5), 4 days after born (P4), 14 days after born (P14), and 56 days after born (P56, adult). 19

2.2 **The 26 available regions in the ADHBA data.** At each stage, up to 26 targeted cortical and subcortical regions are examined by the gene probes. These regions contain most of the functional subregions on the cortex as well as the midbrain structures. 21

2.3 **Example visualization techniques in computational bioinformatics. (Reproduced from [3])** (a) A timeline of the relevant technologies and approaches that introduced into bioinformatics. (b) Visualization of $K-Means$ clusters. (c) Visualization of principal component analysis (PCA) in 3D space. (d) Visualization of parallel coordinates of gene expression regarding in seven dimensions. (e) Visualization of the gene expression clusters in the series of time or conditions. (f) Visualization of the hierarchical clusters in the series of time or conditions. (g) Hypothetical integrative visualization approaches using VR (virtual reality) devices. 22

3.1 **Gene expression visualization of Tbr1 using AIBS's Brain Explorer 2.** Spatial visualization of expression of gene Tbr1 at stages: (a) E13.5, (b) E15.5, (c) P4, (d) P14, and (e) P56 with more details. The expression levels are visualized using a heat map, where the highest and lowest values of expression are marked as red and green respectively. 32

3.2 **Constructing the HOS-Tree.** HOS-Tree used a tree-layout graph to represent the structural hierarchy that using the first developing level as the root of the tree. (a) Example annotated structures of first four levels. (b) Constructed structure hierarchical tree derived from (a). (c) Aligning the structures nodes along annular circles in the preferential growth order, the complete HOS-Tree visualizes the entire brain structures. 34

3.3 **Coloring the HOS-Tree using spatial developing orientations.** (a) The entire HOS-Tree used the ontological structure registration. (b) An example of how we converted the spatial developmental orientations into RGB space. 37

3.4 **Example of the ontological structure registration.** (a) Specific structures CSP (caudal secondary prosencephalon) and RSP (rostral secondary prosencephalon) were developed from SP (secondary prosencephalon) at the previous level; (b) Imprecise location of SP, as the black outline shows, without the structure-based registration; (c) Our structure-based registration approach provides the new location of SP, as the black outline shows. . 38

3.5 **Histogram of the distribution of the DO on each direction.** The DO that strong in z direction has much large number when comparing with in x and y directions. 41

3.6 **Comparing the highlighted structures in the forebrain and hindbrain in HOS-Tree.** (a) The forebrain structures showed an expanding developing property. (b) The hindbrain structures had a strong developing direction in the z-direction. 42

3.7 **Gene profile analysis of gene *Sim*2.** (a) For the pattern at stage E13.5, the strongly expressed areas are mostly mid-range-between levels 9 and 11. (b) The pattern with high expression at stage P56 diffuses into late-range levels-between levels 11 and 13. 44

3.8 **Spatial pattern of the expression profile of gene $Tbr1$.** (a) The pattern at stage P14, wider edges are mostly colored as green and red indicating that gradients exist along x-y-axes. (b) The pattern at stage P56 depicting expression energy is much weaker indicating that $Tbr1$ is functionally expressed before stage P56. 46

4.1 **Skewness distributions.**To present the association among genes in rows and structures in columns. Different distribution types of the temporal patterns provide valuable properties for future analyses: positive skewed distribution (blue curve), the negative skewed distribution (red curve), the near-flat distribution with low variance (green curve), and the normal distribution with high variance (black curve). 57

4.2 **The original BGEFM developed in 2013.** The BGEFM visualization approach designed in our previous work [4]. However, the pattern-matching process was used with 5 pre-defined temporal patterns, and lead to biased result. 61

4.3 **Using silhouette properties to choose the best temporal pattern-learning result for the mouse brain.** For the temporal pattern-learning process in the ADMBA data, we repeat the K-Means cluster for each K from 3 to 13. Based on the result, we choose K as 10 for the following work (values are curved using a smoothing spline). 63

4.4 **The Color-encoded Temporal Patterns Learned from the ADMBA dataset.** The ten temporal patterns in the corresponding colors as well as the fitting curves in the encoded colors. The x anix indicates the stages from E13.5 to P56, the y axis is the normalized gene expression values. 64

4.5 **The Color-encoded BGEFM for the ADMBA dataset.** (a) The color-encoded BGEFM, each entry inside indicates the color-labeled temporal association across genes (in rows) and structures (in columns). (b) The brain regions chosen for the visual analytics, which matched the structure bar on the top of BGEFM. (c) A zoomed-in view of the BGEFM, each entry inside is a color-encoded temporal pattern. 65

4.6 **Two example temporal patterns that learned from the ADMBA could be well-interpreted by the neurodevelopment events.** The 1^{st} temporal pattern starts with a peak expression and then decreases until the E18.5 stage. This temporal property highly matched the neuron-related events. The 9^{th} temporal pattern increases from the P4 stage, and this pattern could be interpreted by the myelination-related events. 71

4.7 **Using silhouette properties to choose the best temporal pattern-learning result for the human brain.** For the temporal pattern-learning process in the ADHBA data, we repeat the K-Means cluster for each K from 2 to 20. Based on the result, we choose K as 8 for the following work (values are curved using a smoothing spline). 74

4.8 **The eight temporal patterns learned from the ADHBA.** (a) We plot the temporal patterns that were learned from the ADHBA data. They are sorted into a decreasing-to-increasing order. For each pattern, the box-plot shows the distribution of gene expressional values (data is processed in log_2 space, and normalized into the range of [0,1], as visualized). (b) The 16 filtered temporal stages. 75

4.9 **The BGEFM visual analytic component of the ADHBA data.** (a) The entire BGEFM of the ADHBA data. 1350 filtered genes are in rows and 16 brain regions are in columns. The genes were clustered into 10 groups as indicated using the gene cluster. The cluster dendrogram of brain regions is shown on top. (b) A regional correlation visualization among these 16 regions. (b) 16 brain regions that corresponded to the columns in (a). 76

4.10 **Correlation analysis between all 16 brain regions.** The cerebellar cortex showed the most different patterns to all other regions, and two subsets of regions showed strong distinctions: hippocampus with amygdaloid complex, and striatum with mediodorsal nucleus of thalamus. The correlation also confirmed that genes in the brain regions that are physically close expressed in similar patterns. 83

5.1 **Pipeline of our visual analytics method.** (a) 3D gene expression data of two example genes at six development stages are shown in sagittal view. (b) 3D atlases of brain structures at each corresponding stages from E13.5 to P56. (c) The 3D coordinate of in our method: $< +x >$ represents the left or *lateral* axis (ADMBA data is only available in hemisphere), $< +y >$ represents up or *dorsal* axis, $< -y >$ represents down or *ventral* axis, $< +z >$ represents front or *rostral* axis, and $< -z >$ represents rear or *caudal* axis. (d) The 3D spatiotemporal gradient of two example genes during the entire brain development are shown in sagittal, coronal, and dorsal views. (e) The designed cone-shape visualization approach provided an intuitive presentation of the anisotropy of the gene gradient. 90

5.2 **Temporal expression values of gene *Oxt* at two brain structures.** The stage-varying expression values of gene *Oxt* (oxytocin) at structure preoptic area (red curve) and cerebellar vermis (blue curve). In the preoptic area, gene *Oxt* shows a consistent increasing pattern. In cerebellar vermis, gene *Oxt* shows a flat-increasing-decreasing pattern. 92

5.3 **GA visualization method of the cone-glyph.** (a) The designing of the cone-shape glyph shows how we calculate the height and radius of the cone using GA value. (b) The maximum GA value is 1, which are represented in red color, while the minimum GA value is 0 in green color. (c) A histogram shows the distribution of GA values of the entire 1753 genes in all brain structures. 97

5.4 **Visualization of gene gradient and the anisotropy.** (a) *Cdh*24 shows relatively stronger density in the $< +x, +y, +z >$ direction. (b) An increasing pattern in the forebrain from E15.5 to E18.5 and in the cortex from P4 to P14 can be clearly seen. (c) Shows a strong anisotropic expression with GA=0.7181. 99

5.5 **Visualization of gene gradient and the anisotropy.** (a) *Rreb*1 expands the expression locations but the gradient is low except in the $< +x, +y, +z >$ direction. (b) This expanding behavior can be clearly seen from the lower hindbrain and the midbrain (13.5 and 15.5) to the entire brain regions and specially in the cortex. (c) This behavior leads to a slight diffusional anisotropy with GA=0.3681. 100

5.6 **Visualization of gene gradient and the anisotropy.** *Tubb*3 shows low gradient n the brain which the GA=0.1527. No significant directionality or anisotropy can be detected. 101

5.7 **Visualization of GA in various local regions.** Each column denotes a single brain region while each row denotes a gene: (a) *Cdh*24, (b) for *Rreb*1, and (c) for *Tubb*3. (d) The histogram of GA distribution of all genes in the corresponding region, with the example genes highlighted. 103

Chapter 1: Introduction

As the controller of the mammalian body, the brain is the most complex biological system. Its role is extraordinarily important and valuable for mammalian life since the brain not only coordinates actions and reactions, but also an impetus for serving as a storage of memories [7]. While the anatomy of the brain and its principal functions have been well-studied in the past, studies of the development of the mammalian brain have gained much traction more recently [8, 9]. Although the brains of all mammals develop in similar ways, the entire development is a complex and regulated process that occurs over a period of time [10]. The brain in humans, for instance, develops from a single fertilized egg to a fully functional organ which contains around 100 billion neurons at birth. The biological progresses is so marvelous that it has to generate about 250,000 nerve cells per minute to construct this 3-pounds-only organ, the brain, from the tip of a 3 millimeter neural tube in just 40 weeks [11, 12]. More remarkably, this process continues long into old age. However, although the entire developmental process of the mammalian brain is complex, fast, and long-lasting, it provides the blueprint for the entire development of mammals in biology and helps biologists to gain deeper knowledge of mammalian lives. Thus, a clear and comprehensive understanding of the developmental process of mammalian brains is essential for predicting different neurological conditions during development and to thereby improve mammalian lives.

Over the past few decades, significant advances have sharply accelerated the exploration and provided insights into the basic regulation and processes of the development of the mammalian brain [8]. In developmental neuroscience, numerous studies have yielded remarkable biological insights that have spanned the exploration of the development of the neural organizations at multiple levels: from the structural, to the cellular, and to the molecular levels. This body of work has provided an increasingly clear picture of brain development, such that the entire developmental process should be considered as "the product of a complex series of dynamic and adaptive processes operating within a highly constrained, genetically organized but constantly changing context" [9].

Taking the central nervous system (CNS) as an example, several studies have revealed that it is one of the longest and the most complex developmental procedures during brain growth [9, 12]. In the mammalian brain, neurons—electrically excitable cells that transmit electro-chemical signals to other cells—are the core components and the basic working units [13]. In brain and spinal cord, neurons communicate with another 100 billion neurons, as well as with the other cells, via synapses, hence forming the central nervous system CNS [14]. Starting from a slipper-shaped neural plate at the earliest stages of embryogenesis, the CNS develops into the most complex organ that coordinates and affects the activity of the entire mammalian body [15, 16]. Thus, the development of CNS that consists of various neural organizations is one of the most important and essential developing procedures. Obviously, it is indispensable to grasp this complex biological process in order to comprehend the entire development of the mammalian brain [9, 17, 18].

As stunning as they are in the complexity, the developmental processes of the mammalian brain, however, requires much further study in order to be completely learned. There is no doubt that more questions need to be addressed by biologists and neuroscientists in

the related areas. How is such an intricate system that contains trillions of cells in different kinds constructed in the first place? How are the myriad of functions that the mammalian brain reliably carries out formed during the entire assembly? How is the specificity of various brain functions assigned to certain types of cells or distinct locations during the growth? How does the establishment of brain functions and the developing regions reflect the existing developmental regulation and timeline [12]?

From these questions, the increasing importance for learning the developmental processes, and the so-called developmental regulation in the mammalian brain can be clearly seen. In biology and neuroscience, regulation, as a concept that indicates the "management of complex systems according to a set of rules and trends", is used to determine the program of the development [19]. This program codes the entire developing process of the mammalian brain, from the embryonic stage, through birth, to adulthood. and finally turns it into a complex organ that contains thousands of distinct structures. During the developmental processes in the mammalian brain, multiple biological events and functionalities such as neurogenesis, neuronal differentiation, cell migration, cell differentiation, synaptogenesis, and synaptic transmission are all necessary for providing the shapes and functions to the brain and the CNS. Remarkably, this entire development follows a predetermined program which orchestrates the development synchronously at various levels—from the structural to the molecular levels [20]. The most significant challenge for learning this predetermined program, however, is that the diversity of these biological events and functionalities and at several levels occurs in both space and time simultaneously.

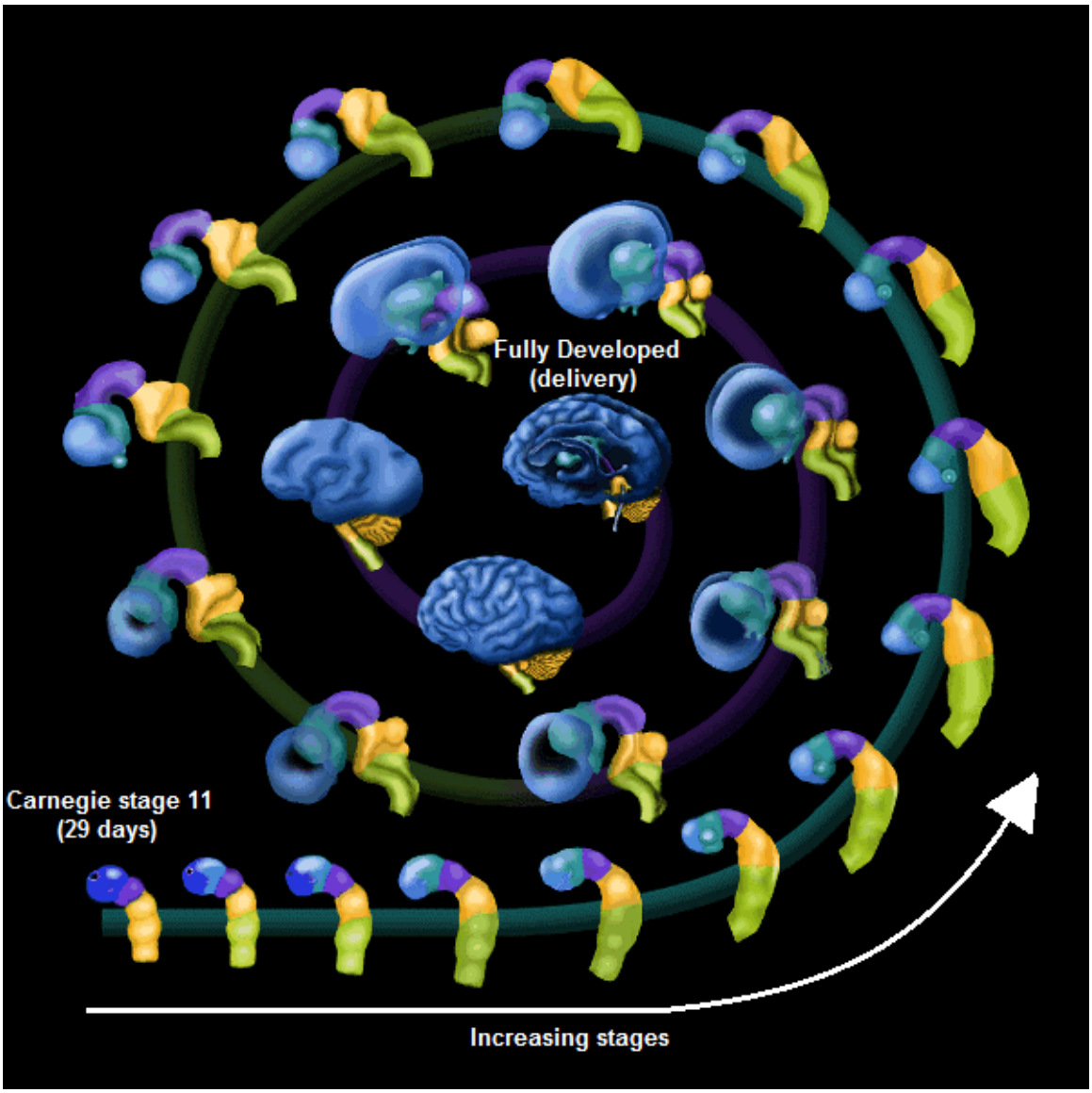

Figure 1.1: **Developmental progress of the neuronal system in the human brain.** (Reproduced from [1]). Starting from the Carnegie stage 11 (29 days) at the bottom-left corner, the neuronal system in the human brain develops in years until delivery (in the center). At the earliest stage, only two embryological structures can be defined: *mesonephric duct* and *sinus venosus*. In the following stages, neurons and connections are growing while more subdivisions appear to form the major regions of the brain. As less as a quarter of its adult size, the human brain is relatively immature at birth and it requires a significantly longer time to develop continually. Commonly, the development of the human brain continues until the age of 10.

In developmental neuroscience, natural but distinctive *developing patterns* are the essential information sought which actually reveal this program [20]. Hence, detecting, exploring, and understanding the developmental patterns in the mammalian brain are of primary importance, and as such, they deserve comprehensive approaches. As described, the mammalian brain is composed of functional regions and distinct cell types. However, both structural anatomy and functionality change uninterruptedly throughout the entire period of growth. Thus, how to learn the salient patterns of these changes becomes the essential point for exploring the comprehensive developmental patterns.

Fortunately, the changes in structural anatomy can be understood by grasping the developmental ontology of the brain structures. This ontology defines the patterns of structural changes through time, location, size, and shape. For instance, Figure 1.2 shows the developmental ontology of the mouse brain from Level 00 to Level 05. Starting from the neural plate (NP) at Level 00, neuroanatomical subdivisions are continually generated in an increasing complexity through development. Early proto-segments such as forebrain (F), midbrain (M), hindbrain (H), and spinal cord (SC) are clearly depicted in Level 01 and more subdivisions start to appear in the following levels [2]. Thus, based on various anatomical and ontological atlases, marked anatomical structures can be tracked precisely, and hence these salient structural changes can be explored by comparing marked anatomical structures at different time epochs, finally converting them into structural patterns. As a result, various structural patterns describing developmental anatomy have been introduced into the field of developmental neuroscience.

In addition, the predetermined program is detailed by genome, which is one of the most important impact factors in organizational development, and the functional changes can be inferred by the change of gene expression [21, 22, 23, 24, 25]. In genetics, expression is the

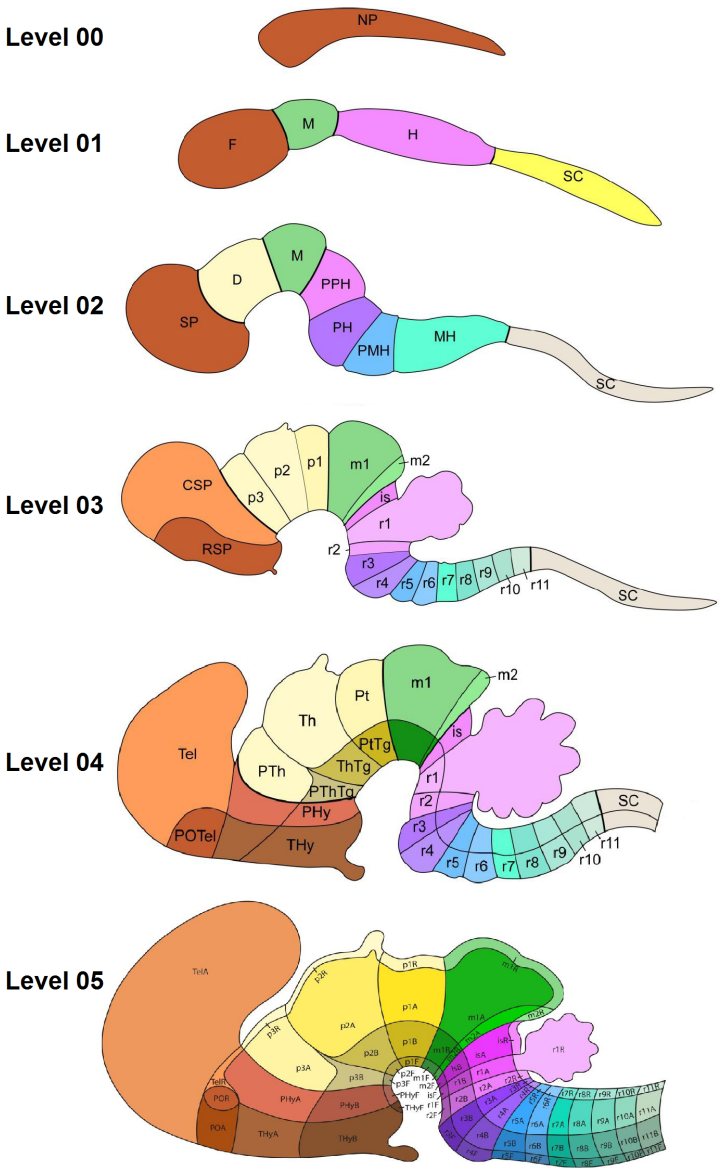

Figure 1.2: **The developmental ontology of mouse brain in the first six levels.** (Reproduced from [2]). The developmental ontology of the mouse brain from Level 00 to Level 05. Starting from the neural plate (NP) at Level 00, neuroanatomical subdivisions are continually generated in an increasing complexity through development. Early proto-segments such as forebrain (F), midbrain (M), hindbrain (H), and spinal cord (SC) are clearly depicted in Level 01 and more subdivisions start to appear in the following levels. The entire developmental ontology of mouse brain contains 14 levels (00-13).

process by which the genetic code of a gene is activated in order to control protein synthesis. According to the *central dogma* of molecular biology by Francis Crick, "DNA can be copied to DNA (DNA replication), DNA information can be copied into mRNA (transcription), and proteins can be synthesized using the information in mRNA as a template (translation)" (DNA and mRNA refer to Deoxyribonucleic acid and messenger Ribonucleic acid respectively) [26]. Thus, the entire expression process controls which information should and will be used in order to construct functional gene products including RNA, hence it gives rise to the phenotypes such as a variety of cells of different shapes and types and organs' functionalities [24, 27]. Thus, gene expression reflects the activation status of gene products such as RNA and targeted proteins and hence the time-varying gene expressions derive the temporal profiles of the organs' biological functions activation status. Therefore, the temporal patterns of gene expression—the developing patterns demonstrating the changes of gene expression over time—delineate the timeline of the biological functionalities of organs and structures in the developing mammalian brain. Thus, the exploration of gene expression over time and structures can significantly enhance the understanding of the development processes of the mammalian brain, and developmental neuroscience has attached great importance to it.

In this dissertation, I will present an extensible visual analytics framework for the exploration of the comprehensive patterns of gene expression in the developing mammalian brain. In this chapter, I will first elaborate on the motivation and aim of this work in Section 1.1. The thesis statement is provided in Section 1.3, and I present proposed solutions in Section 1.4. Finally, the layout of this entire dissertation organization is presented in Section 1.5.

1.1 Motivation

Although numerous studies in this area have introduced more developing patterns, neuroscientists have focused mainly on two forms as mentioned above: structural patterns and temporal patterns [28, 29]. In 2013, the IEEE Scientific Visualization Contest focused on the exploration of the various patterns of gene expression in the developing mouse brain [30]. In order to help the neurosciences to investigate the spatial and temporal changes of gene expression during the brain development, several question were posed:

1. "**Gradient Identification:** which genes exhibit directional expression patterns? For example, in the cortex, rostral-to-caudal gradients indicate involvement with development, whereas exterior-to-interior gradients are involved with layering. Which categories do these genes belong to?

2. **Structural Patterns:** which genes show strong expression in a small set of structures but little expression elsewhere? How do these patterns change throughout development? Which categories do these genes belong to?

3. **Structure Consistency:** which structures have the most consistent expression patterns over time? Which structures are the least consistent? Which structures most resemble each other? This is particularly interesting if the structures are not neighboring. How do the answers change if you restrict the expression patterns to a single gene category?

4. **Complementary Patterns:** which genes have expression patterns that complement each other within a structure? Are these patterns persistent during development? Which categories do these genes belong to?" [30]

Clearly, the various patterns regarding the questions above are not always presented consistently during the entire developmental process, and more importantly, are potentially mutually influenced and restricted. The structural patterns, for instance, not only reveal the developmental behaviors of organs at various brain locations but also provide the anatomical ontology of structures at various time epochs or stages. The temporal patterns (refer to the Structure Consistency), on the other hand, represent the durative activation states of genes while the expression profiles at each stage also affect the structural functionalities and development. Moreover, for any single gene, its current gene expression status depends on its previous state strictly in both time and location during brain development. Thus, the overall patterns that have integrated both structural and temporal patterns could exhibit regional genomic behavior through time. Hence an integrative approach for these various patterns is of primary importance and amenable for exploration and analysis.

Spatiotemporal data and patterns, which are used to represent the objects in both space and time, have attracted considerable attention in the previous decade [31, 32, 33, 34, 35, 36, 37]. As a broad concept (as long as the data contains both spatial and temporal information), spatiotemporal data is worth its weight in gold in many fields, and especially in developmental neuroscience [20, 38, 39, 40, 41]. As mentioned above, gene expression data can reflect distinct biological events and functionalities. Thus, for genes, their time-varying properties of expression at various brain regions imply the temporal patterns of regional functionalities. At the same time, their space-varying properties of expression also establish the spatial patterns which indicate the evolutions of these functionalities among brain regions. Obviously, in order to gain a deeper and clearer understanding of the development of the mammalian brain, it is important to explore the natural spatiotemporal patterns of gene expression, hence to answer the questions of the contest.

1.2 Challenges and aims

In computer science, data mining is an interdisciplinary subfield which focuses on exploring interesting and valuable patterns from large and complex data. Advancing significantly in biology and neuroscience, plentiful approaches in data mining have been developed in order to discover various types of patterns [42, 43]. Supervised learning (includes classification and regression), clustering, dimensionality reduction, structured prediction, and anomaly detection are common classes of techniques that are involved in data mining [44, 45]. Each of them can serve as modeling or knowledge exploration approach for targeting data types and tasks. Thus, how to choose proper approaches to thereby design a robust pattern-learning methodology for revealing various types of patterns is a prerequisite for the exploration of the spatiotemporal patterns.

Furthermore, after learning patterns in various types, integrating them into a data structure which represents spatiotemporal characteristics is the next challenge. Such a data structure can not only enhance the entire visual analytics significantly, but also provide an efficient and extensible platform that could significantly advance related studies. Since the explored patterns contain region-varying expression information among brain structures as well as the time-varying expression information over the entire development, this data structure deserving of innovational approaches should be robustly constructed in order to represent the spatiotemporal patterns effectively. Obviously, how to design such a data structure, and consequently, to derive the spatiotemporal patterns also challenge the exploration progress. Finally, whether the interpretation of these various types of patterns reflects the biological accomplishment or leads to future hypotheses still needs investigation.

On the other hand, understanding the comprehensive developing process requires not only insightful and efficient analysis approaches to explore, but also interactive and intuitive visualization methods to present. "Visual representations and interaction techniques take advantage of the human eye's broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once. Information visualization focused on the creation of approaches for conveying abstract information in intuitive ways" [46]. Visual analytics, on the other hand, focuses on the visualization process in the entire data analysis as well as the user interaction, and it is widely used in the studies of biological data analysis [47, 48]. However, although several visualization tools and platforms targeting the comprehensive pattern presentation have been designed, none of them have satisfied the requirements of the exploration of the spatiotemporal patterns in developing neuroscience (detailed in Chapter 2). How to design such an eligible visual analytics tool is thus the fourth challenge that I will undertake in this dissertation.

Given the challenges above, I propose an extensible visual analytics framework in order to facilitate the comprehensive spatiotemporal pattern exploration of gene expression in the developing mammalian brain. In this dissertation, I will focus only on two species of mammalian brains: mouse and human. The data I have used were collected from the Allen Developing Mouse Brain Atlas (ADMBA) project and the Allen Developing Human Brain Atlas (ADHBA) project, both of which contributed by The Allen Institute for Brain Science (AIBS) (detailed in Chapter 2.1). In summary, this dissertation will address the above challenges in studies of the following specific aims:

Aim 1. Develop an interactive visual analytics method to present spatial information of structures, based on the AIBS datasets for murine models.

Aim 2. Develop a robust method for learning and visualizing the inherent temporal patterns of gene expression in developing mammalian brains, based on the AIBS datasets for both mouse and human brains.

Aim 3. Define an integrative method for providing the intuitive presentation of the spatiotemporal pattern of gene expression in the mouse brain.

In my approach to addressing **A**im 1, I will present a tree-layout visual analytics system to present the spatial pattern of brain structural development. To address **A**im 2, I propose a matrix-layout system to perform robust visual analytics of region-based temporal patterns of gene expression. Finally, for **A**im 3, I will present a gene gradient approach to integrate various patterns into the spatiotemporal pattern. It should be noticed, while this dissertation is mainly focused on the exploration of spatiotemporal patterns for the AIBS data, it can serve as an efficient and robust visual analytic solution for other developing brain data or other multidimensional data as well.

1.3 Thesis statement

The mammalian brain develops through a predetermined program which contains roles that are concurrent in of both space and time. The spatiotemporal patterns of the developing mammalian brain reveal novel biological insights from comprehensive data, to explore these roles. An integrative visual analytics framework enables effective and precise exploration of such spatiotemporal patterns, and overcomes existing challenges in related fields. The explored patterns can reveal valuable developmental processes of the brain as well as roles that can enhance the understanding of mammalian brains.

1.4 Outline of Solutions

My overall approach offers an efficient visual analytic framework which allows for various explorations of the spatiotemporal patterns of gene expression in the developing brain. The proposed framework used three operating modules of the visual analytics to overcome the existing challenges and achieve the aims above—spatial pattern exploration, region-based temporal pattern exploration, and integrative gene gradient-based spatiotemporal pattern exploration. Next, I briefly describe each of the proposed modules.

- **HOS-Tree—Hierarchical Orientation Structural Tree**

  I have developed the HOS-Tree, a visual analytics approach which allows the user to explore the spatial patterns of structural development. The HOS-Tree utilizes prior knowledge of hierarchical ontology to generate a tree-layout visualization of brain structures. The edges in the tree are color-coded to reveal the growth directionality of brain structures during development. At the same time, gene expression can be integrated with the tree-layout hierarchy to indicate the spatial change of developing structures. I will demonstrate the usefulness of the HOS-Tree approach through two case studies. Thus, in summary, the HOS-Tree system uses a tree-layout visualization to present the structural hierarchy, and uses color to indicate the developing orientations, hence to serves as the exploration module for the spatial pattern.

- **BGEFM—Bi-clustered Gene Expression Flow Matrix**

  I have developed the BGEFM, a visual analytics approach which focuses on the association between gene and structure in a temporal manner. This method uses a data-mining approach to learn the inherent temporal patterns of gene expression, and then presents the patterns in a matrix format for all given genes and structures.

Since bi-clustering, a data mining approach can seek the correlative signals between the rows and columns of a matrix (refer to genes and structures domain, detailed in Chapter 4), we apply such approach to the generated matrix in order to explore the correlated region-based temporal patterns of gene expression. Also, I use neurodevelopment events to interpret the learned temporal patterns, while using the enrichment of biological function to investigate the region-based temporal patterns. In summary, the BGEFM uses data-mining approaches for providing an interactive pattern-presentation among genes and structures, hence to serve as the exploration module for the region-based temporal patterns.

- **Spatiotemporal Gradients of Gene Expression**

  The gene gradient is a proposed data structure that measures how a gene expression changes across the developmental structures in the brain, thus representing the spatiotemporal patterns. I use two measurements to indicate the gradients of gene expression: the main gradient orientation (MGO) and gradient anisotropy (GA). Also, I developed a visualization approach to present the gene gradient in an intuitive way. Again, I have investigated the functional enrichment of genes which have significant and distinct gradient properties and demonstrate how the spatiotemporal pattern reflects the brain development. Therefore, in summary, I use the gradient of gene expression to define a data structure which derives spatiotemporal genomic characteristics, as well as provides an intuitive visualization of the 3-D coordinates, hence to furnish the exploration of the spatiotemporal pattern.

As described above, for each module, I have investigated the result by seeking the relations between patterns and biological properties. The analyses show that several explored patterns are well supported by the preliminary knowledge of brain development. Thus, I

assert that the proposed visual analytic framework demonstrates an excellent precision of gene characteristics during brain development. Therefore, it enables the efficient and robust exploration of spatiotemporal patterns of gene expression in the developing brain for both mouse and human. Last but not least, although not all explored patterns are well-explained by existing knowledge, I still firmly believe, however, that they potentially lead to future studies in developing neuroscience fields.

It should be noticed, once again, that although my proposed framework focuses on the spatiotemporal pattern exploration for both the ADMBA and ADHBA data, it could also serve as an efficient and robust visual analytic solution for other multidimensional data for comprehensive pattern exploration.

1.5 Organization of this dissertation

The remainder of this dissertation is organized as follows. In Chapter 2, I will review existing work in the presentation of spatial and temporal patterns, as well as the related works which have focused on ADMBA and ADHBA data visualization and analysis. In Chapters 3 and 4, I will describe the HOS-Tree and BGEFM respectively. In each chapter, I will first detail the design rationales of the visual analytics modules, and they illustrate the robustness and efficiency by investigating interesting patterns. The gene gradient-based spatiotemporal pattern exploration will be described in Chapter 5. Finally, in Chapter 6, I will summarize this dissertation and discuss future works on this topic. Again, although my proposed framework focuses on the spatiotemporal pattern exploration for both the ADMBA and ADHBA data, it can also serve as an efficient and robust visual analytic solution for other multidimensional data for comprehensive pattern exploration.

Chapter 2: Related Works

With rapid growth during recent years, the exploration of overall genomics patterns in mammalian brain has become increasing valuable in biological and neuroscience studies. In this chapter, before the discussion of related studies, I will first detail the data collected from the AIBS projects in Section 2.1. Next, I will survey studies that are most related to the scope of my topic in two categories. First, in Section 2.2, I will review related techniques in high-dimensional data visualization, and especially, in spatiotemporal pattern observation. Following this, I will review salient studies which focus on the topics of observing expression patterns in brains, especially for the AIBS data portal, in Section 2.2.1. Finally, the chapter ends with a brief summary in Section 2.3.

2.1 Allen Brain Neuroscience Datasets

In order to understand how the human brain works in states of health and disease and to answer various challenging questions in neuroscience, the Allen Institute for Brain Science (AIBS) has published an online resources data portal that provides extensive gene expression and neuron-anatomical data [49, 50]. This public data-driven resource enables the discovery and exploration of fundamental brain properties by scientists and researchers. I will introduce both the data contained in the Allen Developing Mouse Brain Atlas and Allen Developing Human Brain Atlas.

2.1.1 The Allen Developing Mouse Brain Atlas

As one of the discovery projects of AIBS, the Allen Developing Mouse Brain Atlas (ADMBA) focuses on providing a detailed description of the various gene expressions in the developing mouse brain. To complete such a comprehensive genome-scaled mouse brain atlas, high-resolution (approximately $20\,\mu m^3$ per voxel), deep-level (approximately 2600 structures, for a total of 14 levels of hierarchical structural ontology) anatomic reference atlases at significant growing stages were generated. Figure 2.1 shows example slides in the sagittal plane chosen from the ADMBA. At the same time, the entire structural ontology during brain development from embryo to adult was provided. By focusing on the neuroanatomical and signaling pathways of targeted genes, the ADMBA project has enabled discovery of the change from neuron development-related functions for the field scientist. This complete neuron development-targeted dataset was provided in the form of the gene-structure pair, where the structures followed a developmental hierarchy. Described in detail, the collected Allen Developing Mouse Brain Atlas data consists of:

- Six developing stages: three embryonic and three early postnatal Theiler Stages are selected: 13.5 days after embryo (E13.5), 15.5 days after embryo (E15.5), 18.5 days after embryo (E18.5), 4 days after born (P4), 14 days after born (P14), and 56 days after birth (P56, adult).
- 2691 structures with hierarchical structure ontology.
- 2105 neuron development-targeted genes.
- 3-D structure annotation information, one for each stage.
- Detailed anatomic reference atlases, one for each stage.

- High resolution *in situ* hybridization image data across all stages.

In summary, the ADMBA data uses the genes-per-structure approach to provide a characterization of gene expression in the brain from mid-gestation to young adult mouse. At the same time, each developmental stage has its reference structure annotation volume, and structure labels are assigned to every voxel in the brain. Since the gene expression values were derived from *in situ* hybridization of brain section images, every pixel has been registered as an appropriate reference structure. Thus, the expression data of ADMBA is multidimensional, and for any given gene, a user can access the detail expressing values in entire brain locations across all developing stages. Finally, the structural annotations follow a hierarchical ontology of the evolving brain.

2.1.2 The Allen Developing Human Brain Atlas

Similar to the ADMBA project, the Allen Developing Human Brain Atlas (ADHBA, or BrainSpan) project was introduced for describing transcriptional mechanisms involved in the developing human brain. However, in contrast with the ADMBA data, in which genes are chosen for neuroanatomical functions, the ADHBA data examines almost all known probes, in order to provide a complete data portal for field scientists. In the ADHBA data, two data modalities were used for gene examination, including RNA sequencing and exon microarray hybridization. Also, the expression data were generated based on 42 brain samples, including both males and females, and spanning the prenatal and postnatal stages of brain development. For each brain specimen, up to 26 targeted cortical and subcortical structures were examined by more than 52,000 gene probes. Thus, described in detail, the collected Allen Developing Human Brain Atlas data consists of:

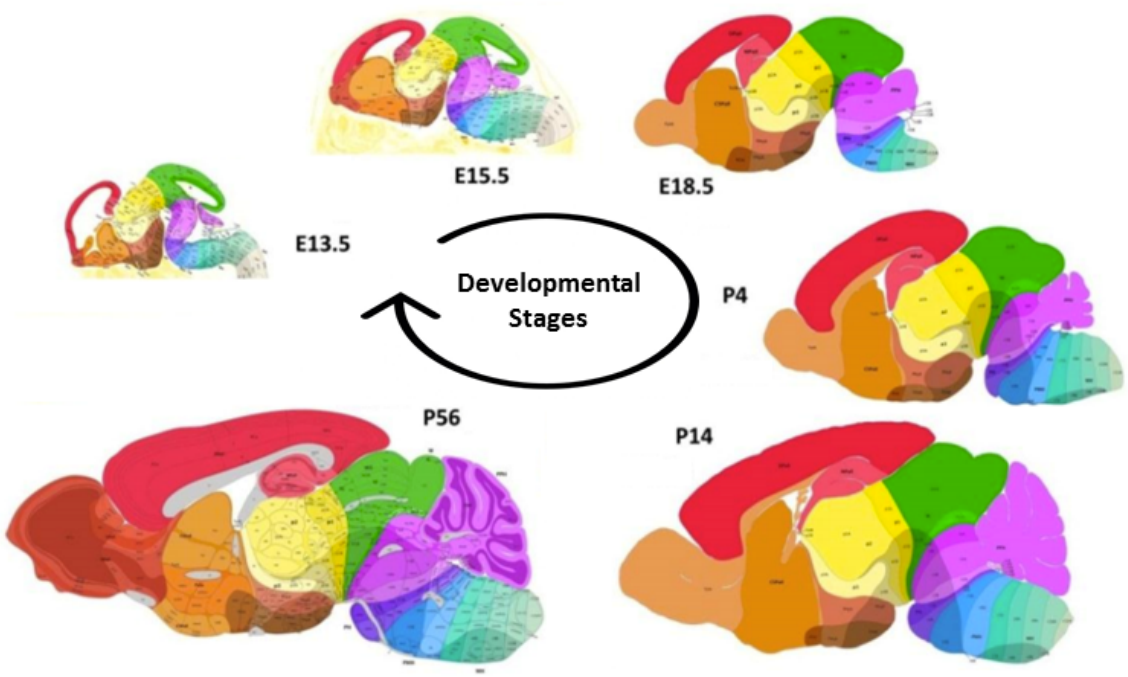

Figure 2.1: **The atlas of developing mouse brain at three embryonic and three postnatal stages [2]**. Example atlas slides in the sagittal plane chosen from ADMBA: 13.5 days after embryo (E13.5), 15.5 days after embryo (E15.5), 18.5 days after embryo (E18.5), 4 days after born (P4), 14 days after born (P14), and 56 days after born (P56, adult).

- 42 brain specimens, spanning prenatal and postnatal development in both males and females.
- All brain specimens were distributed into 31 developing stages.
- Up to 26 targeted cortical and subcortical regions across human brain development.
- 52,376 genes examined in each of the accessible regions.

In summary, the ADHBA data used the genes-per-region approach to provide gene characterization in the human brain from early-gestation to adult. Each available region was examined by entire gene probes for each brain sample, while these brain samples belong to 31 developing stages. Thus, as with the ADMBA, the expression data of the ADHBA is multidimensional. For any given gene, users can access the detailed expression values in available brain regions throughout all developing stages. It should be noticed, however, that not all regions were available for examination in each brain specimen. Thus, the ADHBA data contains more empty data entries than the ADMBA.

2.2 Visualization Approaches for Spatial, Temporal, and Hierarchical Patterns

Visualization can facilitate the discovery and exploration of structures, patterns, and associations among data, and can enable users to gain an intuitive understanding of the data structure. More importantly, qualitative studies outweigh others in the early stage of data analysis, in which analytical visualization is especially significant. Certainly, gene expression pattern analysis and visualization are challenging topics due to their dimensionality, noisy environment, and pattern varieties.

Figure 2.3 shows well-known approaches of visualization and analysis targeting gene expression data [3]. Among the studies using these visualization approaches, numerous of

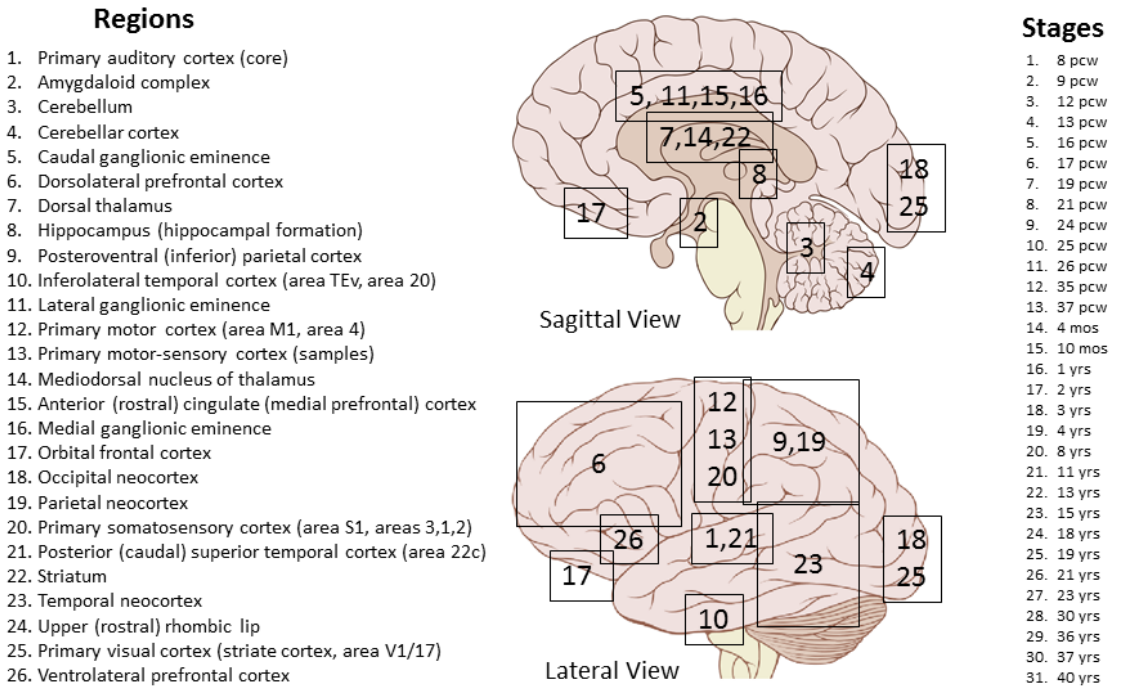

Figure 2.2: **The 26 available regions in the ADHBA data.** At each stage, up to 26 targeted cortical and subcortical regions are examined by the gene probes. These regions contain most of the functional subregions on the cortex as well as the midbrain structures.

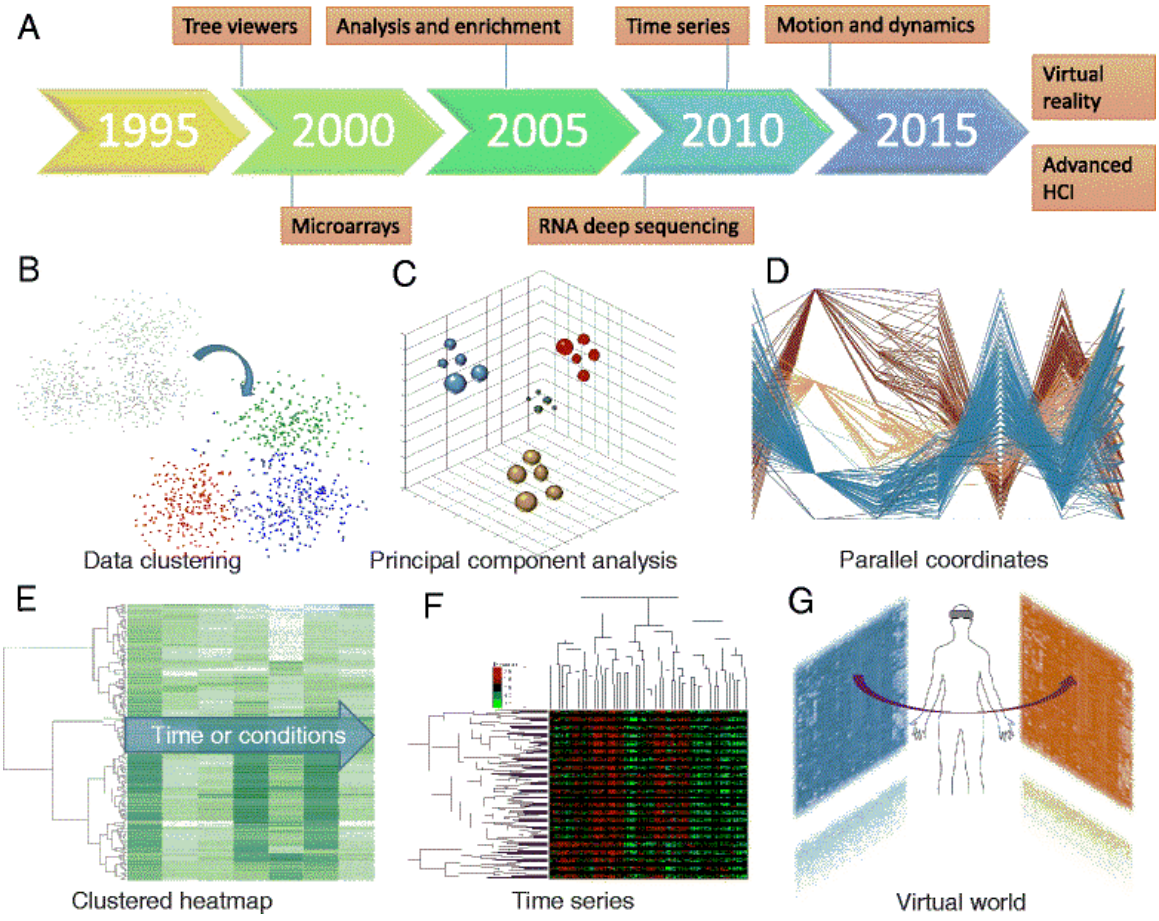

Figure 2.3: **Example visualization techniques in computational bioinformatics. (Reproduced from [3])** (a) A timeline of the relevant technologies and approaches that introduced into bioinformatics. (b) Visualization of $K-Means$ clusters. (c) Visualization of principal component analysis (PCA) in 3D space. (d) Visualization of parallel coordinates of gene expression regarding in seven dimensions. (e) Visualization of the gene expression clusters in the series of time or conditions. (f) Visualization of the hierarchical clusters in the series of time or conditions. (g) Hypothetical integrative visualization approaches using VR (virtual reality) devices.

them exist on the observation of temporal patterns of gene expressions [51, 52, 53, 54]. Clustering methods is one of the most widely used analysis methods in the gene expression related field [55]. By clustering, genes which co-express or share similar expression patterns are grouped together. Further, as a dimension-reduction method widely used in bioinformatics, self-organizing maps (SOM) provide a visualization approach which projects high-dimensional expression data into a low-dimensional viewpoint (usually in 2-D space) [56]. On the other hand, among visual analytic methods, the parallel coordinates (PC) can potentially be considered as the simplest approach to displaying the various gene expression patterns [57]. In PC, a separate axis represents each dimension, and data are linked between the axes in order to detect the relationships. In this case, PC is well-known and widely-used in temporal pattern visualization in the biological field [58]. At the same time, TreeView is one of the most popular approaches to visualizing the hierarchical patterns. The tree-layout presentation reflects the relationships between clusters by a hierarchical dendrogram [59].

2.2.1 Pattern Exploration for the Allen Brain Atlas

Understanding the mammalian brain is significant in order to learn the regulation of life. In order to explain the organization of the mammalian brain, several studies have focused on identifying developmental patterns in the brain. The recognition of developmental patterns in the brain can be obtained from the gene expression. Abundant approaches and techniques have targeted the topic of analyzing and profiling the expression data in genome-wide range [60, 61]. Recently, several studies have focused on identifying the spatial gene expression patterns in the particular mammalian brain regions [62, 63, 21, 64]. At the same time, ABIS has provided a genome-wide atlas of gene expression in the brains of many species. These valuable datasets were accessible by the public by using the ABIS data portal. Especially,

among these different datasets, both the developing mouse brain atlas and the developing human brain atlas were put under the spotlight since they provided comprehensive data for scientists in developing neuroscience and bioinformatics [65, 66, 67, 68, 69, 70].

Studies of the Mouse Brain

Among studies that have been performed for the AIBS datasets, multiple studies have focused on the mouse species datasets, including the ADMBA and AMBA (Allen Mouse Brain Atlas), with various scopes of targets. Some have focused on providing a global comprehensive pattern exploration, including the gene expression pattern [71, 72] and the connection networks [73, 74, 75, 76, 77], while others have performed gene-targeted analyses [78, 79, 80, 81, 82, 83] or region development analyses [84, 85, 86, 87, 88]. Among the abundant studies in this field, several works have targeted the exploration of the general patterns in the developing mouse brain [89, 90, 91, 92, 93]. Especially, Ji's lab made a significant contribution on this topic. They used a tree-layout structure to denote the structures, similar to mine, while using HOSVD (High-order singular value decomposition) to complete the data dimension-reduction and to perform visualization [4, 94, 95]. However, in spite of this accelerated spatiotemporal pattern exploration, there are still limitations to their progress. First, although they used a tree-layout structure to present the developing hierarchy, compared with my work, it did not manifest enough spatial information. And second, the spatiotemporal patterns observed were region- or structure-based, and lacked a more global and integrative spatiotemporal pattern definition and observation.

Studies of the Human Brain

On the other hand, more valuable work has been performed targeting the human species datasets, including the ADHBA and AHBA (Allen Human Brain Atlas). From these studies, some have proposed the limited solutions for the global gene expression network and region-connectivity challenges [96, 97, 98, 99, 100, 101, 102, 103], while others have focused on distinct targets. For example, several types of research have focused on local function and development in the human brain and have found distinct relationships between brain structures and cell types during growth [104, 105, 106, 107, 108, 109, 110]. Moreover, some studies tried to identify the developing differences between human and other species, and found that the differences came from not only biological function but also anatomical structures [111, 112, 113]. More recent studies also focused on diseases analysis based on the AHBA data to provide a better understanding of their impact on the brain [114, 115, 116, 117, 25]. Since the majority of the brain is the neuron system, only the neurodevelopment-related diseases were chosen, such as autism disorders and Parkinson's disease. However, a global spatiotemporal pattern exploration approach is still missing for the developing human brain data.

2.2.2 Released Tools for AIBS Data Portal

At the same time, constructed by the AIBS data portal and the above studies, extensive data exploration tools have already been developed. Brain Gene Expression Analysis (BGEA) provides computational techniques for gene expression analysis [118, 119]. NeuroBlast, which is derived from the principles of the Basic Local Alignment Search Tool

(BLAST), is used to identify similar genes based on sequences. In NeuroBlast, the expression patterns can be observed in 3-D space, and users can also explore similar genes by generating a ranked list based on the similarity of expression patterns [120]. Brain Explorer, constructed by AIBS, serves as a visualization tool for visualizing mouse brain anatomy and gene expression data. It provides a 3-D rendering of mouse brain structures and the gene expression patterns [2, 121]. Anatomic Gene Expression Atlas (AGEA), on the other hand, was built based on computed spatial correlations across gene expression data [122]. AGEA offers users the exploration of the genome-wide neuroanatomical relationships as well as molecular-level brain organization. ALLENMINER is an open-source software that provides the 3-D exploration of mouse brain data [123]. Enhanced to the Brain Explorer, ALLENMINER enables the custom regions of interest (ROI) analysis for gene expression and structure. Finally, in 2013, our team developed HOS-Tree, a visual analytic system that enables the exploration of the spatial expression pattern over a structure-wide hierarchical layout [4].

2.2.3 Existing Challenges in Capturing Spatiotemporal Gene Gradients

Although this body of work in visualization has provided various approaches for revealing various types of gene expression patterns, few of them have focused on the topic of the exploration of the spatiotemporal pattern and so called the gene gradient [4, 124, 122]. While several studies have provided limited discoveries for specific diseases and genes [125, 126, 127], few of them have concerned the spatiotemporal pattern across the entire development of mammalian brains [23, 70, 100]. Finally, although our preliminary work provided the exploration of the spatiotemporal fold change of gene expression in the

developing mouse brain, no reported work has provided a visualization approach which facilitates the comprehensive visualization of the spatiotemporal gene gradient [128].

2.3 Summary

This chapter reviews the related works in the fields of visual analytics of AIBS datasets. I started by reviewing the related technologies and proposed methods for comprehensive pattern visual analytics, and followed with a discussion of previous studies for the AIBS data portal. I listed the advantages and limitations of several interesting approaches. In the next three chapters, I will present three visual analytics components for my proposed visual analytics framework. In each chapter, I will discuss the visual techniques that I have chosen, and will use explored patterns to perform investigation.

Chapter 3: Exploration of the Spatial Pattern of Brain Development

In this chapter, we present an interactive visual analytics method to explore the spatial patterns in the AIBS datasets. Especially, in 2013, we developed HOS-Tree (Hierarchical Orientation Structure Tree) to participate the 2013 IEEE Data Visualization Contest [4]. HOS-Tree focuses on the spatial pattern visualization of both developing orientations among all brain structures and gene expression inside. In order to provide an intuitive and efficient way, the HOS-Tree utilizes the prior knowledge of the hierarchical ontology to generate a tree-layout visualization for the brain structures. This data-driven visual analytics overcomes the first challenge in Chapter 1.1: how to enable the learning of the natural spatial patterns of structures in the developing brain.

To satisfy one of the core prerequisites for spatiotemporal pattern exploration, we designed HOS-Tree. By describing each structure as a node and the development relationship as edges, the HOS-Tree used a tree-layout representation to indicate the entire hierarchical ontology in an intuitive way. More important and innovational, the edges in the tree are color encoded to reveal the growth directionality during structural development. At the same time, the gene expression can be integrated with the tree component to indicate the spatial change over developing structures. We demonstrated the efficiency of HOS-Tree through two case studies.

The chapter is structured as follows. Section 3.1 provides background and motivations. Section 3.2 describes the data collection and tasks. Section 3.3 describes the design of the HOS-Tree system. Next, we demonstrated the efficiency of HOS-Tree through two case studies in Section 3.4. We close this chapter by summarizing our work in Section 3.5. It should be noticed, since the ADHBA data did not provide the spatial region information, we only perform the HOS-Tree to the ADMBA data, which contains entire structure volume and location information in 3-D space.

3.1 Background and Motivation

Developmental neuroscience is a subfield of neuroscience which records and analyzes changes in the developing brain from embryogenesis to end-of-life. The development of the mammalian brain provides useful and valuable information for neuroscience in the form of patterns of changes in structure and function. Patterns, in turn, serve as essential information to characterize brain development. Therefore, it is pivotal and critical to observe, identify, visualize and analyze the inherent patterns of change that can be used to infer genomic regulation. A systemic and comprehensive understanding of the rules during brain development will lead neuroscientists to biologically relevant insights and treatment strategies of neurological disorders and diseases.

Several studies have demonstrated that underlying genomics has the most impact on the growth of the mouse brain. Until recently several efforts has been expended on studying spatiotemporal gene expression (mRNA) patterns; developmental neuroscience has traditionally focused on either the underlying spatial or temporal patterns. Understanding spatiotemporal patterns of gene expression will help clarify the roles that various genes play

during brain development. Thus, understanding the spatial pattern of developing structures becomes more important.

Data repositories from the Allen Institute for Brain Science (AIBS) inherently captured structures and associated changes in gene expression in both adult and the developing brain at numerous stages. It now remains to discover these associations. The topic of the 2013 IEEE Data Visualization Contest targeted the existing domain of developmental neuroscience on the mouse brain. It is essential to record and characterize the significant patterns of mouse brain organization and development and to analyze their relationships to gene expression patterns. However, to aid neuroscientists to explore the evolution and regulation of such complex processes and to help grasp the multitude of gene expression-structural relationships, it requires not only a complete and comprehensive data processing system but also an interactive and efficient visualization approach.

In order to observe, identify, visualize and analyze the comprehensive patterns of gene expression during brain development, we created a visual analytic system based on the ADMBA dataset. Our system was crafted in response to the spatial pattern exploration challenges posed by the contest. We next describe the posed challenges and specifically describe the required analytic tasks.

3.2 Data and Tasks

The Allen Institute has organized several salient studies that capture and curate a large number of genes in the developing brain and made this data publicly available. It is now feasible to characterize various relationships between gene expression and structural patterns over space and time. In this section, we briefly describe the available data set, i.e., the

Allen Developing Mouse Brain Atlas or the ADBMA, as well as the tasks and challenges posed by the 2013 IEEE Data Visualization Contest [4].

3.2.1 Data Collection

The ADMBA provides the expression levels of 2105 genes in 2691 anatomical structures across three embryonic and three early postnatal stages: E13.5 — embryonic days 13.5, E15.5 — embryonic days 15.5, E18.5 — embryonic days 18.5, P4 — postnatal days 4, P14 — postnatal days 14, and P56 — postnatal days 56 (adult). Three-dimensional reference mouse brain atlases for each stage and the corresponding annotated volume with structure labels are also available. Additionally, a hierarchical annotation of various evolving brain structures is available. Finally, the scrutinized genes are organized into 11 categories allowing for further appropriate enrichment studies.

3.2.2 Tasks

Based on the data provided by the Allen Institute, the 2013 IEEE Scientific Visualization Contest posed several tasks and challenges, and two of them were targeting on the spatial pattern of structures.

Task 1. **G**radient Identification. Which genes exhibit directional expression patterns? Which categories do these genes belong to?

Task 2. **S**tructural Patterns. Which genes show strong expression in a small set of structures but little expression elsewhere? How do these patterns change throughout development?

Indeed, the characterization of such complex structural patterns requires not only an efficient visual analytic approach to representing the entire developmental process but also

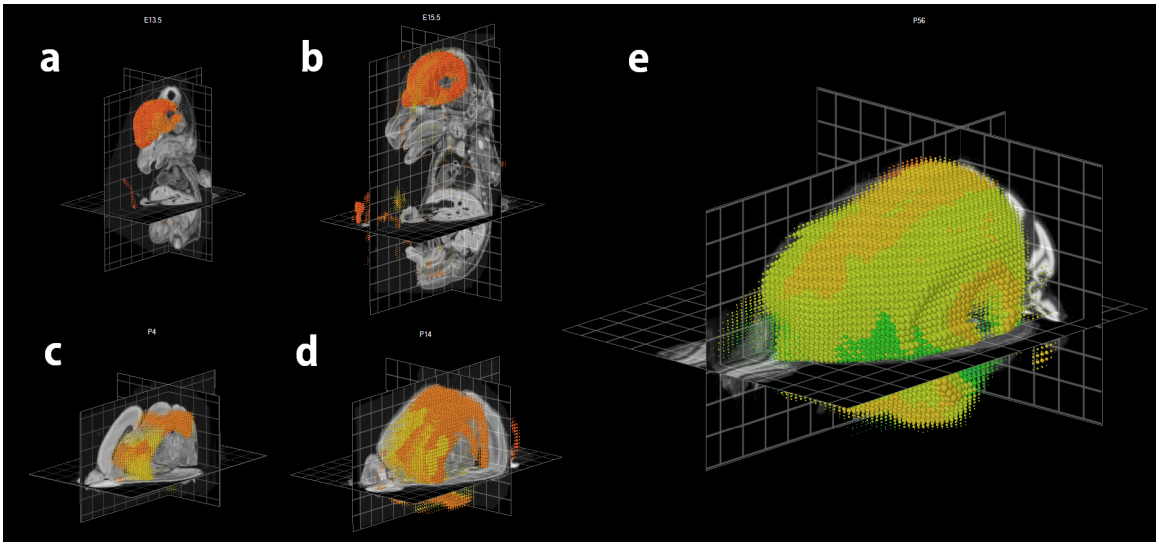

Figure 3.1: **Gene expression visualization of Tbr1 using AIBS's Brain Explorer 2.** Spatial visualization of expression of gene Tbr1 at stages: (a) E13.5, (b) E15.5, (c) P4, (d) P14, and (e) P56 with more details. The expression levels are visualized using a heat map, where the highest and lowest values of expression are marked as red and green respectively.

a flexible interrogative method to enable effective interactive queries. The Allen Institute designed a visualization tool, Brain Explorer 2, to explore gene expression levels and mark their 3-D locations in the mouse brain at a given stage of development. Figure 3.1 shows the visualization pertinent to the expression of gene *Tbr1* by using Brain Explorer [121]. However, this tool is too limited to observe spatiotemporal patterns hidden in the data or to complete any of the two tasks listed above. These tasks can only be realized when significant patterns in various phenotypes are exhumed from data collected. Therefore, to explore the spatial patterns in a developing mouse brain and to address the challenges posed by the contest, we created a visual analytics system described in some detail in the following section.

3.3 Visual Analytics Design

In order to address the challenges posed in the tasks above, we created a visual analytic solution that rests on the Hierarchical Orientation Structural Tree or HOS-Tree. This component organizes the brain anatomy as an oriented tree and allows for the visualization of the development process in an intuitive manner. Eventually, the proposed visual analytics framework integrated all these components (including the component in Section 4) together interactively and subsequently the developmental genomic patterns in a mouse brain could be revealed. The structure-gene cluster patterns found in the second component (refer to the BGEFM in Section 4) are used to drive the visualization of the patterns during development and hence the HOS-Tree. Further, a collection of gene expression patterns can be traced back to various anatomical structures. Here, we first describe the designing of the first component, the HOS-Tree.

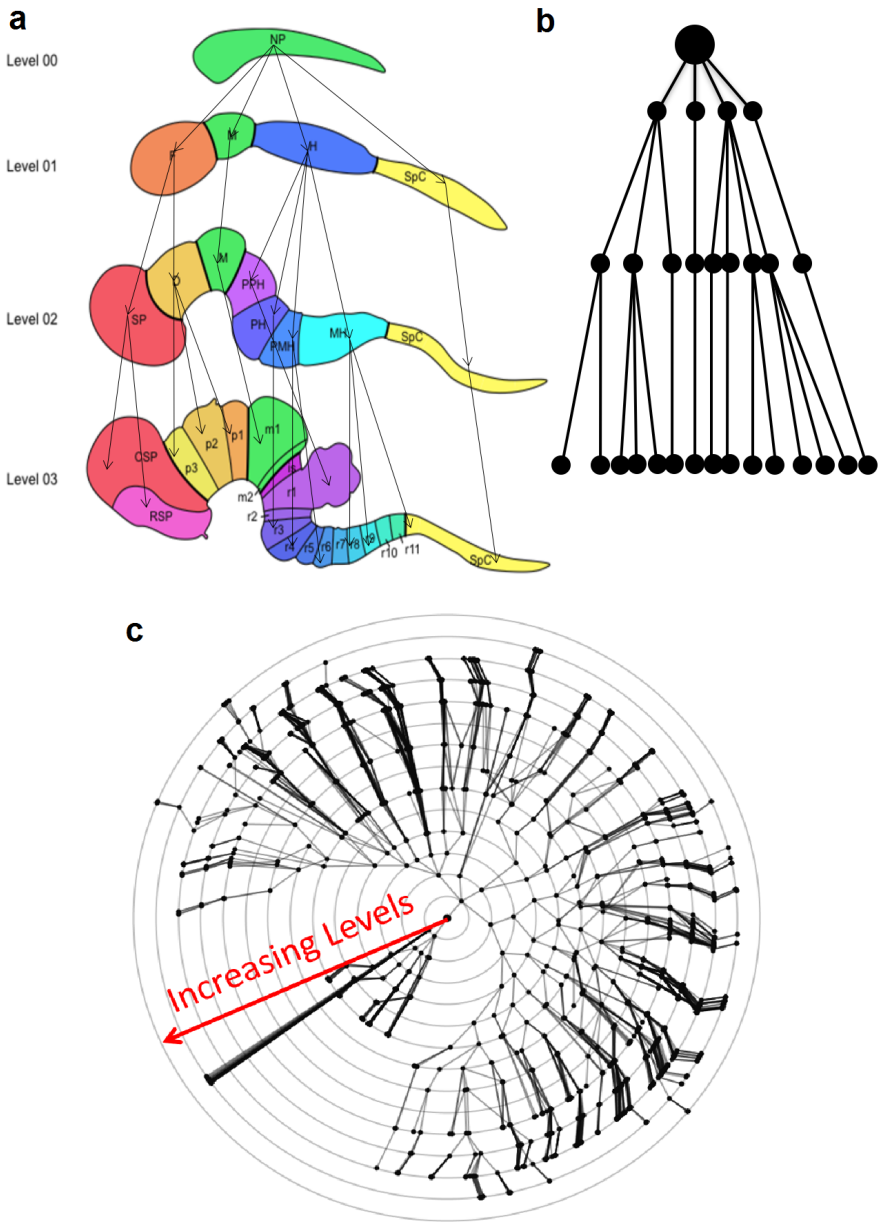

Figure 3.2: **Constructing the HOS-Tree.** HOS-Tree used a tree-layout graph to represent the structural hierarchy that using the first developing level as the root of the tree. (a) Example annotated structures of first four levels. (b) Constructed structure hierarchical tree derived from (a). (c) Aligning the structures nodes along annular circles in the preferential growth order, the complete HOS-Tree visualizes the entire brain structures.

3.3.1 Using Tree-layouts to Present Structures

The HOS-Tree imposes a hierarchical structure over the developmental stages and is well suited to observe the evolution of structural patterns. Since brain structures were selected for the spatial pattern visual analytics, it is natural to use their developmental orientations and hierarchical levels to represent such patterns. As done in the previous work, we used a tree-layout graph to represent the structural hierarchy. For a mammalian brain study, it is natural to use the first developing level as the root of the tree. Further, aligning the structures along annular circles presented the preferential growth of the structures in a hierarchy. Moreover, various brain structures form the nodes, while the edges among them displayed the developmental relationships (connected predecessor with successor structures). Figure 3.2(a) and (b) shows an example with various annotated structures at the first four levels. Figure 3.2(c) is the constructed hierarchical structure tree derived from this annotation. In the HOS-Tree, solid lines were used between structural nodes attributed with expression levels while dashed lines marked those nodes with missing measurements of expression levels. Additionally, the thickness of edges provided an indication of the normalized expression values at any given stage. This depiction allowed for a synergistic depiction of the structure of function, and vice-versa, a significant relationship in developmental biology.

3.3.2 Coloring the HOS-Tree Using Developing Orientation

Additionally, in order to visualize the spatial patterns intuitively, we next color-encoded this tree-layout graph by the developmental orientation. The developmental orientation of each structure could be obtained through the spatial locations of itself and its predecessor. However, Simply calculating the 3D vector from the centroids of the predecessor to

successor structures in different coordinates could lead to inaccurate results. Even worse, 3-D imaging registration could neither be used in our case due to the significant variances among the volumes' sizes and shapes. To solve this problem, we designed an ontological structure registration that assumed a structure's location at a higher hierarchical level or a later stage. In this method, we used the union of successor structures at the next level to predict the location of the predecessor structure. Equation 3.1 and 3.2 showed how we calculated the predicted volume (PV) of structure:

$$PV_{S_i} =< V_{S_i} \bigcup (\bigcup_{j}^{S_j \succ S_i'} V_{S_j}) > \tag{3.1}$$

Here, S_i denotes the i^{th} brain structure, V_{S_i} denotes the 3D volume of structure S_i in the atlas as well as $< V_{S_i} >$ denotes its spatial centroid , S_i' denotes the predecessor of S_i, and $S_j \succ S_i$ denotes that S_j is a successor structure of S_i. Thus, the spatial developmental orientation (DO) of S_i is defined as:

$$\vec{DO}_{S_i} =< V_{S_i} > - < PV_{S_i} > \tag{3.2}$$

where V_{S_i} denoted the spatial centroid of the brain structure S_i:

$$< \vec{V}_{S_i} > = (C_x, C_y, C_z) | S_i \tag{3.3}$$

and the coordinates of the centroid is the average of every voxel's coordinates in this brain structure volume at each unit direction (x, y, and z):

$$C_x = \frac{\sum_i \bar{x} V_i}{\sum_i V_i}, C_y = \frac{\sum_i \bar{y} V_i}{\sum_i V_i}, C_z = \frac{\sum_i \bar{z} V_i}{\sum_i V_i}, \tag{3.4}$$

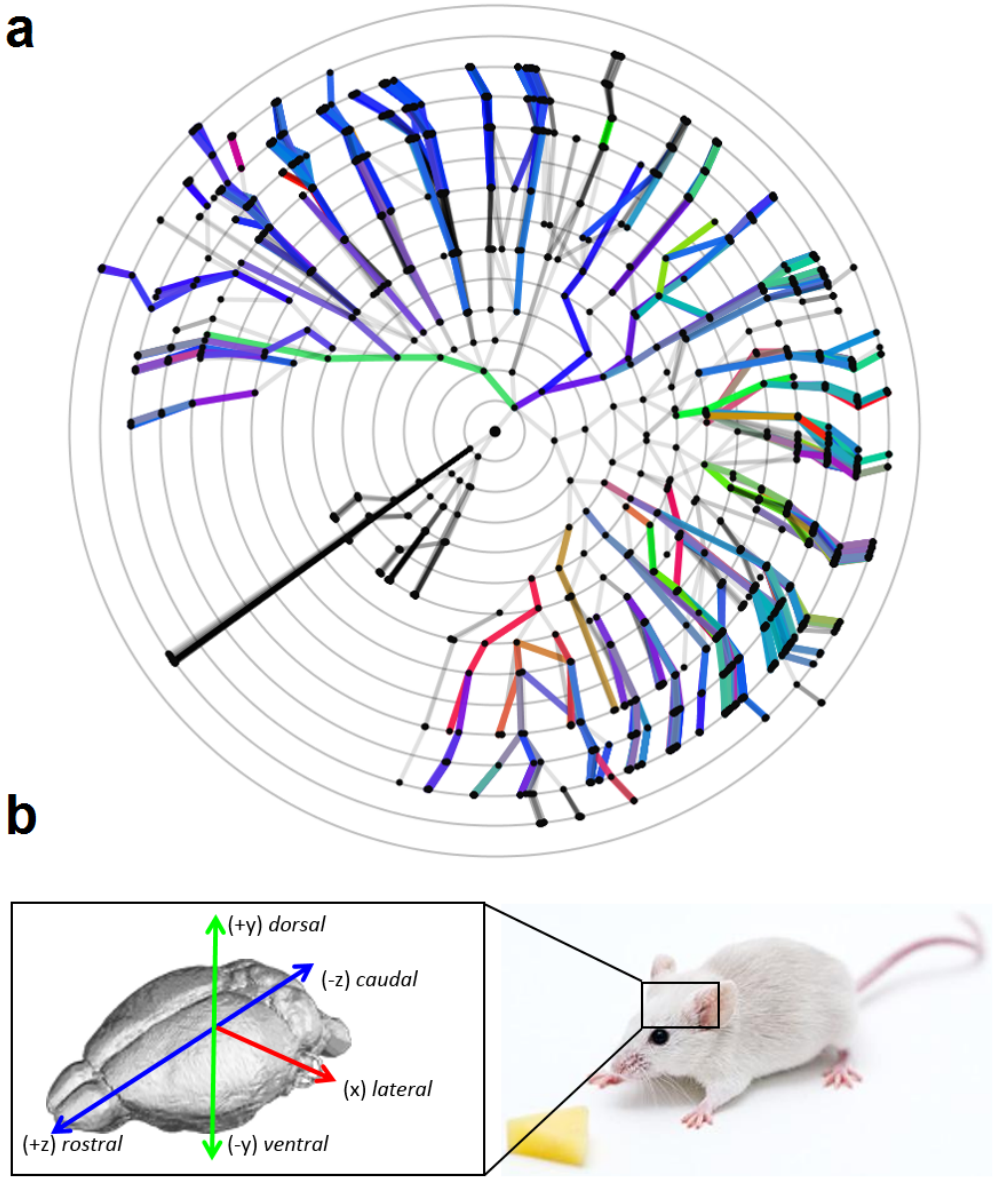

Figure 3.3: **Coloring the HOS-Tree using spatial developing orientations.** (a) The entire HOS-Tree used the ontological structure registration. (b) An example of how we converted the spatial developmental orientations into RGB space.

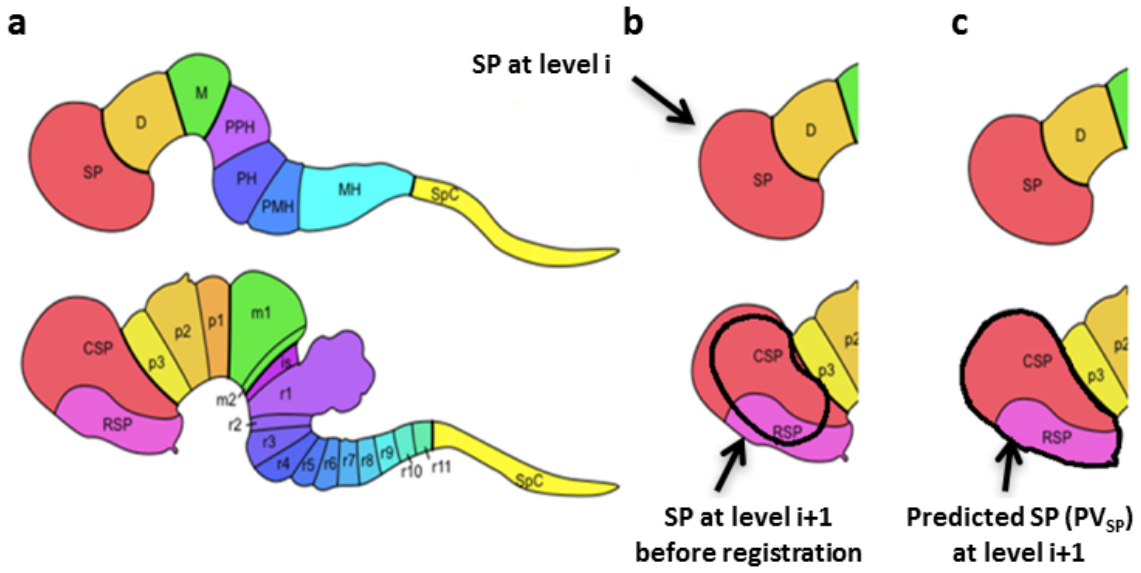

Figure 3.4: **Example of the ontological structure registration.** (a) Specific structures CSP (caudal secondary prosencephalon) and RSP (rostral secondary prosencephalon) were developed from SP (secondary prosencephalon) at the previous level; (b) Imprecise location of SP, as the black outline shows, without the structure-based registration; (c) Our structure-based registration approach provides the new location of SP, as the black outline shows.

Next, the spatial developmental orientations were normalized into RGB space, as the same method used in DTI (Diffusion Tensor Imaging) visualization. In this color system, the intensity of each primary color denotes the measurement on each color axis as determined by the projection of the development orientation vector: red is along the x-axis (medial-lateral), green is along the y-axis (ventral-dorsal), and blue is along the z-axis (caudal-rostral). Figure 3.3(a) shows the entire color-encoded HOS-Tree while (b) indicates the color-encoding method. Figure 3.4 shows the entire HOS-Tree using the ontological structure registration, in which we believed to provide better robustness on DO calculations: (a) structure CSP and RSP were developed from SP at previous level; (b) the location of SP is not precise when calculating the developing orientation without the structure registration; (c) the structure registration uses the union of children structures of SP, CSP and RSP, to predict the new location of SP for developing orientation calculation. In summary, the HOS-Tree to represent the spatial patterns of the entire brain structures, and it formed the first component of our visual analytics.

3.3.3 Interactive Functions in HOS-Tree

Our overall approach offered an efficient visual analytic system that allows for various exploration by integrating both visual analytics components (the BGEFM and the HOS-Tree). The interactive functions will be discussed later since they include the next visual analytics component. Among these functions, there is one that was designed especially for the HOS-Tree: gene profile exploration (GPE). The GSP function focuses on the spatial expression patterns of any given gene at various stages. After the user selects a single GOI (gene-of-interest) by using GPE, HOS-Tree can present its expression profile by highlighting the edges among the structural nodes, and hence displays the spatial expression patterns

at any given stage. In this case, since the colors of edges indicate the developing orientation while the highlighted edges measure the gene expression levels, the spatial expression patterns could be inspected from the HOS-Tree. Moreover, users can compare various directional expression patterns at each stage to find the expression changes throughout the development process.

3.4 Observations and Results

In this section, we will demonstrate the robustness and effectiveness of our prototypical visual analytic system by analyzing several explored spatial patterns. Figure 3.5 shows a distribution histogram of the DO on each direction. From the chart, DO that strong in z direction has much large number when comparing with in x and y directions, and this can be easily interpreted by the major developmental direction of the mouse brain is rostral-caudal (along z axis). By investigating these trends, we will find that the HOS-Tree provides the solutions for the contest tasks. Moreover, although the tasks were targeting on the spatial pattern of structures, the HOS-Tree overcomes the limited demand and provides an effective visual analytics for the spatial pattern exploration of both structure and gene expression. Especially, as described in the section above, GPE represents pertinent gene expression energy by the width of graph edges. Thus, the spatial patterns of both developing structures and gene expression are observed by comparing the variance of spatial expression patterns through the various developmental stages.

3.4.1 Exploring Global Spatial Pattern of Structures using HOS-Tree

In our first case study, we use our approach to explore the spatial pattern of developing structures. Since the HOS-Tree presents the entire structures, the spatial developmental pattern could be observed. Figure 3.6 showed an example analysis of spatial pattern by

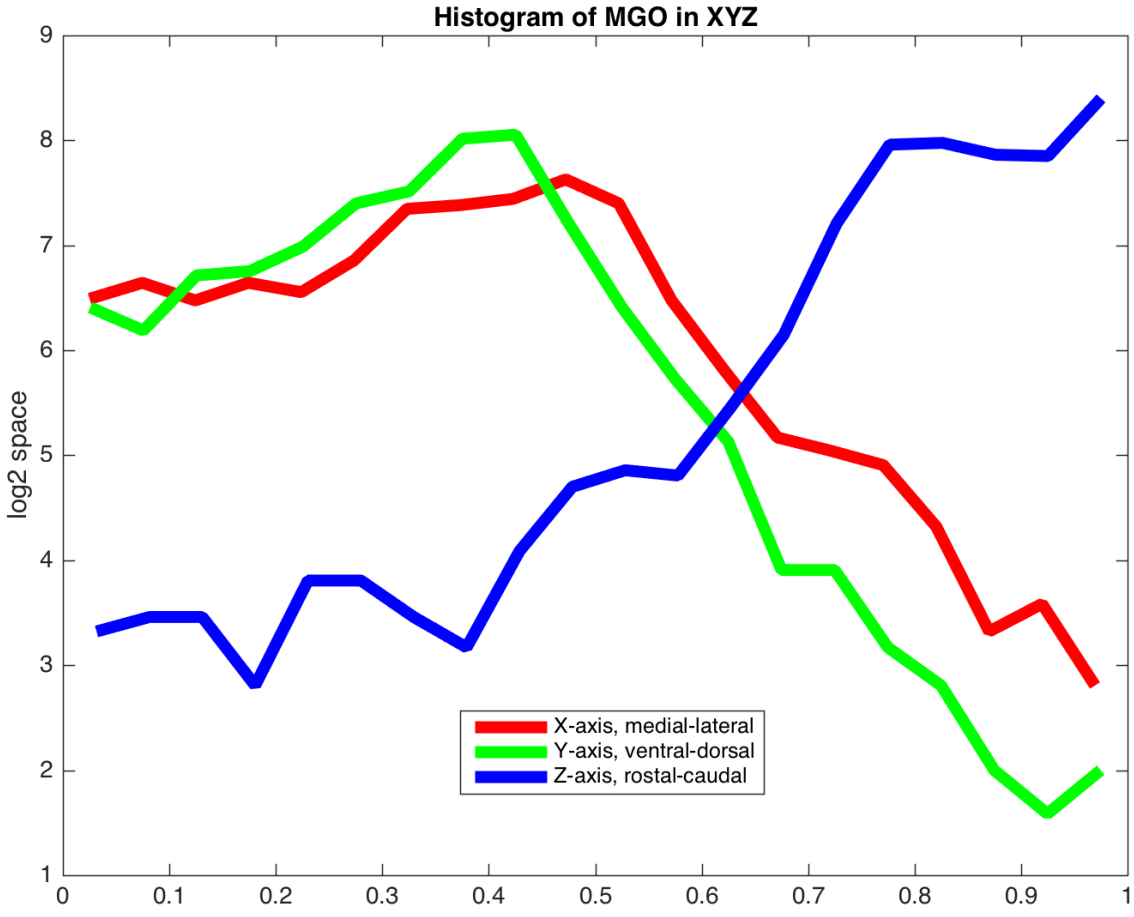

Figure 3.5: **Histogram of the distribution of the DO on each direction.** The DO that strong in z direction has much large number when comparing with in x and y directions.

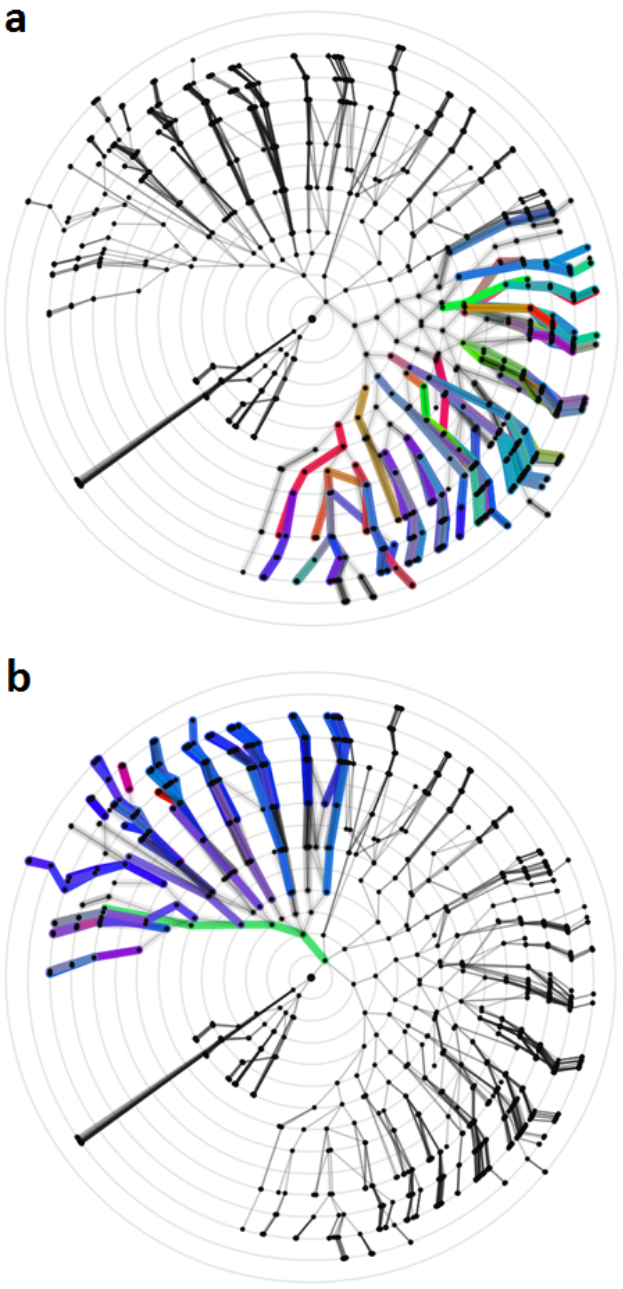

Figure 3.6: **Comparing the highlighted structures in the forebrain and hindbrain in HOS-Tree.** (a) The forebrain structures showed an expanding developing property. (b) The hindbrain structures had a strong developing direction in the z-direction.

comparing the structures in the forebrain and hindbrain. Since the edges were colored using the 3-D developing orientation, three color units (refer to Red, Green, and Blue) showed the orientation of the developing vector in each coordinate. Thus, based on the HOS-Tree visualization, we found the forebrain structures (highlighted in Figure 3.6(a)) showed an expanding developing property, while the hindbrain structures (highlighted in Figure 3.6(b)) had a strong developing direction in the z-direction ($caudal - rostral$). This observation could be interpreted by the biological ontology that the forebrain, majorly contains the frontal cortex, has strong expanding developing direction horizontally. On the other hand, the hindbrain, which contains the cerebellum and pons, majorly develop in an extending way. Thus, the global spatial pattern of developing structures in mouse brain is explored, and hence, the precondition of *Task 1: Structural Patterns* is solved. Therefore, we believed the HOS-Tree component was suitable and robust for spatial pattern exploration of ADMBA data, and next, we move on to the following step: explore the spatial pattern of gene expression.

3.4.2 Exploring Gene Profiles at Various Stages

Next, we demonstrate how the HOS-Tree component provides the exploration of the spatial pattern of gene expression. We first consider gene *Sim*2, well known as the homolog of one of Drosophila single-minded (*Sim*) gene. It expresses preferentially in the diencephalon during early embryogenesis to mediate neuroendocrine hormone gene expression. Since function GPE seeks the desired spatial pattern of a single gene across stages, here we use this function to investigate its spatial patterns at various stages. Figure 3.7 shows the spatial expression patterns, (a) at stage E13.5 (left), and (b) at stage P56 (right). As a reminder, HOS-Tree uses colors to measure developing orientations, widths

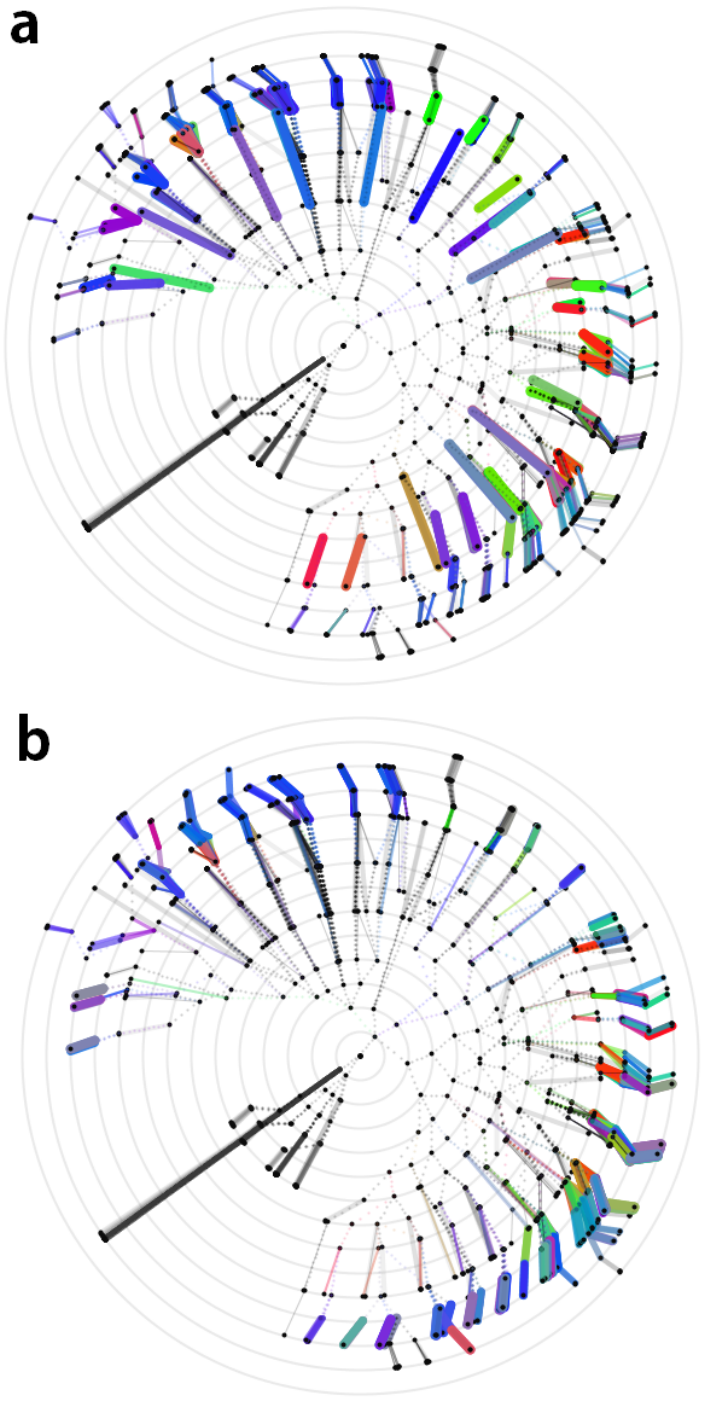

Figure 3.7: **Gene profile analysis of gene** *Sim*2**.** (a) For the pattern at stage E13.5, the strongly expressed areas are mostly mid-range-between levels 9 and 11. (b) The pattern with high expression at stage P56 diffuses into late-range levels-between levels 11 and 13.

to measure the gene expression energy, and corresponding radius to measure hierarchy levels. Thus, the nodes in the HOS-Tree with wider edges connected are the structures *Sim*2 strongly expressed. Thus, it is clearly observable that the highly expressed areas are in mid-range levels (between levels 9 and 11). Compared to the pattern at stage P56, the highly expressed areas are diffusing into late-range levels (between levels 11 and 13).

We also choose gene $Tbr1$ (T-box, brain, 1) to perform another investigation. This analysis focuses on the comparison between distinct spatial expression patterns at postnatal stages P14 and P56. Gene $Tbr1$ serves as an important transcription factor in vertebrate embryonic development, and it is critical for neuron differentiation and migration in brain development. Furthermore, $Tbr1$ is considered to be one of the essential genes in regulating development of the human cortex. Figure 3.8 shows the spatial pattern of the expression profile of gene $Tbr1$. It can be clearly observed that many of the wider edges are colored as either green or red. By the definition of the HOS-Tree coloring scheme, this pattern provides strong evidence that the observed expression energy is much stronger along the x- and y- axes than the z-axis. Since x and y units estimate horizontal and vertical spatial gradients respectively, the observed spatial expression pattern matches often observed major cortical development closely. On the other hand, when compared to the spatial expression pattern at stage P56, the expression energy observed (per the widths of edges) are much weaker. Based on this observation, we believe gene $Tbr1$ is functionally expressed before the postnatal stage P56.

We now state our solution to *Task 1* and *Task 2:* by analyzing gene profiles at various stages using the GPE (gene profile exploration) function, the spatial pattern of expression of any given gene could be detected by examining the width of edges in the HOS-Tree.

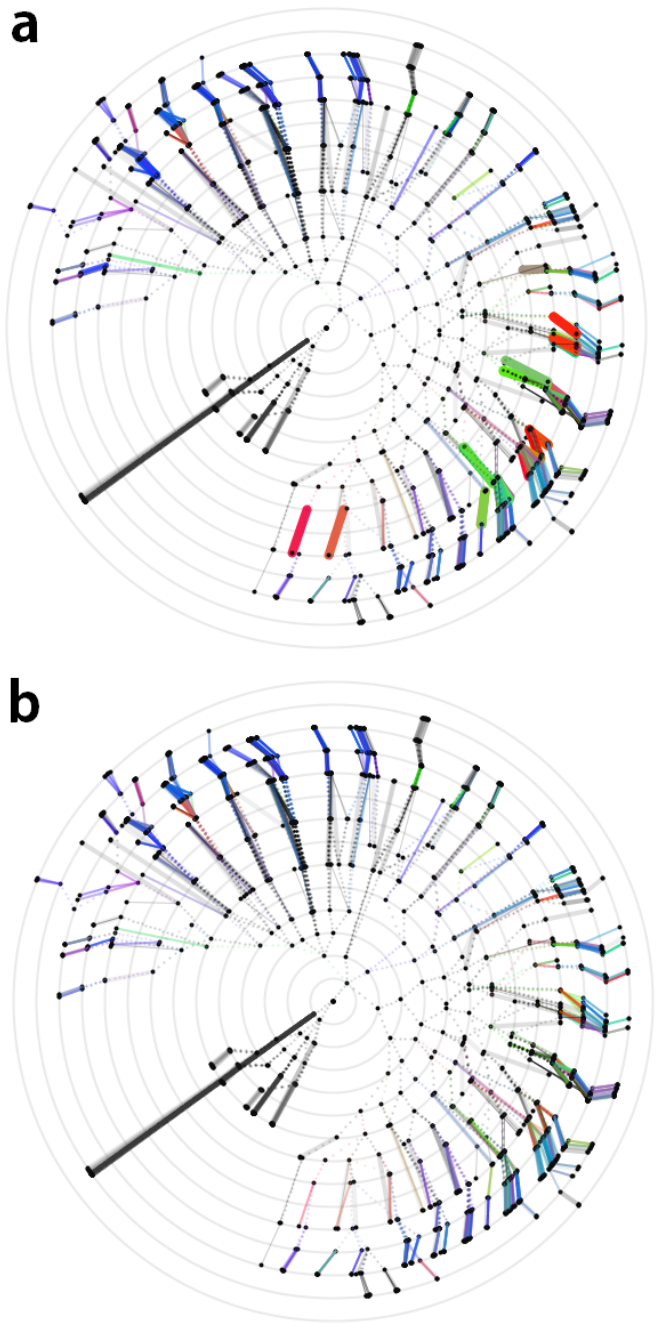

Figure 3.8: **Spatial pattern of the expression profile of gene** $Tbr1$**.** (a) The pattern at stage P14, wider edges are mostly colored as green and red indicating that gradients exist along x-y-axes. (b) The pattern at stage P56 depicting expression energy is much weaker indicating that $Tbr1$ is functionally expressed before stage P56.

Hence, we firmly believe that the HOS-Tree visual analytic system is robust and efficient in observing the spatial patterns of both developing structures and gene expression.

3.5 Summary

In this chapter, we present HOS-Tree, an interactive visual analytics method to explore the spatial patterns of structures in the ADMBA data. HOS-Tree enables the spatial pattern exploration not only for developing orientations of all brain structures but also for gene expression among them. The HOS-Tree generates a tree-layout visualization for the brain structures, and use the spatial development orientation to color-encode the tree. This data-driven visual analytics overcomes the first challenge in Section 1.1: how to enable the learning of the natural patterns of structures in the developing brain.

We used two case studies to investigate the performance of the HOS-Tree. The first case study focused on the global spatial developing pattern of all structures, while the second one picked two example genes to analyze the spatial expression pattern at various stages. Both of them indicate the HOS-Tree provides efficient solutions for the parts of 2013 IEEE Data Visualization Contest tasks and, more importantly, is suitable for delivering efficient and intuitive spatial pattern exploration approaches to the users.

However, the HOS-Tree still has its limitations. The major problem is the overlay of the structure nodes on the interface due to the large amount of brain structures. It is a limitation coming from the iteration when calculating the position of the nodes on the tree. Since each node was placed on the certain circle to indicate the developing hierarchy level, the overlay problem in the HOS-Tree cannot be Simply solved until another force-based algorithm was used. Thus, in the future, we plan to look for a robust force-directed method to revise the node overlay problem in the HOS-Tree.

Chapter 4: Exploration of the Temporal Pattern of Brain Development

In this chapter, we present another interactive visual analytics method to explore the temporal patterns in the AIBS datasets. When we participated the 2013 IEEE Data Visualization Contest, we developed another visual analytics component that focused on visualizing the gene-structure associations in a temporal pattern manner, especially, we named it the BGEFM which short for Bi-clustered Gene Expression Flow Matrix. This data-driven visual analytics overcomes the second challenge in Section 1.1: how to enable the learning of the inherent temporal patterns of gene expression in the developing brain.

The BGEFM serves as an integrative visual analytics component, and satisfies the other core prerequisites for spatiotemporal pattern exploration. In order to implement the BGEFM, we first used the data-mining approach to learning the inherent temporal expression profiles of genes and consider them as the temporal pattern. We next use these learned patterns to represent the association between genes and spatial structures and used a 2-D matrix to visualize them. Also, we apply bi-cluster process to this matrix to seek the correlated region-specific expression behaviors. Again, we investigate the usefulness of BGEFM through the biological analyses of several patterns explored.

In this chapter, we first describe the background and motivations in Section 4.1. In the following Section 4.2, we describe the data collection for the temporal pattern exploration component. Since this chapter includes both ADMBA and ADHBA data, we first introduce

the methods used for BGEFM in Section 4.3, including data processing and data-mining approaches. Then, we focus on the visual analytics of ADMBA in Section 4.4 and of ADHBA in Section 4.5. In each section, we will discuss the visualization processes and results, and next use several observed patterns to investigate the efficiency of the BGEFM. Finally, we summarize our work in Section 4.6.

4.1 Background and Movitation

Exploring inherent genomic characteristics of the developing mammalian brain is critical for developmental neuroscience and an important topic in translational bioinformatics. Over the past few decades, significant advances have provided substantial insights in this and related fields [8, 43]. Spanning the studies at multiple levels from structural, to cellular, and to molecular levels, numerous studies have characterized functions and regulations of genes in the brain developmental processes [9]. While this body of work has provided an increasingly clear picture of genomic characteristics in brain development, however, a comprehensive and deeper understanding of genes requires much further study due to the complexity of the developing brain.

In developmental neuroscience, distinct spatial and temporal patterns of gene expression are essential for revealing the genomic characteristics of the brain development [20, 129, 22]. In genetics, expression is the process by which the genetic code of a gene is activated for protein synthesis, hence the gene expression, which reflect the activation status of genes, result in the phenotypes such as a variety of shapes and types of cells and their functionalities as well as their organizations into various anatomical structures in the brain [24]. Various types of patterns of gene expression thereby provide valuable information for delineating the cell lineage and organization patterns as well as the functionalities

of structures during brain development. Numerous neuroscience studies have focused on two major forms of genetic patterns: temporal patterns and structural patterns. The first one represents the locally consistent temporal patterns of gene expression in certain brain region(s) or structure(s). The second one, on the other hand, reveals the differences of expression patterns across different brain locations. However, since the brain is composed of different functional regions and various cell types, both structural anatomy and functionality change uninterruptedly throughout the entire period of growth, and more importantly, these changes interact and influence each other. Thus, an integrative pattern that exhibits the regional genomic behaviors through time can provide a more comprehensive observation of the genomic characteristics in the developing brain. Consequently, detecting, exploring, and understanding these patterns are of great interest and importance to the research field and a comprehensive visual analytic strategy combining data analysis and visualization is needed to achieve this goal.

Visual analytics plays an important role in bioinformatics and neuro-informatics due to its capacity for organizing, transforming, analyzing, and representing a large amount of data in a visually intuitive fashion [46, 48, 130]. In 2013, the IEEE Scientific Visualization Contest targeted the exploration of the comprehensive patterns of gene expression in the developing mouse brain [30]. In order to tackle the posed challenges and tasks, we designed and developed an integrative visual analytics system to finish several posed challenges [4]. As one of the core visualization components in that system, a bi-clustered gene expression flow matrix (BGEFM), was implemented to allow the observation and identification of the region-specific temporal patterns of gene expression during brain development. The BGEFM enabled users to explore distinct temporal genomic characteristics across all brain structures during all developmental stages and thereby gain a comprehensive understanding

of the developmental processes. Using the BGEFM, we discovered multiple types of salient patterns including: gradient pattern which exhibits genes with directional expression; structural pattern which identified the structures with highly expressed genes; consistency pattern which highlight the structures with the most consistent expression patterns over time; and complementary pattern which include genes exhibiting complementary patterns inside a structure.

As described, the AIBS data portal has published the developing brain data for both mouse and human. The topic of the 2013 IEEE Scientific Visualization Contest targeted the existing domain of developmental neuroscience about the mouse brain, and we developed the HOS-Tree to finished several challenges. Here, we propose another visual analytics component to aid neuroscientists explore the hidden relationships between gene expression and structures, and more importantly, provide an interactive and robust visualization approach. We created the BGEFM, the second visual analytics component in this dissertation, which enables the observation and identification of the region-specific temporal patterns of gene expression during brain development. It should be noticed, the BGEFM was designed in response to the temporal pattern exploration challenges posed by the contest.

Although the brains of all mammals develop in similar ways, understanding the developmental processes of both mouse and human brain are important [10]. After the ADHBA data published, we shifted our attention to the human brain. ADHBA provides gene-per-region expression data for almost all known gene probes across brain regions, and hence it is best suited for our approaches to perform the exploration of regional temporal patterns. Thus, we also perform the proposed platform to enable an integrative exploration of the regional temporal patterns of gene expression in the developing human brain.

In summary, our proposed framework provides a comprehensive exploration of patterns as well as capabilities for interactive visualization for both ADMBA and ADHBA data. It contains two components: 1) a data-driven pattern learning approach that identifies inherent temporal patterns of gene expression across various brain regions, and 2) the BGEFM for region-based temporal pattern visualization. The entire workflow and analysis of our proposed framework can be described in the following steps:

Step 1. Generate a data structure to represent the gene-per-region expression data at various developing stages.

Step 2. Clean the data in order to retain the most information and distinct expression entries.

Step 3. Use $K-Means$ clustering to learn the regional temporal patterns of gene expression.

Step 4. Visualize temporal patterns learnt from the clusters.

Step 5. Interpret the patterns using neurodevelopmental processes and events [6].

Step 6. Apply BGEFM to present the region-specific characteristics.

Step 7. Investigate the correlated gene-structure sections in the BGEFM using biological enrichment analyses.

Using this framework, we identify several interesting patterns that were explored (step 7), and further examine them using gene enrichment analyses with a focus on neurodevelopmental processes. Based on the results, we have several biologically interpretable observations that can potentially enhance the biological knowledge. These learnt temporal patterns, which interpreted by the neurodevelopment events, reflected the brain developmental

processes. Also, the genes with these patterns clearly play certain roles in various regions during the development, and the temporal patterns captured these roles. Specifically, genes can be categorized into two subsets — 1) region-agnostic, which showed consistent temporal characteristics in the entire brain, and 2) region-specific, which showed distinct patterns at various brain regions. In addition, the brain regions that physically close have similar temporal patterns for most of the genes. We postulate that the explored patterns reflect the comprehensive genomic characteristics of the brain developmental processes. Hence, neuroscientists will be able to gain a deeper understanding of the development of the brain using the proposed framework. It should be noted that although our proposed framework focused on the region-specific temporal pattern exploration for both the ADMBA and ADHBA data, it could serve as an efficient and robust visual analytic solution for other multi-dimensional data that contains large numbers of entries with spatial measurements across time. Next, we briefly describe the posed challenges and tasks one more time and the data collected for this chapter.

4.2 Data and Tasks

As described, the AIBS data portal has published the developing brain data for both mouse and human. In this section, we briefly describe the collected data, the ADMBA (Allen Developing Mouse Brain Atlas) and ADHBA (Allen Developing Human Brain Atlas), as well as the other tasks and challenges posed by the 2013 IEEE Data Visualization Contest [4].

4.2.1 Data Collection

The ADMBA provides the expression levels of 2105 genes in 2691 anatomical structures across three embryonic and three early postnatal stages: E13.5-embryonic days 13.5,

E15.5-embryonic days 15.5, E18.5-embryonic days 18.5, P4-postnatal days 4, P14-postnatal days 14, and P56-postnatal days 56 (adult). Three-dimensional reference mouse brain atlases for each stage and the corresponding annotated volume with structure labels are also available. Additionally, a hierarchical annotation of various evolving brain structures is available. Finally, the scrutinized genes are organized into 11 categories allowing for further appropriate enrichment studies.

Similar to the ADMBA project, Allen Developing Human Brain Atlas (ADHBA, or BrainSpan) project was introduced for describing transcriptional mechanisms involved in the developing human brain. However, differ to the ADMBA data which only contains neuroanatomical genes, the ADHBA data examined 52376 probes which contain almost all known genes to provide a complete data portal for the field scientists. First, In the ADHBA data, two data modalities were used for the gene examination which includes RNA sequencing and exon microarray hybridization. Also, the expression data were generated based on 42 brain samples, including both males and females, spanning prenatal and postnatal stages of brain development. For each brain specimen, up to 26 targeted cortical and subcortical structures were examined by most probes of genes.

4.2.2 Tasks

Besides on the tasks described in Section 3.2.2, the 2013 IEEE Scientific Visualization Contest posted others tasks and challenges for the region-specific temporal pattern exploration.

Task 1. Structural Consistency. Which structures have the most consistent expression patterns over time? Which structures are the least consistent?

Task 2. **C**omplementary Patterns. Which genes have expression patterns that complement each other within a structure? Are these patterns persistent during development?

Although these data have become an invaluable resource for the exploration of interesting temporal patterns of gene expression, a viable systematic visual analytic platform has still failed to materialize. Thus, to help field scientists to access, observe, explore, and analyze the interesting spatiotemporal patterns, we propose a visual analytics component base on ADMBA and ADHBA data. Using the observed patterns that provided by this component, users can understand the genomic characteristics for particular processes in the developing brain, which can potentially lead to further analysis.

4.3 Visual Analytics Design

In order to explore the inherent and temporal patterns of gene expression from the complex data of ADMBA, it is necessary to use an efficient and suitable approach to representing the data before doing the analysis. In this section, we will introduce the design of the proposed visual analytics component, the BGEFM. First, we will describe the data processing steps followed by the temporal pattern-learning process. Then, we will introduce the how we integrate the temporal pattern with BGEFM to provide a global visualization.

4.3.1 Data Processing

We first used a 3-D matrix, $M_{g,s,t}$, to collect the pair-wise expression values for specific genes in the 1^{st} dimension (g in rows) across particular spatial structures in the 2^{nd} dimension (s in columns) at definitive stage in the 3^{rd} dimension (t in depth). Thus, the matrix M

contains the entire collected data, and any data entry could be queried through any given g, s, and t.

In multidimensional data analysis and visualization, one of the most challenging problems is the dimensionality reduction. An efficient reduction process will not only reduce the complexity of the data structure but also enhance the performance of the visual analytics. Focusing on the AIBS data, we first used a data-mining approach to reducing the dimensionality on the time-dimension to learn the time-varying characteristics of gene expression, so-called temporal patterns. Next, we used a 2-D matrix visualization to represent the associations among genes and structures in the temporal patterns manner.

In the data-mining field, clustering process best suited our purpose since by grouping subsets of collected temporal expression profiles in such a way that the genes in the same group shared the highest similarity in time-varying behavior. Thus, we clustered the data matrix M into subsets of temporal profiles and hence to reveal the temporal patterns. In order to increase the cluster performance, we converted the data matrix M into the format of temporal profiles. In this approach, for each gene at each structure, we collected the gene expression values at various stages and reshaped them into a temporal profile in the format of a 1-by-6 vector for mouse, and 1-by-31 for human. For the stages that the structure did not exist, we used the median expression values of its ancestor structures at the corresponding stages to represent those missing data. Since the sampling structures in each profile are in the stage sequential order as well as the developmental hierarchy, the inherent temporal patterns would be preserved from the original data. Thus, this data conversion process generated a collection of complete temporal profiles while keeping the inherent characteristics of genes. Eventually, this process converted the 3D data matrix M into a temporal profile matrix, and we named it $MV_{g,s}$.

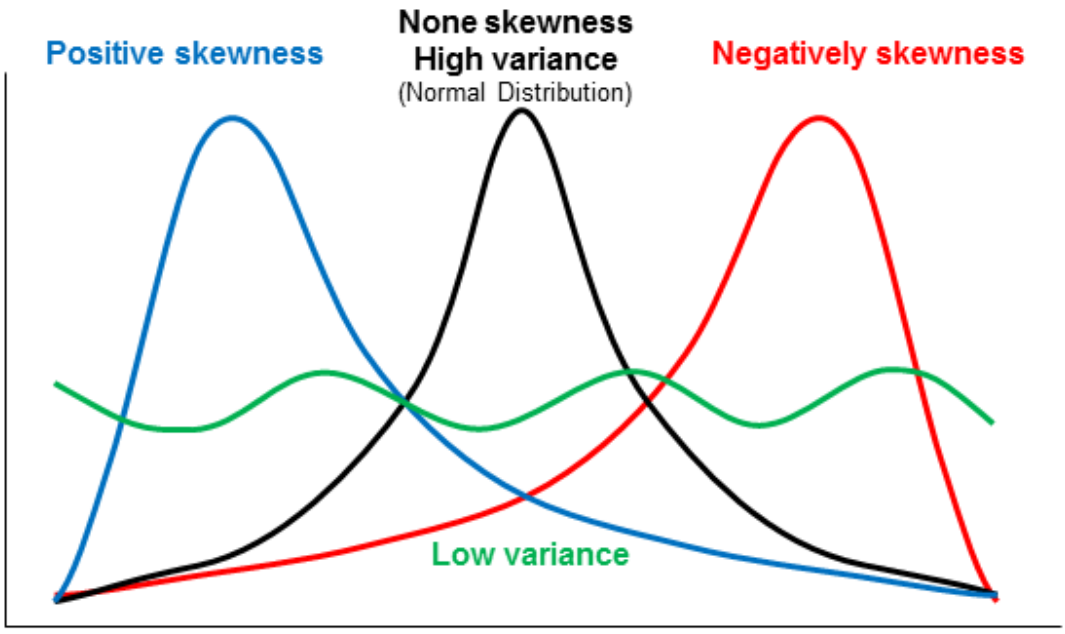

Figure 4.1: **Skewness distributions.**To present the association among genes in rows and structures in columns. Different distribution types of the temporal patterns provide valuable properties for future analyses: positive skewed distribution (blue curve), the negative skewed distribution (red curve), the near-flat distribution with low variance (green curve), and the normal distribution with high variance (black curve).

However, $MV_{g,s}$ cannot be used for further analyses directly due to its huge size. We next filtered it based on the features of the central moments in order to increase the efficiency and the performance of the pattern-learning process in the later step. Here we considered the expressional flows with both the largest variance and most significant skewness for the pattern-learning process. The variance is the 2^{nd} central moment of the temporal flow, which uses the squared standard deviation to measure how the distribution is spread out from the mean. As Figure 4.1 shows, comparing with the low variance distribution (green curve), the normal distribution (black curve) has more variance in statistics and contains distinct temporal property. On the other hand, the skewness, the 3^{rd} central moment, measures the lopsidedness or the asymmetry of the distribution of expressional values in the temporal flow. Different from the normal or the uniform distribution, both positive (blue curve) and negative skewness (red curve) show specific temporal properties, which are significantly valuable for detecting the expressional behaviors in the temporal trend. Thus, both variance and skewness provided the measurements to distinguish the time-varying changes of gene expression across structures and regions and hence enabled an efficient learning of inherent and distinct temporal patterns, and therefore, the observation of inherent temporal patterns is enabled. However, although this process grasped the most significant genomic properties in the developing brain, balancing between accuracy and performance, it is not guaranteed that the whole inherent patterns will be grasped. We believe a better performance supercomputer would make a more precise pattern exploration available in the future.

4.3.2 Learning Temporal Patterns

Based on the filtered $MV_{g,s}$, which contained the most significant gene-per-structure expression flows, we used a data-driven pattern-learning approach to reveal the temporal

patterns. This process has two steps: learning modules from data, and then obtaining patterns from those modules. The implementation of the learning module could be best considered as a clustering process that classifies the expression flows into groups. Thus, simply applying a cluster analysis to the filtered $MV_{g,s}$ could complete the first step. In order to discover the salient patterns, we repeated the clustering process for ensuring the convergence, and used silhouette to choose the best result from multiple parameters. It should be noted, although unsupervised cluster approaches are best suited for the non-known data, depending on the situation of the study, the supervised or semi-supervised clustering could also be considered.

Once the temporal module learning process is completed, the temporal patterns could be observed from the clustering results. We next used the median or mean values at each stage to represent the time-varying flows. Then we used those flows to represent the temporal patterns and hence the associations between the corresponding genes and regions. For each temporal patterns, the x-axis shows the temporal stamps while the y-axis denotes the collected expression values in the corresponding time stamp. Consequently, we sort the temporal patterns in the order of the appearance of the first peak and hence the patterns follow a decreasing-to-increasing order. Moreover, we assigned a pattern index as well as a unique color to each pattern in order to provide a measurable association [131]. Therefore, the pattern-indices-based representation not only reduced the dimensionality on the temporal series but also enhanced the efficiency for future analysis and visualization.

4.3.3 The Bi-clustered Gene Expression Flow Matrix (BGEFM)

Before exploring the region-specific patterns, we design the visualization component, BGEFM (Bi-clustered Gene Expression Flow Matrix) to represent temporal patterns. We believe a 2D matrix could best represent the association of the gene-structure data. Thus,

we used such a matrix to represent the associations between entries in rows and locations in columns. Figure 4.2 shows the BGEFM designed in our previous work. However, when we designed the original BGEFM in 2013, it had significant limitations. First, the temporal patterns of gene expression were predefined. Thus, the temporal associations among genes and structures are biased. For a robust approach, the temporal patterns should capture the natural characteristics that gleaned from the data. Hence, targeting the ADMBA and ADHBA, we use a data-driven approach to learn the inherent temporal patterns for the visual analytics. For each entry inside the BGEFM, we applied a pattern-matching process to seek the index of the most matching temporal pattern based on the Pearson Correlation Coefficient (PCC) and assigned to the entry. Obviously, the size of such a matrix should be the size of entries by the size of the measured locations, and each element in this matrix was the label of the temporal patterns. Therefore, this 2D matrix indicated the associations of the entries and locations in the form of temporal properties.

Bi-clustering, or co-clustering, is a data mining approach to seek the strong correlated sections among rows and columns of a matrix. Over the past decade, many bi-clustering algorithms have been developed and widely applied in bioinformatics in order to explore local patterns [132, 133]. By clustering data entries in rows (genes) and columns (regions) that share the most similar expressional patterns together, bi-clustering is best suited to our purpose in seeking the strongly correlated gene-structure sections. Thus, we next performed a bi-clustering analysis to this 2-D matrix and we named it the BGEFM which is short for Bi-clustered Gene Expression Flow Matrix. In summary, the BGEFM used visualization techniques to encode the variability of the expression patterns over various stages and among regions in the temporal pattern manner.

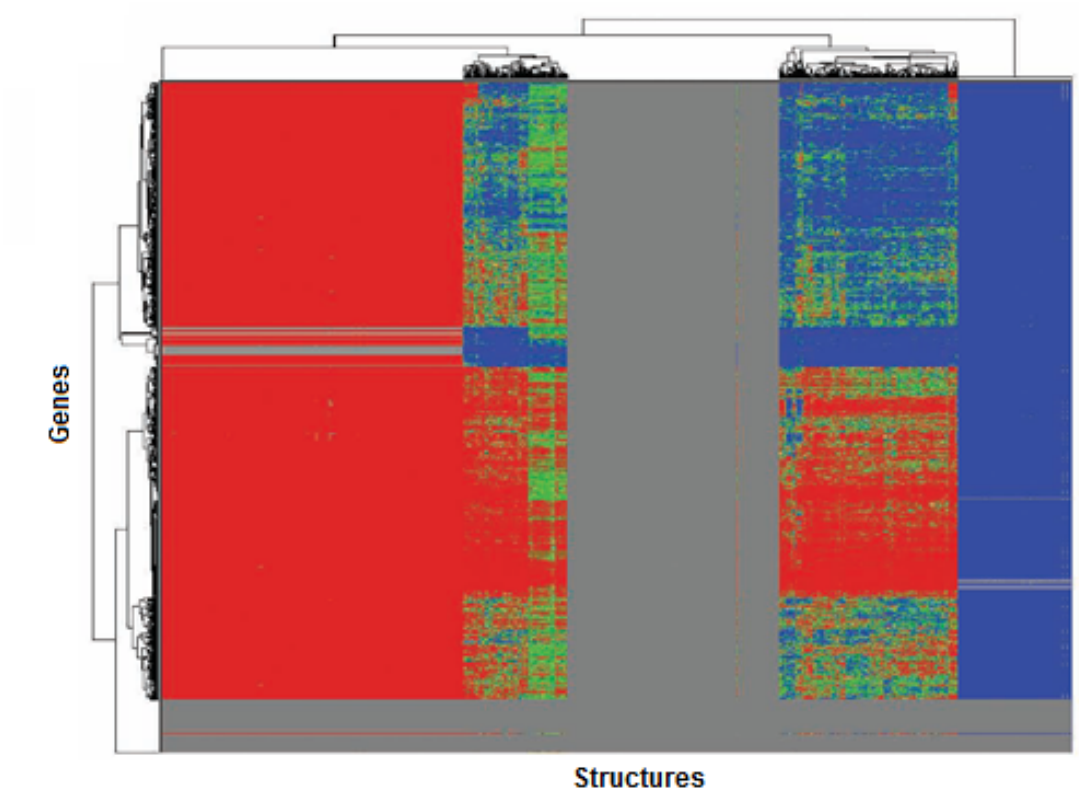

Figure 4.2: **The original BGEFM developed in 2013.** The BGEFM visualization approach designed in our previous work [4]. However, the pattern-matching process was used with 5 pre-defined temporal patterns, and lead to biased result.

4.4 Implementation and Result for the Developing Mouse Brain Dataset

We applied the temporal modules-learning method into ADMBA data to observe the temporal patterns of gene expression.

Based on the proposed method, the key process in this module was an unsupervised cluster analysis to the filtered 3D storage matrix that in the size of 1753*2489*6 (filtered from the original size as 2105*2691*6). As described, in order to maximize the performance and provide the most significant genes for bioinformatics analyses, we used the central moment filter to reduce the data size. Balancing the performance (4-Cores Intel Core i7 CPU and 16GB RAM), we chose the top approx. 2.2 million profiles (approx. 50%) of the temporal profiles that had largest variance and significant skewness for the next pattern-learning module (in log_2 space). However, although this process gleaned the most significant genomic properties in the developing brain, it is not guaranteed that all inherent patterns will be grasped.

4.4.1 Temporal Patterns and BGEFM

For the ADMBA data, we performed the clustering process (setting as described in Section 4.3.2) for each reasonable K ($3 \leq \mathrm{K} \leq 13$), and then, used the silhouette to find the best cluster result. Figure 4.3 showed the max, mean, median, and min values of silhouette for each K. In our case, balanced the performance and robustness, we chose K as 10 for our future analysis. Figure 4.4 showed the ten temporal patterns in the corresponding colors as well as the fitting curves in the encoded colors.

As proposed in our method, we next used BGEFM to visualize the associations between genes (in rows) and structure (in columns). As described, we applied the pattern-matching process to seek the most-similar temporal pattern among the ten learned ones. In our case,

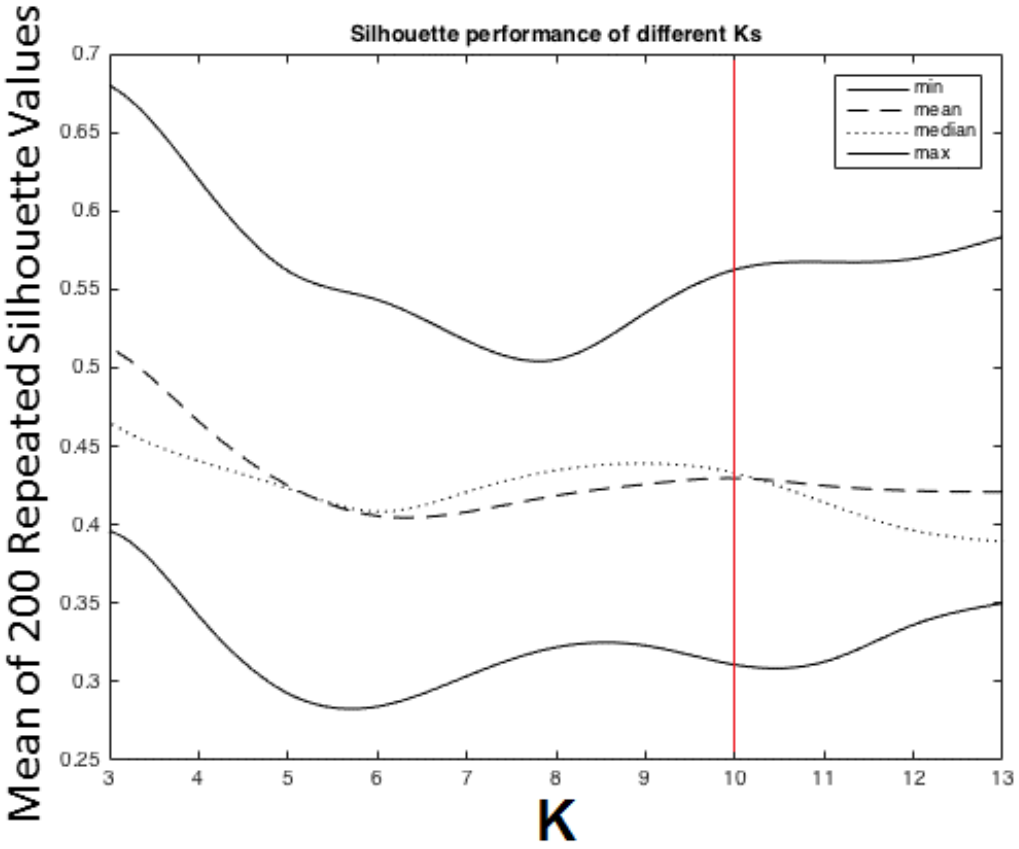

Figure 4.3: **Using silhouette properties to choose the best temporal pattern-learning result for the mouse brain.** For the temporal pattern-learning process in the ADMBA data, we repeat the K-Means cluster for each K from 3 to 13. Based on the result, we choose K as 10 for the following work (values are curved using a smoothing spline).

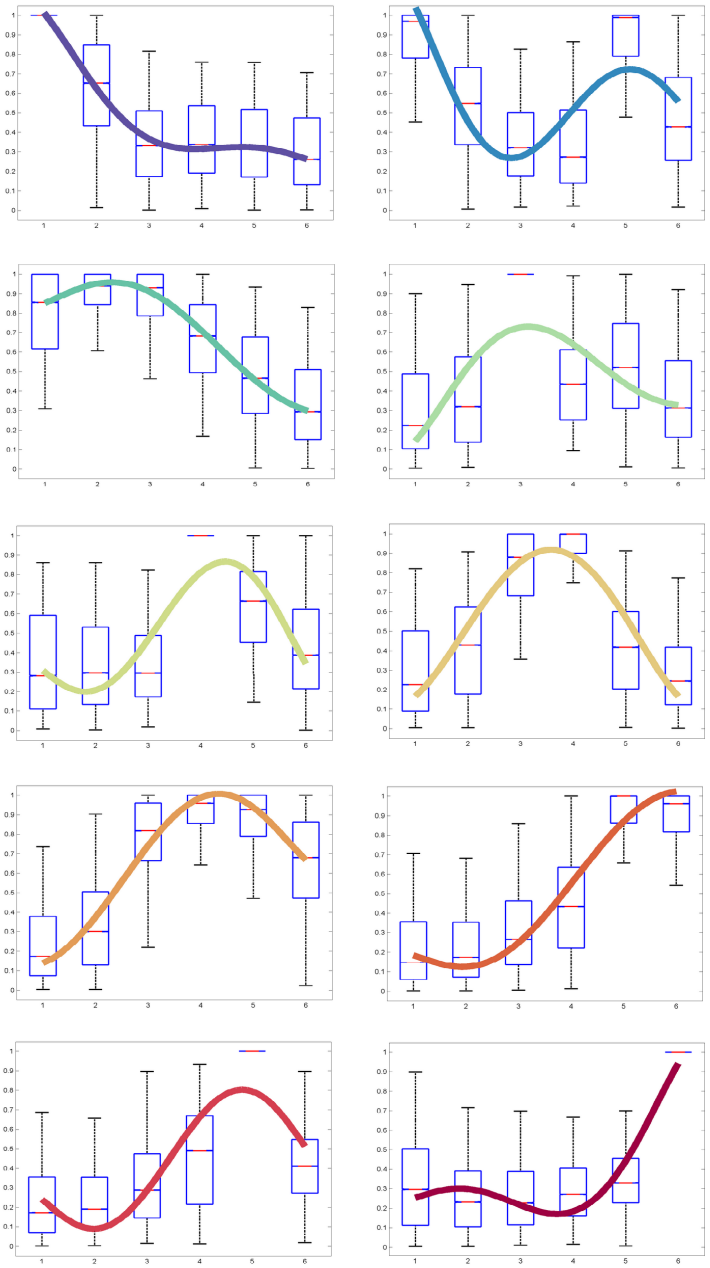

Figure 4.4: **The Color-encoded Temporal Patterns Learned from the ADMBA dataset.** The ten temporal patterns in the corresponding colors as well as the fitting curves in the encoded colors. The *x* anix indicates the stages from E13.5 to P56, the *y* axis is the normalized gene expression values.

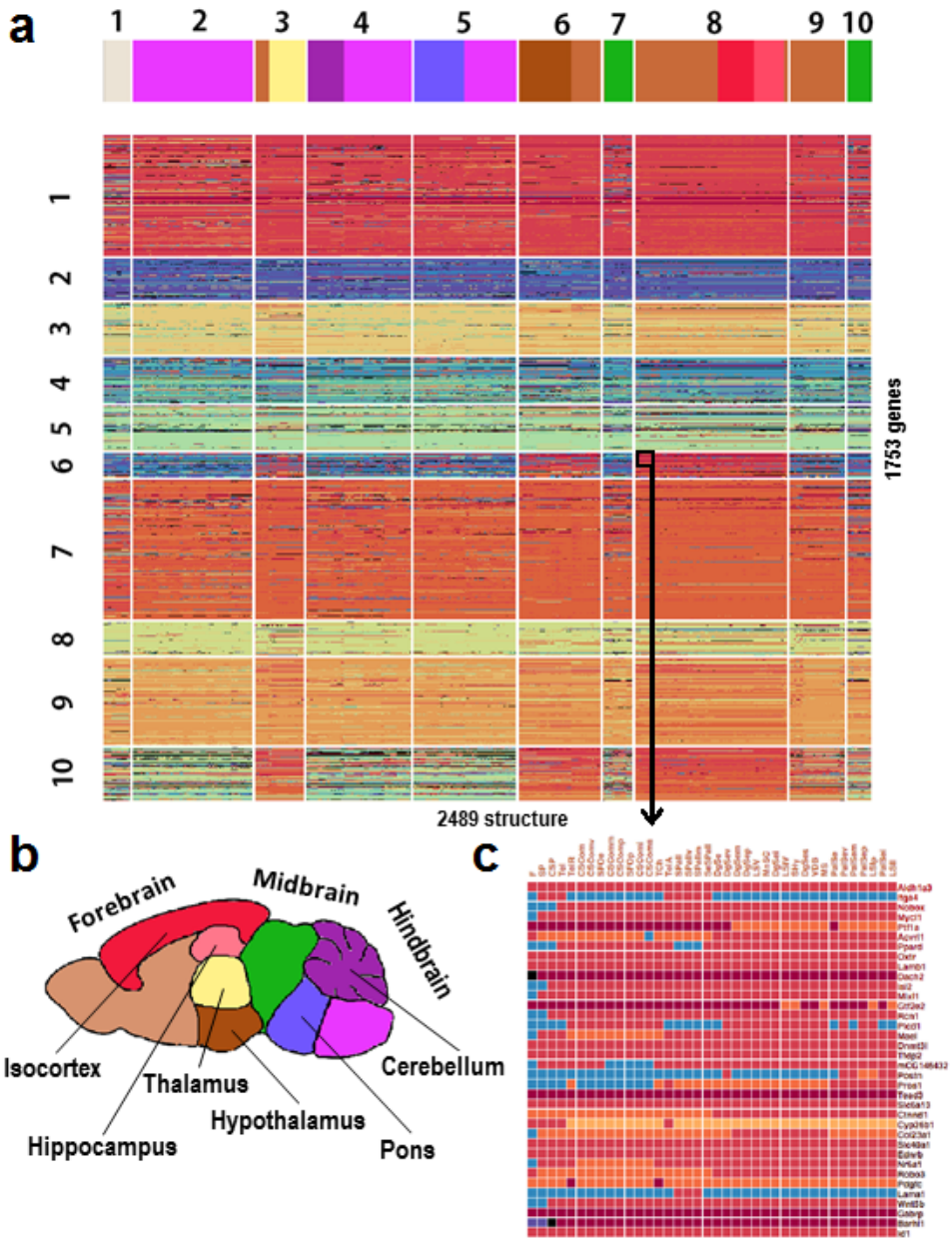

Figure 4.5: **The Color-encoded BGEFM for the ADMBA dataset.** (a) The color-encoded BGEFM, each entry inside indicates the color-labeled temporal association across genes (in rows) and structures (in columns). (b) The brain regions chosen for the visual analytics, which matched the structure bar on the top of BGEFM. (c) A zoomed-in view of the BGEFM, each entry inside is a color-encoded temporal pattern.

approx. 90% entries (approx. 4.2 million of approx. 4.6 million) were well matched (PCC $\geq$ 0.5), and we also assigned a black color to those entries that not well matched or learned. Figure 4.5(a) showed the color-encoded BGEFM that enabled global observation of the temporal associations between genes (in rows) and structures (in columns). At the top part of the BGEFM are the color labels of the entire brain structures, which encoded based on ADMBA reference color code. It should be noted that, for a clear and intuitive visualization, we simplified the structures into three parts (forebrain, midbrain, and hindbrain) and several interesting regions (hypothalamus, isocortex, hippocampus, thalamus, cerebellum, and pons) which is indicated intuitively through the top-right legend (Figure 4.5(b)).

4.4.2 Observations on Exploration of the ADMBA Dataset

We used the BGEFM, as the second visual analytics component in this dissertation, to represent the temporal associations between genes and structures. Since we color-encoded the temporal pattern indices of entire genes across entire structures, users can intuitively observe how and where the gene expressed over time. Based on the visualization result, we explored several gene expression characteristics in the developing mouse brain. We next investigate these patterns through the DAVID (Database for Annotation, Visualization, and Integrated Discovery) enrichment analysis, and manifest how they helped the scientists in their research.

Genes are playing certain roles in the brain development.

The first observation is that the genes expressed in certain ways in the local brain regions. From the BGEFM, it is clear that the major colors of the matrix entries were consistently represented in most of the bi-clustered correlated sections. We investigated this

observation by taking the 3^{rd} structure cluster in the BGEFM as an example, where most structures here belong to the thalamus (structure color labeled as yellow). In this structure cluster, the color of rows (temporal patterns of genes) in each gene clusters are showing specific temporal patterns: the 2^{nd} gene cluster showed strong characteristics of 1^{st} and 2^{nd} temporal patterns (patterns in purple and blue colors); the 3^{rd}, 5^{th}, 8^{th}, and 9^{th} gene clusters showed only high expression values in the middle-stages (patterns in green, yellow, and orange colors); the 1^{st}, 7^{th}, and 10^{th} gene clusters showed strong characteristics of the in late-increasing temporal patterns (patterns in red and cherry colors). These observations are the evidence that specific genes are showing distinct correlations in different brain regions, and those distinct correlations are in the form of the change of expression values through time.

The similarity of gene expressing patterns is correlated with the distance among regions.

In the developing mouse brain, we observed a rule that the closer of two structures in physical space, the similar temporal patterns the genes expressed in. This observation was also grasped from the BGEFM visualization by comparing the primary colors among various correlation sections. We took the 3^{rd}, 6^{th}, 8^{th}, and 9^{th} structure clusters to investigate our observation. Their structure color labels indicated the structures in these clusters only belonged to the forebrain brain regions. Interestingly, the temporal pattern labels in these structure clusters are very similar to each other. For example, most of the genes in the 1^{st}, 6^{th}, 7^{th}, 9^{th}, and 10^{th} gene clusters expressed in the consistently increasing temporal patterns while 3^{rd}, 5^{th}, and 8^{th} gene clusters expressed in the increasing-decreasing patterns and 2^{nd} gene clusters showed decreasing patterns. However, contrasting to the forebrain,

the distinct patterns in the hindbrain (the 2^{nd}, 4^{th}, and 5^{th} structure clusters) were noticeable. The distinction came from the different temporal patterns of the genes in 6^{th} and 10^{th} gene clusters-surging in the hindbrain but consistently increasing in the forebrain. Thus, base on the visualization of BGEFM, the temporal pattern labels in each structure cluster is similar to those that contained the structures in the same or neighboring brain regions. Therefore, based on the above observations, we made a hypothesis that the genes expressed in certain temporal patterns in the whole brain, and the closer in physical space, the fewer patterns change.

Genes could be categorized into two subsets: region-agnostic and region-specific.

The region-agnostic genes were those genes that expressed in certain temporal patterns across all brain regions, and in our case, are those in the 1^{st}, 2^{nd}, 3^{rd}, 7^{th}, 8^{th}, and 9^{th} gene clusters. As clearly represented in the BGEFM, the 1^{st} and 7^{th} gene clusters were showing sharply increasing expression patterns (patterns in hot colors as cherry and red) at entire brain structures. Similarly, the 2^{nd} gene cluster in the BGEFM showed only decreasing expression patterns (patterns in purple and blue colors) in all brain regions. Nevertheless, the genes in the 3^{rd}, 8^{th}, and 9^{th} gene clusters showed fluctuant patterns across all regions. From the enrichment analysis result, we found the genes in the 2^{nd} gene cluster were strongly enriched in neuron and cell development. However, the genes in the 3^{rd} gene cluster had high proliferation- and metabolic-related enrichments. Thus, we could make a biological hypothesis based on the integrative result: the 2^{nd} cluster genes only activated at the early stages to build up the foundation of the brain-the neuron cells, and during this progress, the genes in the cluster started to work for the cellular activities.

On the other hand, region-specific genes expressed depending on the specific locations, and their temporal patterns varied at different brain region. For example, the genes in the

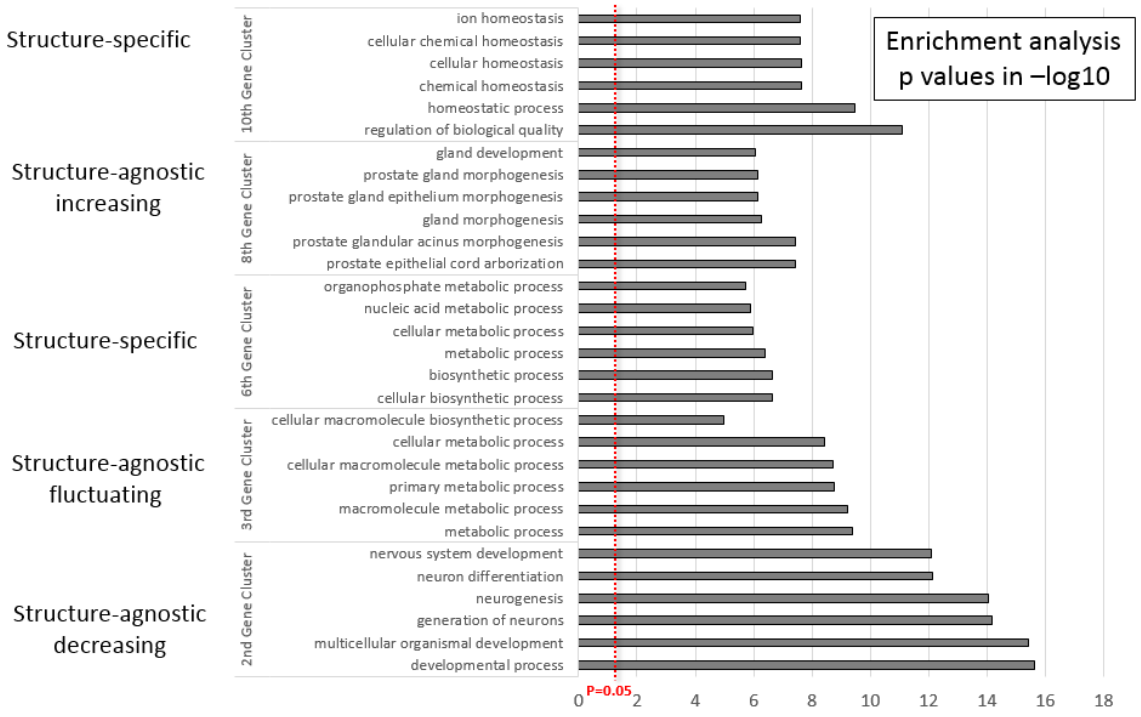

Table 4.1: **Enrichment results of the genes in different gene clusters in the ADMBA data.** The p values are in the $-log_{10}$ space.

6^{th} and 10^{th} gene clusters showed sharply increasing patterns in the forebrain but decreasing patterns in other regions. Moreover, our enrichment analysis result showed these genes enriched in homeostasis firmly. Hence, we may make another biological hypothesis that hindbrain stopped to grow earlier than the forebrain. Table 4.1 gave out the entire enrichment results of the genes in different gene clusters. Therefore, we believed these explored patterns reflected the development of the mouse brain and were suitable for biologists and neuroscientist to enhance their studies.

The temporal patterns could be interpreted by neural events.

The explored temporal patterns could also be interpreted by well-known neurodevelopment events. We analyzed the learned temporal patterns using the mice brain developing events database provided by the TranslatingTime [6]. Since the time-varying functionality could best interpret the temporal genomic characteristic, we only focus on the fluctuant events, i.e., onset-peak-offset. Figure 4.6 showed the two example temporal patterns that we learned from the ADMBA data, the 1^{st} and the 9^{th} one. As showed, the 1^{st} temporal pattern starts with a peak expression and then decreases until the E18.5 stage. As per the mice neural events on TranslatingTime.net, several neural system developments showed offset within the first stages. Interestingly, either the functionality of these events or the affected regions had a strong correlation with the elaboration of the neural system, and also corroborated the enrichment results (genes in the 2^{nd} gene cluster of BGEFM were strongly enriched in neuron and cell development). This revealed that the murine brain is starting to form its core-neuron system at the very early stages and then subsequently constructs the supporting or collaborating structures.

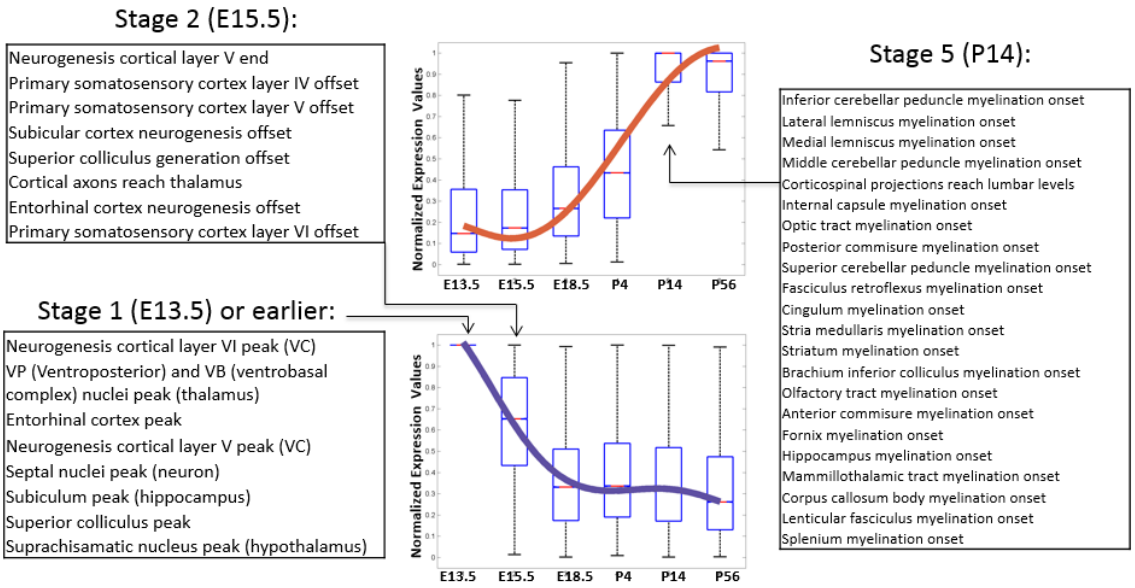

Figure 4.6: **Two example temporal patterns that learned from the ADMBA could be well-interpreted by the neurodevelopment events.** The 1^{st} temporal pattern starts with a peak expression and then decreases until the E18.5 stage. This temporal property highly matched the neuron-related events. The 9^{th} temporal pattern increases from the P4 stage, and this pattern could be interpreted by the myelination-related events.

On the other hand, the 9^{th} (and also the 10^{th}) temporal pattern increases from the P4 stage, and this pattern could be interpreted by the myelination-related events. Also, enrichment analysis showed the related genes (as the genes in the 7^{th} and the 10^{th} gene clusters of BGEFM for the forebrain region) were highly enriched in biological quality and homeostatic processes. Thus, we believe these observations revealed that from the 4^{th} stage onwards, the brain finished the basic construction and subsequently supporting tissue and supporting structures are stabilized and matured. Therefore, we believe the approached framework provided a valuable result for the region-specific temporal patterns exploration and these results potentially lead to future studies.

4.5 Implementation and Result for the Developing Human Brain Dataset

Similar to the process in Section 4.4, we applied the temporal modules-learning method into ADHBA data. Different to the mouse data, the human data contains much more genes and stages, but much fewer regions. As described, ADHBA data contains 52376 genes, 31 stages, and 26 regions, we first generated the 3-D data matrix for the entire data and then applied the data filter. However, since the ADHBA data contains missing entries inside, we first filtered the missing data to improve the robustness for future analyses. For those genes, stages, or regions that were not fully examined, we simply removed them. After filtering these missing entries, a 3-D matrix data in the size of 15867*16*16 was kept (16 available regions are shown in Figure 2.2(c), 16 available stages are shown in Figure 2.2(b)). In order to maximize the performance and provide the most significant genes for bioinformatics analyses, we next used the central moment filter to reduce the data size. Balancing the performance (4-Cores Intel Core i7 CPU and 16GB RAM), we chose the top 1350 genes (approx. 8.5%) of the temporal profiles that had largest variance and significant skewness

for the next pattern-learning module (in log_2 space). However, once again, although this process gleaned the most significant genomic properties in the developing brain, it is not guaranteed that all inherent patterns will be grasped.

4.5.1 Temporal Patterns and BGEFM

Since the data size after the filter is small enough to provide adequate progress, we kept the entire filtered temporal profiles for the pattern-learning process. For the ADHBA data, we repeated the K-Means clustering for each reasonable K ($2 \leq K \leq 20$) and used silhouette to choose the best K. Figure 4.7 showed the max, mean, median, and min values of silhouette for each K. As shown in Figure 4.7, to balance between performance and precision, we chose K as 8 as the number of temporal patterns. Figure 4.9 shows the eight temporal patterns that were learned from the ADHBA data as well as the box-plot showing the distribution of normalized gene expressional values for each pattern. Figure 4.7(a) shows the BGEFM and the cluster dendrogram for the brain regions. The bi-clustering process in BGEFM grouped the genes into 10 clusters. In addition, we also performed a correlation analysis between the brain regions as shown in Figure 4.7(b).

Similar to the process in Section 4.4, we next visualize the temporal associations between genes and structure using BGEFM. Since we only have 16 brain regions, when seeking the correlated sections, we applied hierarchical cluster to explore the dendrogram among the brain regions. Figure 4.6(a) showed the eight color-encoded temporal patterns while (b) displays the BGEFM where its entries indicate the temporal pattern types. The structure names are shown on the top of BGEFM.

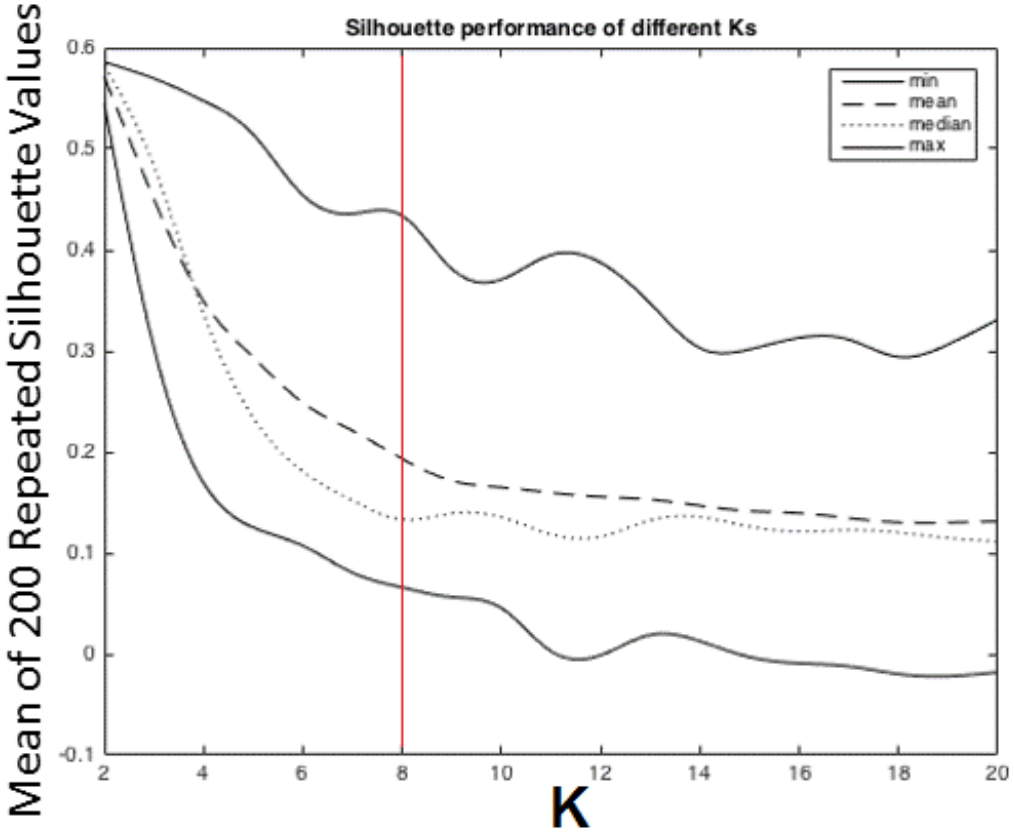

Figure 4.7: **Using silhouette properties to choose the best temporal pattern-learning result for the human brain.** For the temporal pattern-learning process in the ADHBA data, we repeat the K-Means cluster for each K from 2 to 20. Based on the result, we choose K as 8 for the following work (values are curved using a smoothing spline).

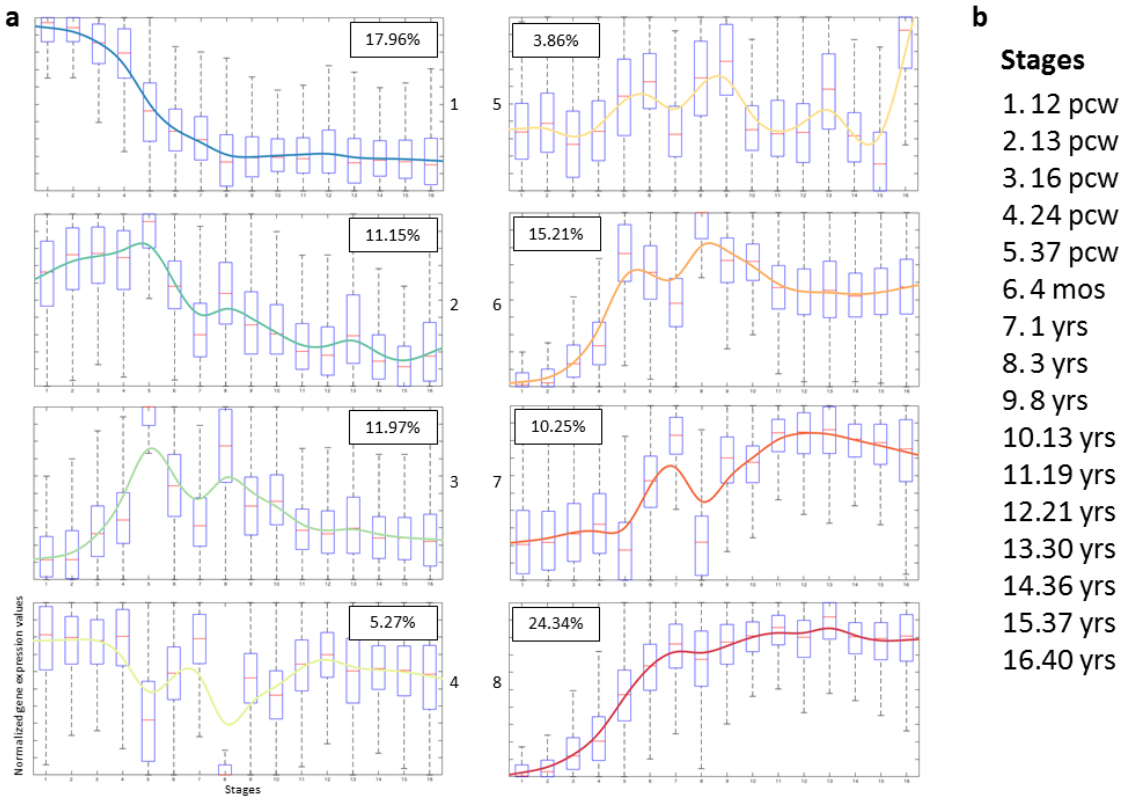

Figure 4.8: **The eight temporal patterns learned from the ADHBA.** (a) We plot the temporal patterns that were learned from the ADHBA data. They are sorted into a decreasing-to-increasing order. For each pattern, the box-plot shows the distribution of gene expressional values (data is processed in log_2 space, and normalized into the range of [0,1], as visualized). (b) The 16 filtered temporal stages.

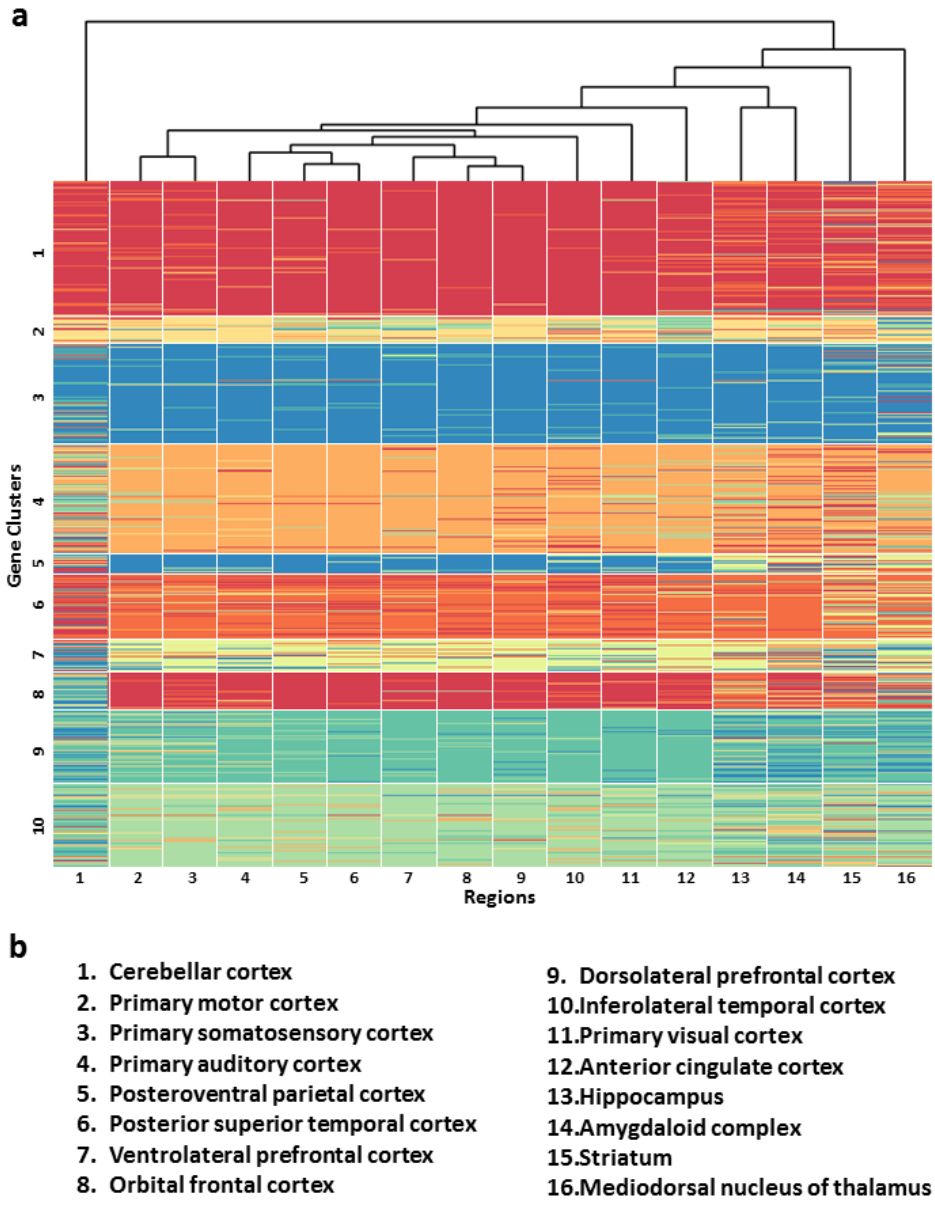

Figure 4.9: **The BGEFM visual analytic component of the ADHBA data.** (a) The entire BGEFM of the ADHBA data. 1350 filtered genes are in rows and 16 brain regions are in columns. The genes were clustered into 10 groups as indicated using the gene cluster. The cluster dendrogram of brain regions is shown on top. (b) A regional correlation visualization among these 16 regions. (b) 16 brain regions that corresponded to the columns in (a).

4.5.2 Observations on Exploration of the ADHBA Dataset

We used BGEFM to visualize the color-coded temporal associations between genes and brain structures as shown in Figure 4.9(a) and then examined some identified patterns to demonstrate the usefulness of the proposed visual analytics framework. To interpret the temporal patterns, we used neurodevelopment events from *TranslatingTime.net* [6]. Also, for gene-of-interest (GOI), we used the NIH Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 to perform the enrichment analyses [5]. Moreover, through the analyses, we also show how these region-specific temporal patterns enhanced the biological knowledge on brain development.

Genes play certain but various roles in different brain regions during the development.

The first observation is that the genes expressed in certain ways in local brain regions. From the BGEFM, it is clear that the colors of the matrix entries were consistently represented in most of the bi-clustered correlated sections. For instance, the 1st gene cluster showed a very consistent red color among all brain regions. According to our color-coding method, these genes had an increasing expressional pattern (the 8^{th} temporal pattern in Figure 4.8). Starting from the early embryonic stages, these GOIs started to perform protein synthesis that controlled the developmental phenotypes and reached the peaks at early childhood (7^{th} stage). More importantly, the expression values of these GOIs kept at high level and continued long into old ages. Besides this, the genes in the 6^{th} gene cluster also showed consistent patterns that majorly based on the 6^{th} and 7^{th} temporal patterns (increasing but in different ways), and the enrichment analysis showed these genes strongly enriched in biological functions of synaptic transmission and signaling as well (Table 4.2).

Stages	Events
1 - 4	Hippocampal commisure appears
	Corpus callosum appears
	Pulvinar projections in the intermediate zone of the developing pre striate iso cortex are present
	Range of rapid synaptogenesis (VC) start
	Pulvinar projections in the cortical plate of the developing prestriate isocortex are present
5 (birth)	Anterior commisure myelination onset
	Brachium inferior colliculus myelination onset
	Optic radiation myelination onset
	Fornix myelination onset
	Auditory radiation myelination onset
	Lateral Geniculate Nucleus Myelination Onset
	Mammillothalamic tract myelination onset
	Cingulum myelination onset
	Fasciculus retroflexus myelination onset
6 - 7	Optic tract myelination end
	Visual cortex peak synaptic density
	Range of rapid synaptogenesis (VC) end
	Corpus callosum body myelination end
	Middle cerebellar peduncle myelination end

Table 4.2: **Significant neurodevelopment events of the human in various stages.** The neuron-related regions and functions started to develop at very early stages and stopped at the 4^{th} stage, and only the 1^{st} and 2^{nd} temporal patterns best matched this. The myelination functions have significant change in the 5^{th}, 6^{th}, and 7^{th} stages. The 3^{rd} and 6^{th} temporal patterns best matched these changes.

In contrast, the genes in the 3^{rd} gene cluster showed an opposite pattern — a consistent decreasing pattern (the 1^{st} temporal pattern in Figure 4.8). Interestingly, these GOIs started from their peak expression levels at the earliest stage, and continually decreased until infancy, then kept this condition to the end. Thus, we believe the genes in the 1^{st} and 3^{rd} clusters have complement functionalities in the entire brain development. Table 4.2 shows the enrichment analysis for the targeted GOIs. From the result, the genes in consistent increasing pattern — those in the 1^{st} cluster — strongly enriched in synaptic signaling and transmission, while the genes in consistent decreasing pattern — those in the 3^{rd} cluster — strongly enriched in neurogenesis and neural system development. This can be best interpreted by the well-known developmental processes that happened in the human brain: from the beginning of the development, brain generates neurons in a high speed until birth; during the neuron generation and neuron system development, synapse starts to connect these neurons into the neural network as well as enables the functionalities in different regions; synapse transmits signals among the neurons in the brain as long as the brain is functioning. Thus, this observation showed the evidence that genes were playing certain roles during the brain development, and more importantly, their temporal expressional characteristics indicated the specific correlations.

The temporal patterns match neurodevelopment events.

We used the neurodevelopment events of human, which are listed on TranslatingTime.net, to interpret learned temporal patterns with key events listed in Table 4.2. While the neurodevelopment events only contain the early stage in human development (i.e., up to approx. 800 post-conceptional (PC) days), we focused on the flows of the temporal patterns before the 8^{th} stage which is around 3 years old.

We first notice that the neural-related regions and functions started to develop at very early stages and stopped at the 4^{th} stage while only the 1^{st} and 2^{nd} temporal patterns matched this condition. Both of them had very high expression values at the beginning stages and started to decrease from the 3^{rd} and 5^{th} stage separately. The second observation included the 5^{th}, 6^{th}, and 7^{th} stages, which strongly matched the myelination functions. In the 5^{th} stage, the 260-300 PC days or the birth of the child, the myelination functions started to onset, while the expression of related genes should have high values or peaks. On the other hand, the myelination functions started to offset from the 6^{th} to 7^{th} stages (350-800 PC days) and the myelination genes should start to decrease the expression levels. Combining these two results above, the 3^{rd} and 6^{th} temporal patterns are best matched. However, the only difference between them was that the gene expression values began to decrease after birth in the 3^{rd} pattern, while in the 6^{th} pattern the expression values still kept in a high level and continued to old age. Based on this difference, we strongly believed these distinct GOIs could lead to future biological studies that targeted on the new myelination biomarkers searching.

Genes could be categorized into two subsets: region-agnostic and region-specific.

Based on the visualization result of the BGEFM, we classified the genes into two categories: region-agnostic and region-specific. The region-agnostic genes expressed in certain temporal patterns across all brain regions while the region-specific genes expressed in different patterns in the specific brain regions.

The region-specific genes expressed in different temporal patterns depended on the locations, and in our case for the ADHBA data, most of the genes showed this behavior. From the visualization of BGEFM, the 2^{nd}, 4^{th}, 5^{th}, 6^{th}, 7^{th}, 8^{th}, 9^{th}, and 10^{th} gene clusters showed strong region-specific property (Figure 4.9(a)). From the hierarchical cluster

dendrogram at the top of the BGEFM, we found that these genes show different patterns in the 1^{st}, 13^{th}, 14^{th}, 15^{th}, and 16^{th} brain region (cerebellar cortex, hippocampus, amygdaloid complex, striatum, and mediodorsal nucleus of thalamus) comparing to the other brain regions (2^{nd} to 12^{th} regions). Interestingly the 2^{nd} to 12^{th} regions are all cerebral cortex regions while the 1^{st}, 13^{th}, and 16^{th} regions are not.

These genes are also associated with important brain functions. For example, the genes in the 8th gene cluster showed very different temporal pattern in the different brain regions: gradually increasing in the 2^{nd} to 12^{th} regions, irregular in 15^{th} and 16^{th} regions, increasing-to-decreasing in the 1^{st} region, but more drastically increasing around the time of birth (37 PCW) in regions 13 (hippocampus) and 14 (amygdaloid complex). As shown in Table 4.3, the enrichment of the corresponding genes showed some learning-behavior related functions, which exactly matches the learning and cognitive functions associated with hippocampus and amygdala.

On the other hand, the region-agnostic genes only were those in the 1^{st} and 3^{rd} gene clusters. From the BGEFM, it is clear that the genes in the 1^{st} gene cluster belonged to the 8^{th} temporal pattern while the genes in the 3^{rd} gene cluster belonged to the 1^{st} temporal pattern. As described, the 1^{st} temporal pattern showed a consistent decreasing trend before the birth and maintained at low expressing values till the last stage. The 8^{th} temporal pattern, however, showed a consistent increasing trend before the birth and kept expressing at high values in all following stages. Since these two temporal patterns were completely complementary, we firmly believe those two groups of genes have complementary functions, as well as they reflected the regulation of the human brain development that the neuron system is only constructed during embryonic period and the synapse starts to be functional around the birth and till old age.

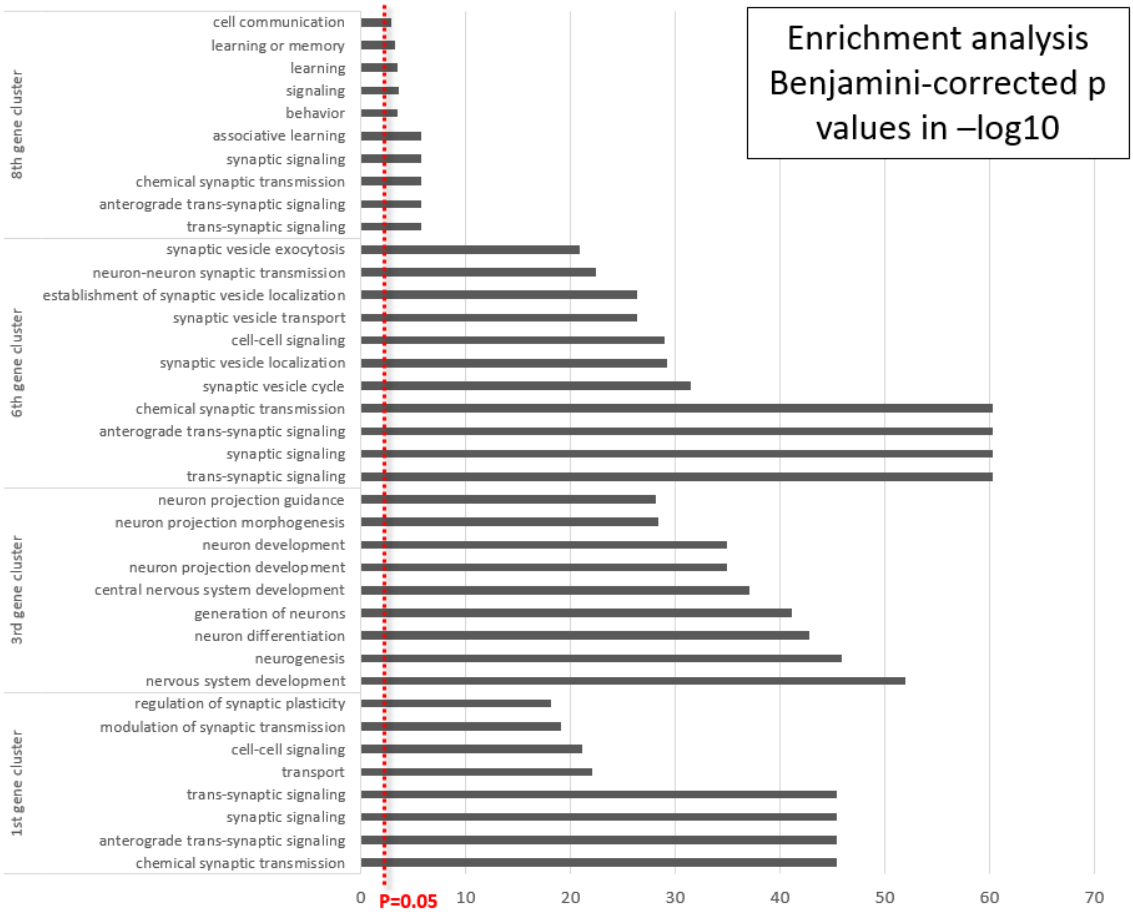

Table 4.3: **Enrichment results of the genes in different gene clusters in the ADMBA data.** The p values are in the $-log_{10}$ space and used the Benjamini correction. The p=0.05 is shown using red-dot line. The genes in the 1^{st} and 6^{th} gene clusters are strongly enriched in synaptic signaling and transmission. The genes in the 3^{rd} gene cluster are strongly enriched in neural-related function and development. The genes in the 8^{th} gene cluster also show enriched functions in learning-behaviors.

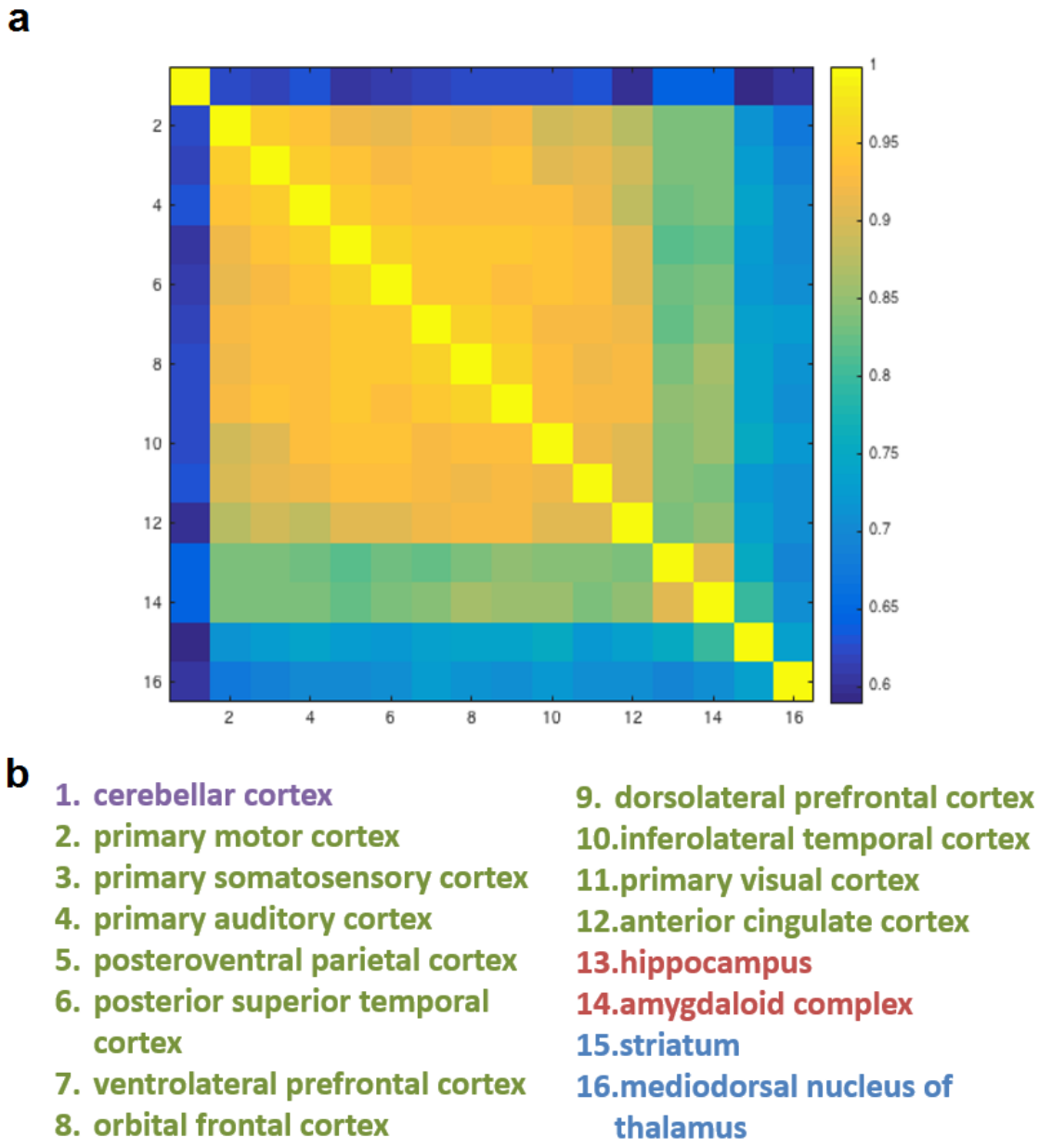

Figure 4.10: **Correlation analysis between all 16 brain regions.** The cerebellar cortex showed the most different patterns to all other regions, and two subsets of regions showed strong distinctions: hippocampus with amygdaloid complex, and striatum with mediodorsal nucleus of thalamus. The correlation also confirmed that genes in the brain regions that are physically close expressed in similar patterns.

Genes in the brain regions that are physically close expressed in similar patterns.

From the BGEFM, as described in the above observation, most genes expressed in different temporal patterns depended on the brain regions. However, not each region was distinguished, and they could be grouped into four subsets. These first distinct region(s) subset only contained the 1^{st} region (cerebellar cortex); the third subset contained the 2^{nd} to 12^{th} regions (primary motor cortex, primary somatosensory cortex, primary auditory cortex, posteroventral parietal cortex, posterior superior temporal cortex, ventrolateral prefrontal cortex, orbital frontal cortex, dorsolateral prefrontal cortex, inferolateral temporal cortex, primary visual cortex, anterior cingulate cortex); the third subset contained the 13^{th} region (hippocampus) and the 14^{th} region (amygdaloid complex); the fourth subset contained the 15^{th} region (striatum) and the 16^{th} region (mediodorsal nucleus of thalamus). Figure 4.10 shows the correlation between all regions. Clearly, the cerebellar cortex showed the most different patterns to all other regions, and two subsets of regions showed strong distinctions: hippocampus with amygdaloid complex, and striatum with mediodorsal nucleus of thalamus. Next, we investigated these regions in brain anatomy. Among these 16 brain regions, 15 of them belonged to the forebrain and midbrain except the cerebellar cortex, which is far away in space. On the other hand, hippocampus, amygdala, and thalamus belonged to the striatum which was part of the midbrain. Especially, the thalamus and putamen belong to the main body of the striatum, while amygdala and hippocampus belong to the tail of caudate nucleus. Thus, this observation was interpreted by the biological grand truth and hence we strongly believed that the region-specific temporal patterns reflected the brain development regulations and could be used for future studies.

4.6 Summary

In this chapter, focusing on the ADMBA and ADHBA data, we proposed a visual analytics component that enabled an integrative exploration of the regional temporal patterns of gene expression in the developing brain. As a part of the overall proposed framework, this component consisted of a data-driven pattern-learning approach and the BGEFM visualization component, and provided a comprehensive pattern exploration as well as an interactive visualization. We first used a data structure $MV_{g,s}$ to collect the entire gene-per-region expression data in the manner of temporal flows, and filtered using the largest variance and the skewness properties in statistics. We next used the $K-Means$ clustering approach to seek the temporal patterns as well as color-coded them in order to perform the visualization. Finally, the BGEFM, which was initially designed in our previous work, served as the overall visual analytics of the regional temporal patterns of gene expression. After implementing to both the ADMBA and ADHBA data, the proposed visual analytics framework provided several salient patterns and observations, and we investigated several of them using the neurodevelopmental events and gene enrichment analyses. Based on the result, several learned patterns could be well-interpreted by the neurodevelopmental events and reflect the developmental processes of the human brain. On the other hand, we also observed several interesting genomic characteristics. First is that we found genes were expressing in certain but various temporal patterns in the brain. More importantly, some of the patterns were consistent in all brain regions while others were not, hence we named them region-agnostic and region-specific genes. Furthermore, we also found that the brain regions that are physically close contained similar temporal patterns of gene expression, while the similarity was correlated negatively with the physical distance among them. These patterns and observations clearly provide scientists valuable information to

gain a deeper understanding of the brain development. Therefore, we claim that our second visual analytics component overcomes the second challenge in Section 1.1: how to enable the learning of the inherent temporal patterns of gene expression in the developing brain.

Chapter 5: Exploration of Spatiotemporal Patterns of Brain Development

In this chapter, we present an integrative method to provide an intuitive presentation of the spatiotemporal pattern of gene expression. We used gene gradient to combine the two types of patterns (spatial patterns and temporal patterns) explored in Section 3 and Section 4. The gene gradient measures how a gene expression changes across the structures in the developing hierarchy. In order to provide an intuitive visualization, we use two measurements to present the gene gradient: the main gradient orientation (MGO) and gradient anisotropy (GA). Again, based on the explored patterns, we will investigate the functional enrichment of genes with distinct gradient properties, and demonstrate how the spatiotemporal patterns reflect the brain development. Once again, since the ADHBA data did not provide the spatial region information, we only perform the spatiotemporal pattern exploration to the ADMBA data.

In this chapter, we first repeat our motivations in Section 5.1. Then, we focus on the spatiotemporal pattern and the gene gradient measurements in Section 5.2, as well as present an intuitive visualization method. Next, based on several explored patterns, we investigate the usefulness of the gene gradient-based spatiotemporal pattern in Section 5.3. Finally, we summarize our work in Section 5.4.

5.1 Movitation

In developmental neuroscience, discovering and visualizing the spatiotemporal gradient of gene expression in the developing brain is an increasingly important topic. The spatiotemporal gradient of gene expression, or so-called expression directionality, indicates the changes of expression behaviors of genes across both space and time. Such a gradient normally provides plentiful and valuable information for scientists. Especially in studies with a focus on the brain development, gradients are commonly used to divide neurogenic regions into distinct functional domains. In the cortex of the mouse brain, for example, rostral-to-caudal (front-to-bottom) gradients are usually involved with tissue development, while exterior-to-interior (outside-to-inside) gradients indicate the cortex layering [30]. Thus, by exploring such gradients in gene expression, scientists are able to capture comprehensive patterns of target transcriptionic characteristics and hence gain a better understanding of the entire program of brain development.

Over the past few decades, numerous studies have developed several approaches to reveal patterns of gene expression [134]. These approaches have yielded remarkable biological insights that span the exploration of these patterns of gene expression in various forms: from regional to global, from temporal to spatial. However, although this body of work has enabled the discovery and exploration of specific types of patterns, few of them have provided an integrative approach for exploring the gradients of gene expression.

In computational bioinformatics, fold change of gene expression is widely used for describing the genetic variation between two situations. Thus, fold changes of expression over developmental stages can denote temporal genetic activity during organ development, and fold changes among brain structures can indicate the spatial gene activities. Therefore,

the integrative spatiotemporal fold change will describe not only how much a gene expression changes but also where these changes occur, and the rate of such a fold change thereby can be best considered as the spatiotemporal gradient. At the same time, anisotropy is a well-known measurement to present the isotropic property in the visualization of spatial diffusion. In diffusion tensor imaging (DTI), fractional anisotropy (FA) — a scalar value between zero and one — is widely used to determine the neural fibers in white matter [135, 136]. For isotropic diffusion, when FA=0, the gradient is uncertain in directionality (equally restricted or unrestricted in all directions); for a direction-restricted diffusion, then FA=1. Thus, since the FA is well suited for the measurement of directionality, we leveraged such a concept and enabled the determination of the directionality of the gene gradient. There is some work in spatiotemporal visualization, however, there is little work in visualizing gradients of gene expression in the developing brain.

However, in order to capture such directionalities as well as perform a visualization, more challenges have to be overcome. For instance, how to capture the expression changes in various structures and across time; how to provide the spatiotemporal fold changes; how to convert the fold changes into gradient; how to provide integrative data structures to represent them; how to statistically measure the gradients; how to leverage FA concept to define and measure the anisotropy; how to visualize the gradient and anisotropy intuitively; and whether the explored gradient and anisotropy are interpretable in biology are questions which still need to be answered.

Focusing on the Allen Developing Mouse Brain (ADMBA) data, we propose a visual analytics method to facilitate the comprehensive exploration of the spatiotemporal anisotropy of the gene gradient in the developing mouse brain. Figure 5.1 shows the entire pipeline of our proposed method: using the gene expression data (as shown in (a)) and the

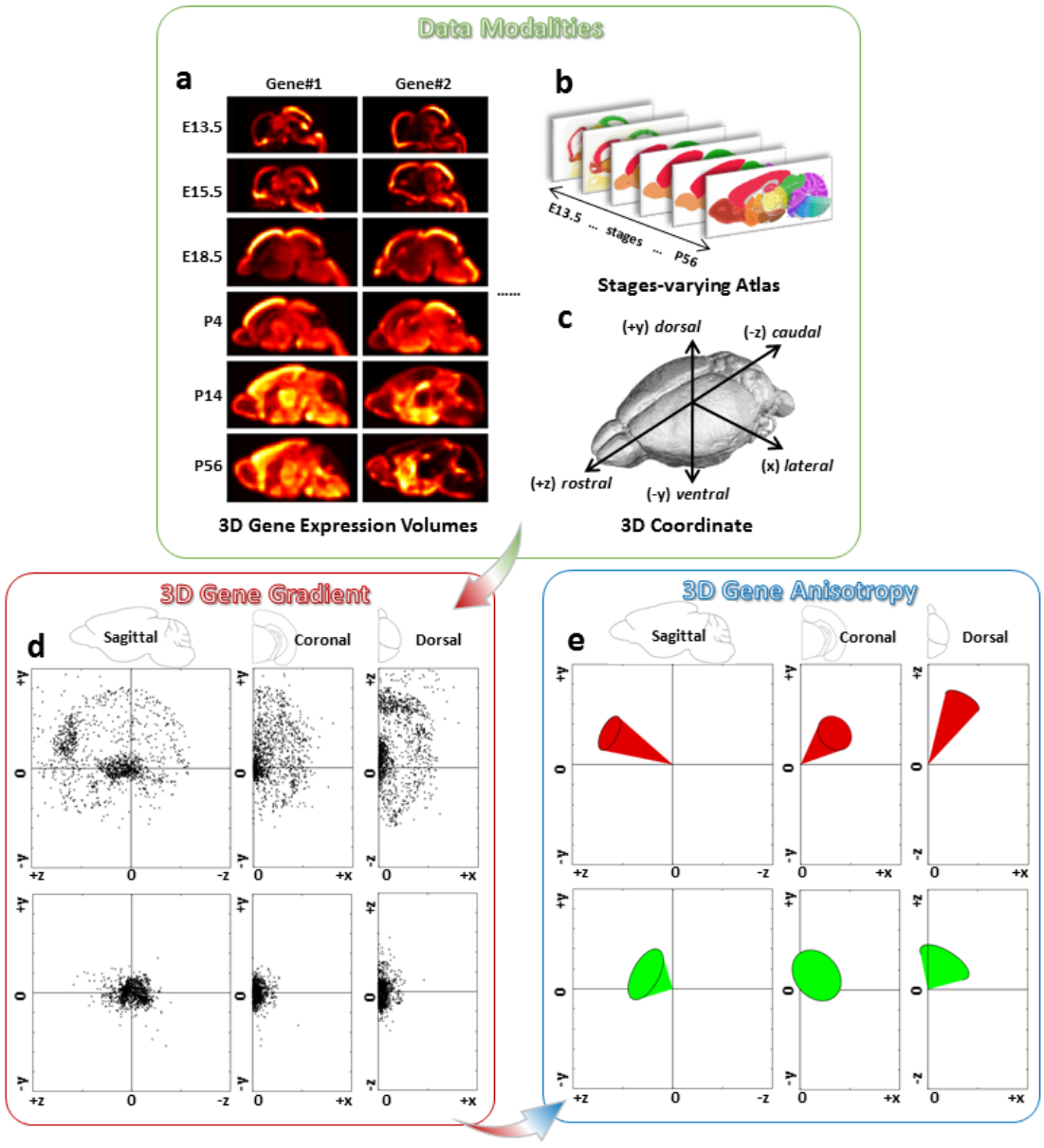

Figure 5.1: **Pipeline of our visual analytics method.** (a) 3D gene expression data of two example genes at six development stages are shown in sagittal view. (b) 3D atlases of brain structures at each corresponding stages from E13.5 to P56. (c) The 3D coordinate of in our method: $< +x >$ represents the left or *lateral* axis (ADMBA data is only available in hemisphere), $< +y >$ represents up or *dorsal* axis, $< -y >$ represents down or *ventral* axis, $< +z >$ represents front or *rostral* axis, and $< -z >$ represents rear or *caudal* axis. (d) The 3D spatiotemporal gradient of two example genes during the entire brain development are shown in sagittal, coronal, and dorsal views. (e) The designed cone-shape visualization approach provided an intuitive presentation of the anisotropy of the gene gradient.

corresponding atlases of structures (as shown in (b)), we first generate the developmental orientation-weighted fold changes of the gene expression in the 3D coordinate (as shown in (c)). Next, for each gene, we integrate its fold changes in the entire structures into the spatiotemporal gradient (as shown in (d)). Moreover, we define GA (gradient anisotropy) and MGO (main gradient orientation) which measure the anisotropy of the spatiotemporal gradient. In addition, we design a cone-shape visualization approach to provide an intuitive presentation of the GA and MGO (as shown in (e)). Finally, we investigate the spatiotemporal anisotropy and the gradient of several genes to examine the precision of the proposed method.

5.2 Visual Analytics Design

In order to explore the spatiotemporal gradient and anisotropy of gene expression, our proposed method uses the entire spatiotemporal fold changes of gene expression across all stages to generate the gradient, and uses two measurements to present it, hence to provide a visualization. We first give a description of the approaches we developed for the exploration of the spatiotemporal fold change. Next, we will introduce the methods that integrate these fold changes into the gradient as well as the measurements that present the anisotropy. Finally, we will describe a 3D visualization component which provides an intuitive presentation of the spatiotemporal anisotropy of gene expression.

5.2.1 Exploring Fold Changes of Gene Expression

In order to explore and visualize the spatiotemporal fold changes of gene expression as described, our proposed method integrates the 3-D development orientation (DO, described in Equation 3.2 in Chapter 3.3) and the fold-change of the gene expression. Specifically, for each gene at the corresponding structures and stages, we use the fold-change of gene

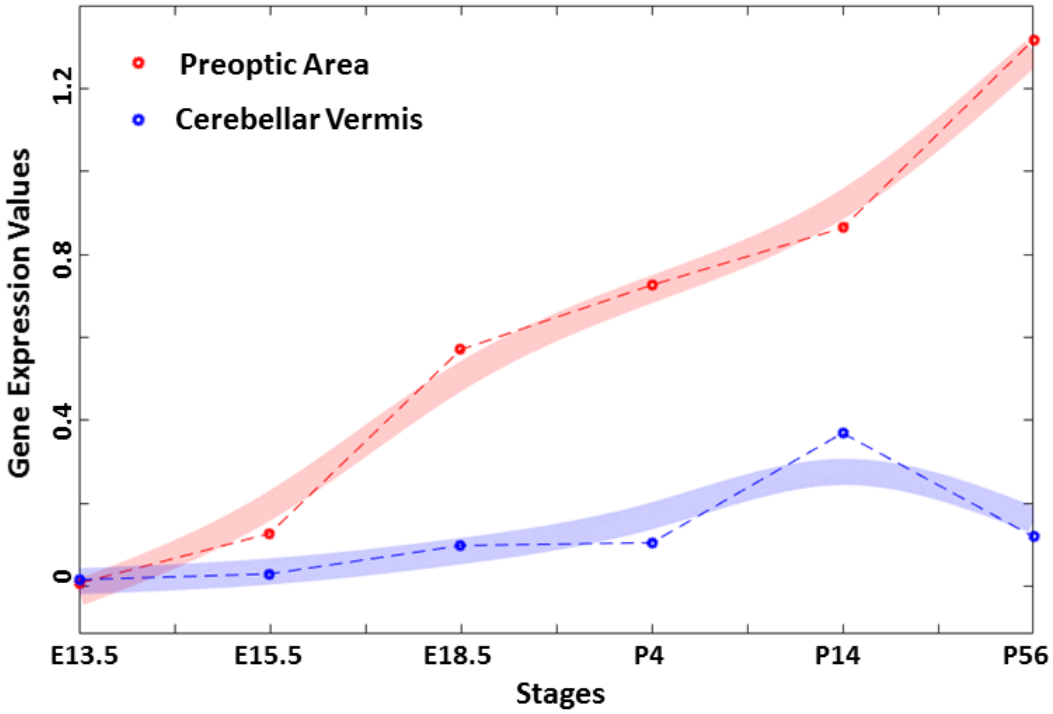

Figure 5.2: **Temporal expression values of gene *Oxt* at two brain structures.** The stage-varying expression values of gene *Oxt* (oxytocin) at structure preoptic area (red curve) and cerebellar vermis (blue curve). In the preoptic area, gene *Oxt* shows a consistent increasing pattern. In cerebellar vermis, gene *Oxt* shows a flat-increasing-decreasing pattern.

expression values to modulated the DOs. In this section, we will describe these steps in greater detail.

We integrated DO with the gene expression values to generate the spatiotemporal fold changes in the form of 3D vectors. Thus, for each DO, its magnitude is weighted by the log_2 fold changes of gene expression between stages and in turn we name them gradient orientations (GOs). Figure 5.2 shows the stage-varying expression values of gene *Oxt* (oxytocin) at structure preoptic area (red curve) and cerebellar vermis (blue curve). In the preoptic area, gene *Oxt* shows a consistent increasing pattern and the log_2 fold changes: from E13.5 to E15.5 is 3.95; from E15.5 to E18.5 is 2.16; from E18.5 to P4 is 0.35; from P4 to P14 is 0.25; and from P14 to P56 is 0.40. In cerebellar vermis, however, gene *Oxt* shows a flat-increasing-decreasing pattern and the log_2 fold changes: from E13.5 to E15.5 is 0.77; from E15.5 to E18.5 is 2.17; from E18.5 to P4 is 0.20; from P4 to P14 is 1.59; and from P14 to P56 is -1.14. Thus, for any given stage t, spatiotemporal fold change $GO(\vec{G_p}, S_i, t)$ denotes the directionality of who gene G_p expressed in structure S_i, and hence the GOs are computed as:

$$GO(\vec{G_p}, S_i, t) = \begin{cases} |log_2(\frac{E_{p,i,t}}{E_{p,i,t-1}})| \cdot D\vec{O}_{S_i}, & \text{if } S_i \text{ exist at both} \\ & \text{stages } t \text{ and } t-1 \\ < 0, \vec{0}, 0 > & \text{otherwise} \end{cases} \tag{5.1}$$

5.2.2 Generating the Integrative Gene Gradient

Since the collection of GOs of genes had revealed the spatial expressing trend in the developing mouse brain, we thereby integrated them into the spatiotemporal gradient of gene expression. For any given gene, this gradient indicated the average change of the spatial expression directionalities across entire developmental stages. Temporal gradient orientation (TGO), which is also included in our preliminary work, indicates the sum of the GOs

for all stages for any given gene. Also, we summarized the TGOs into a vector-layout data structure WGO (weighted gradient orientation), which provided the inherent spatiotemporal patterns of gene expression during brain development, for future visualization:

$$\begin{aligned} &\vec{TGO}(\vec{G}_p, S_i) = \sum_{t=2}^{N} \vec{GO}(\vec{G_p}, S_i, t) \text{ , and} \\ &\vec{WGO}_p = < TGO(\vec{G}_p, S_1), TGO(\vec{G}_p, S_2), \ \ldots \ , TGO(\vec{G}_p, S_N) > \end{aligned} \tag{5.2}$$

Here, N is 2489 due to the total number of brain structures in ADMBA data and $E_{p,i,t}$ is the expression value of the p^{th} gene at the i^{th} structure at the t^{th} stage.

5.2.3 Measuring the Spatiotemporal Anisotropy of the Gene Gradient

Measuring and comparing the learned spatiotemporal gradients of genes, i.e., the WGOs, are our next challenges. In order to provide scale measurements, inspired by the FA (fractional anisotropy) concept from DTI (diffusion tensor imaging), which a scalar value between zero and one (when FA=0, the gradient is diffusional or uncertain in directionality; when FA=1, the gradient has direction-restricted directionality) [135, 136]. Here, we used two definitions to measure the gradient and anisotropy of the WGO for each gene: MGO (main gradient orientation) and the GA (gradient anisotropy). The MGO is the main orientation of the WGO, which is essentially presented as the sum of the entire TGOs. It indicates the direction where the largest gradient happened in the developing brain, hence represents the orientation of the spatiotemporal gradient of gene expression. Therefore, the MGO of each gene G_p can be captured as:

$$\vec{MGO}_p = \sum_i \vec{TGO}(\vec{G}_p, S_i), \tag{5.3}$$

On the other hand, the gradient anisotropy (GA) measures the directionality of expression gradients of any given gene. Here we defined the gradient anisotropy of for each gene G_p as the *log* ratio of the Euclidean norm of the MGO and the variance of the WGO:

$$GA_p = log_2(\frac{\left\|M\vec{G}O_p\right\|}{var(W\vec{G}O_p)}), \tag{5.4}$$

and the variance of the weighted gradient vectors $var(W\vec{G}O_p)$ is calculated by the angles of entire gradient vectors $TG\vec{O}_{G_p,S_i}$ and the $M\vec{G}O_g$:

$$var(W\vec{G}O_p) = \frac{1}{n} * \sum_{i=1}^{n} angle(M\vec{G}O_p, TG\vec{O}_{G_p,S_i}), \tag{5.5}$$

$$angle(M\vec{G}O_p, \vec{GO_{G_p,S_i}}) = \arccos \frac{M\vec{G}O_p \cdot \vec{GO_{G_p,S_i}}}{\left\|M\vec{G}O_p\right\| * \left\|\vec{GO_{G_p,S_i}}\right\|}, \tag{5.6}$$

By the definition described above, GA can indicate the anisotropy property of the gradient of any given gene. After rescaling the GA values into the range from 0 to 1, a lower GA value (closer to 0) means the gene expresses in an expanding way that with low directionality, while a higher GA value (closer to 1) indicates the gene expression has a strong directionality in the corresponding MGO direction.

5.2.4 Visualizing the Spatiotemporal Anisotropy

After we defined and captured the spatiotemporal gradient and the anisotropy of gene expression, we designed a visualization method to enable an intuitive presentation. Focusing on the 3D vector-layout gradient, 3D rendering approach in the Cartesian space is the best method for such a presentation. Thus, we use x, y, and z to indicate the sagittal, dorsal, and coronal axes for our visualization method. In such a 3D space, the origin represents the centroid of the mouse brain, $< +x >$ represents the left or *lateral* axis (ADMBA data

is only available in hemisphere), $< +y >$ represents up or *dorsal* axis, $< -y >$ represents down or *ventral* axis, $< +z >$ represents front or *rostral* axis, and $< -z >$ represents rear or *caudal* axis.

The designed visualization method consisted of two modules: gradient rendering, which provides 3D visualization of the spatiotemporal gradient of gene expression and anisotropy rendering, and uses the cone-glyph to present the anisotropy of the gradient. In the first module, the TGOs of every given gene were used for the visualization. Focusing on the two measurements — MGO and GA, the second module used a cone-glyph to visualize their spatiotemporal properties. In order to indicate a spatial orientation, it is natural to use the direction of MGO to control the central axis of the cone, and the height and radius of the cone reflect the GA value:

$$\begin{aligned} height &= 0.6 * \|GA\| + 0.2, \\ radius &= \frac{\sqrt{1 - height^2}}{2} \end{aligned} \tag{5.7}$$

Moreover, we used red-green colors to encode the cone-glyph in order to provide a more visual understanding of the GA. In this color system, green stands for lower GA while red stands for higher GA. Finally, the histogram of the GA values is also provided, for both regional structures and the entire brain. Figure 5.3 shows the 3D visualization of the gradients and anisotropy of example genes in various brain regions. We will discuss the results in detail in the next section.

5.2.5 Role of Validation

Effective evaluation is necessary for visualization techniques, and most visualization studies require validation processes such as user studies, surveys, case studies, and experiments [137, 138]. However, in the targeted fields of our proposed visualization method,

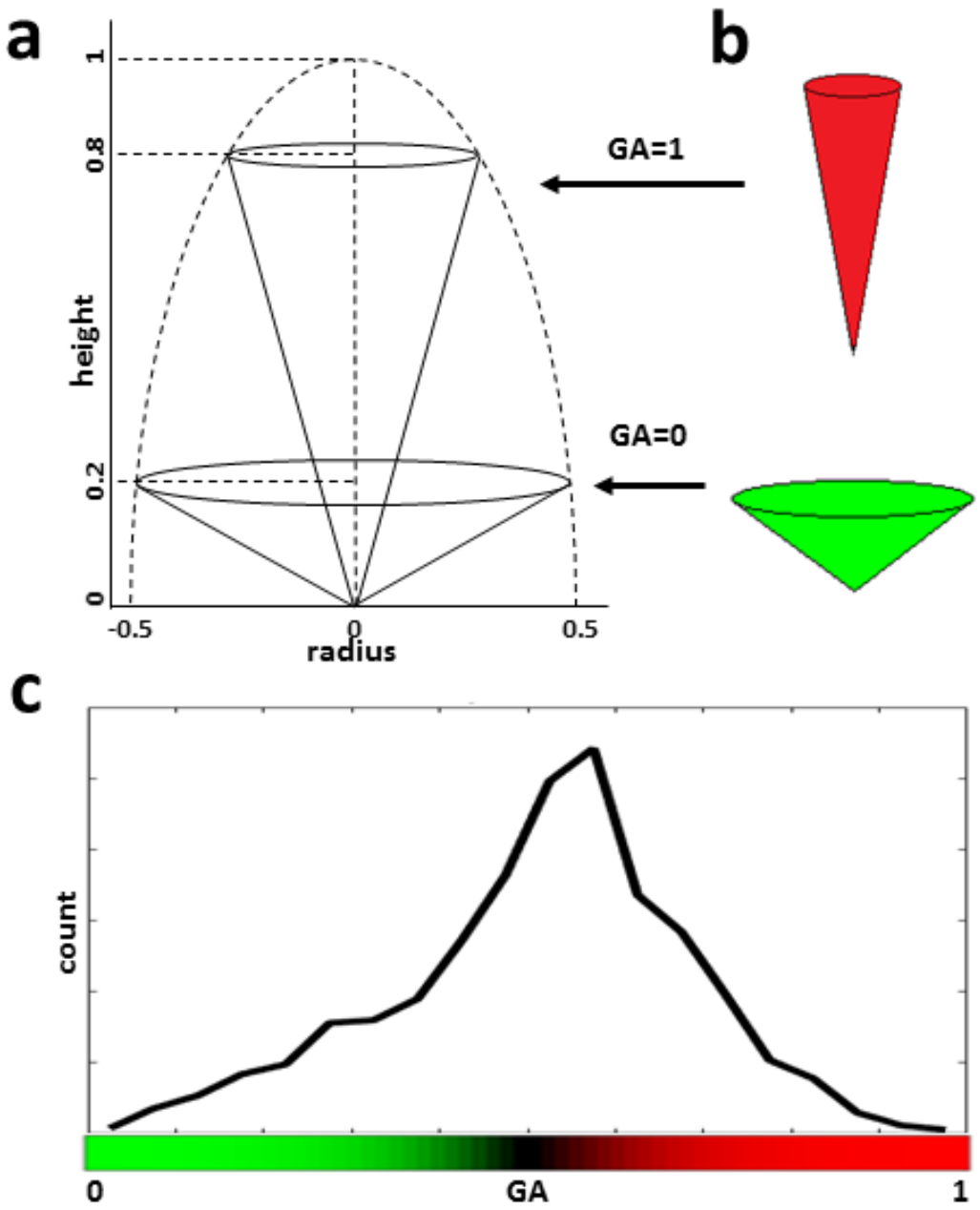

Figure 5.3: **GA visualization method of the cone-glyph.** (a) The designing of the cone-shape glyph shows how we calculate the height and radius of the cone using GA value. (b) The maximum GA value is 1, which are represented in red color, while the minimum GA value is 0 in green color. (c) A histogram shows the distribution of GA values of the entire 1753 genes in all brain structures.

the accuracy of either the gene gradient or the anisotropy can not be guaranteed since the ground truth in this topic is lacking. As described in Chapter 2, no reported work has provided similar visualization approaches in this topic and hence no gradient-relative results have been published in developmental neuroscience which can verify the observations. Thus, in this case, validation of accuracy and user studies are not provided in this work. However, the proposed visualization method can serve as a platform for exploring the gene gradient for scientists in related areas and lead to potential biological hypotheses.

5.3 Result

In order to investigate the performance of the proposed visualization method, we selected several genes that showed distinct spatiotemporal patterns of gradient and anisotropy to perform gene-annotation enrichment analyses.

5.3.1 3D gradient and anisotropy provided a visualization of directionality of gene expression.

Figure 5.4, Figure 5.5, and Figure 5.6 provide the complete gradient visualization of gene *Cdh*24 (Cadherin 24, Figure 5.4), *Rreb*1 (Ras-Responsive Element-Binding Protein 1, Figure 5.5), and *Tubb*3 (tubulin beta 3 class III, Figure 5.6) in the entire developmental period from E15.5 to P56. The first column shows the normalized 3D expression data of these three genes in the sagittal view; the second, third, and forth column present the 3D rendering of the WGO of these example genes in three different views (sagittal, dorsal, and coronal); The final column provides the cone-glyph GA visualization of the anisotropy of the WGO in the sagittal view.

The patterns of the WGO of the example genes, can be clearly seen in the 3-D visualization. As shown in Figure 5.4(a), gene *Cdh*24 has very strong expression gradient in

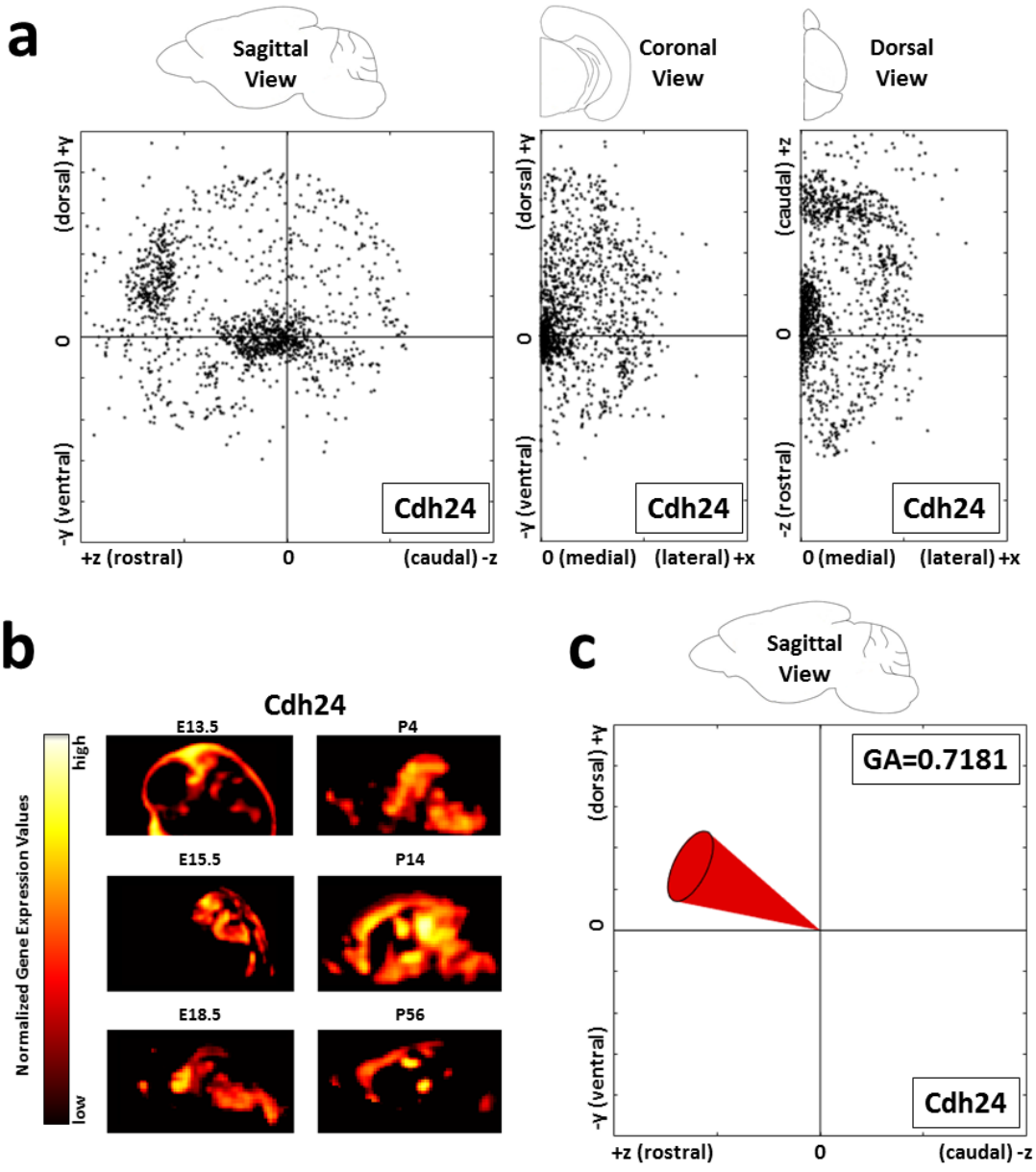

Figure 5.4: **Visualization of gene gradient and the anisotropy.** (a) *Cdh*24 shows relatively stronger density in the $< +x, +y, +z >$ direction. (b) An increasing pattern in the forebrain from E15.5 to E18.5 and in the cortex from P4 to P14 can be clearly seen. (c) Shows a strong anisotropic expression with GA=0.7181.

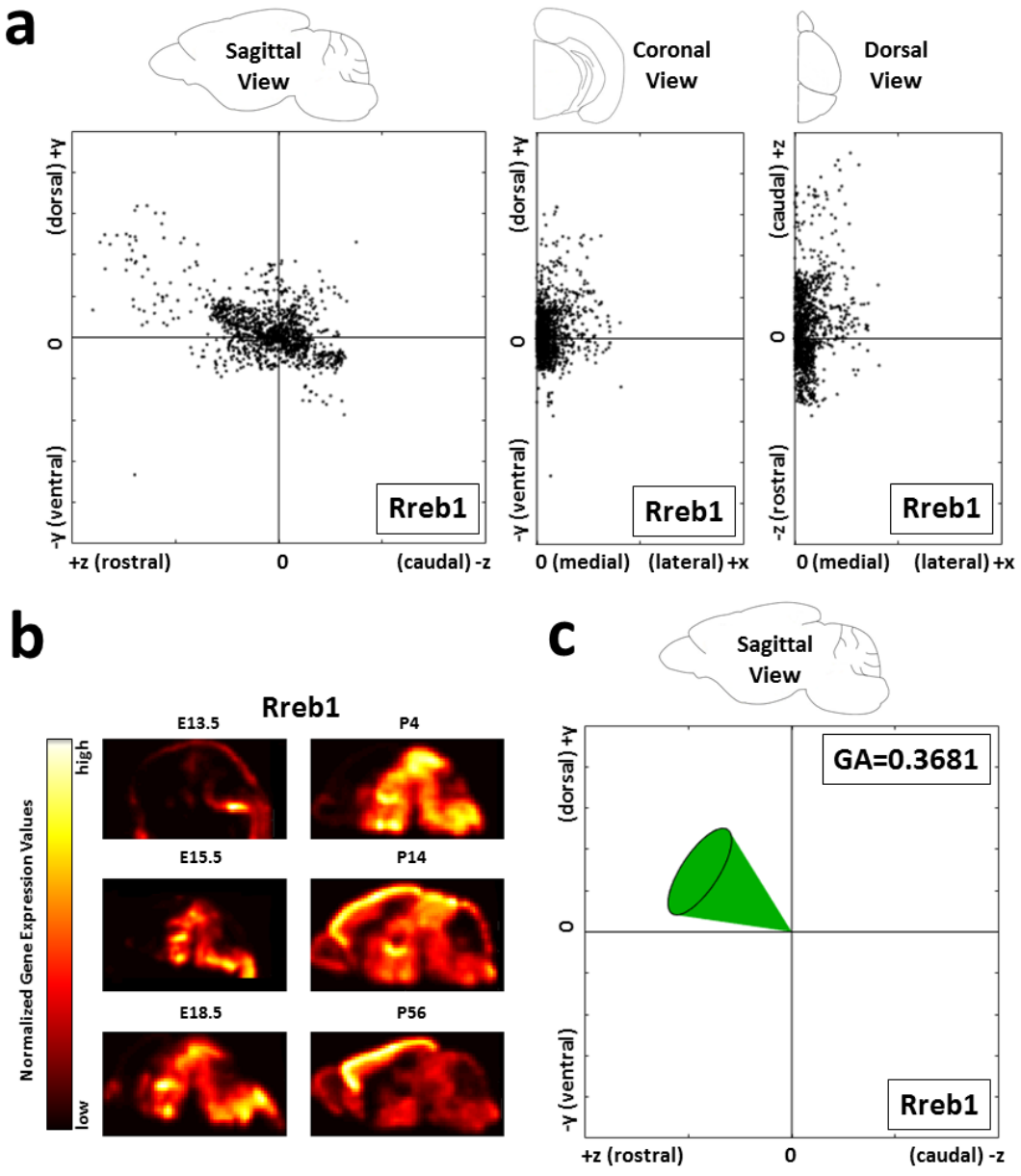

Figure 5.5: **Visualization of gene gradient and the anisotropy.** (a) *Rreb*1 expands the expression locations but the gradient is low except in the $< +x, +y, +z >$ direction. (b) This expanding behavior can be clearly seen from the lower hindbrain and the midbrain (13.5 and 15.5) to the entire brain regions and specially in the cortex. (c) This behavior leads to a slight diffusional anisotropy with GA=0.3681.

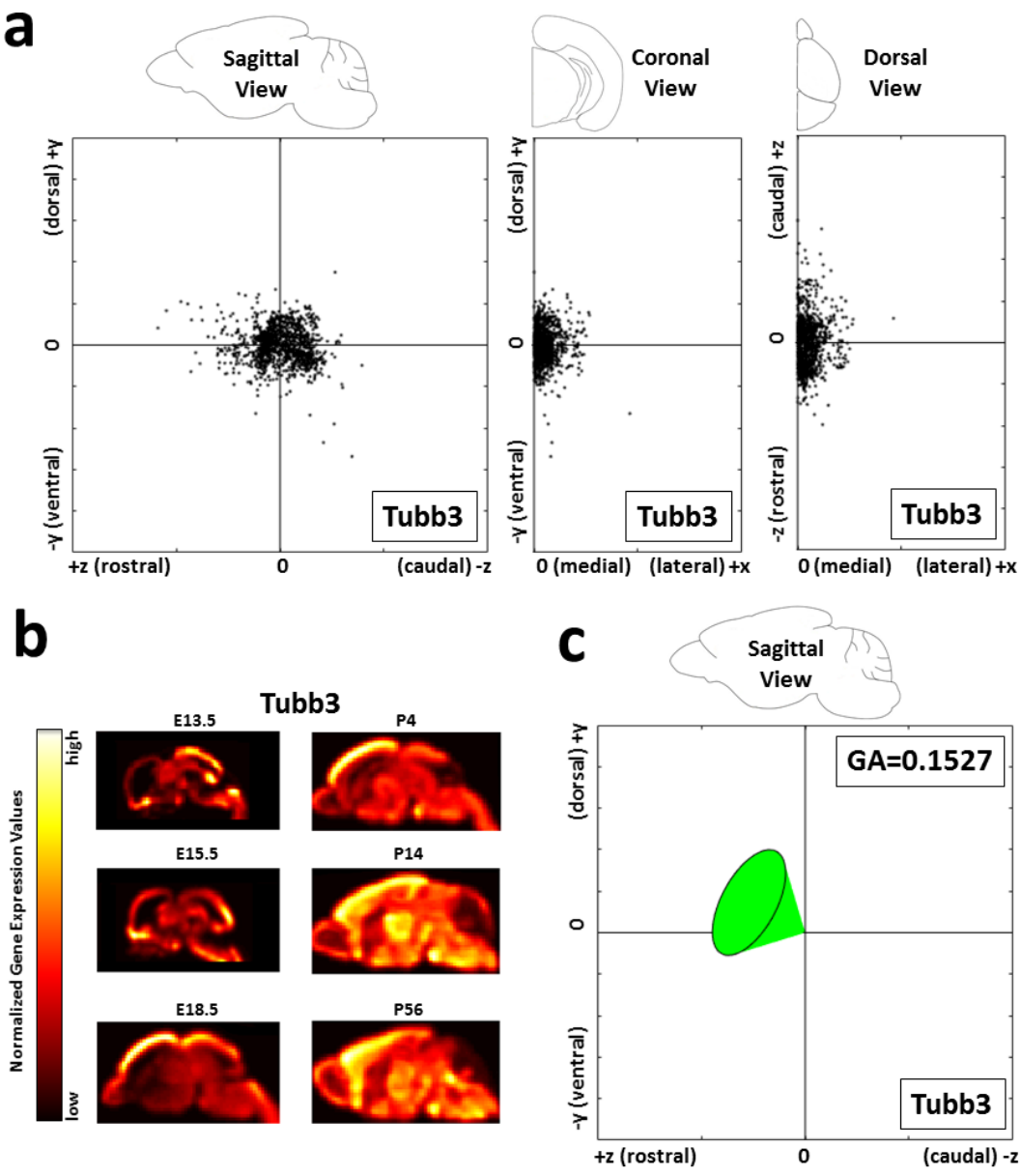

Figure 5.6: **Visualization of gene gradient and the anisotropy.** *Tubb*3 shows low gradient n the brain which the GA=0.1527. No significant directionality or anisotropy can be detected.

the brain and there by lead to the exploration that the expression of *Cdh*24 changed largely and in various directions during the development. More importantly, the gradient of *Cdh*24 shows relatively stronger density in the diagonal $< +x, +y, +z >$ direction. This can be verified through the 3D expression values Figure 5.4(b): a clear increasing pattern in the forebrain from E15.5 to E18.5 and in the cortex from P4 to P14. The cone-glyph visualization of the anisotropy, Figure 5.4(c), presented this observation in 3D rendering in a sagittal view: a thin cone which indicates a high directional anisotropy in red color with the corresponding GA=0.7181. In Figure 5.5(a), gene *Rreb*1 did not only show large expression gradient but also present clear directionalities. However, the expanding behavior of the expression can be clearly detected in Figure 5.5(b): from lower hindbrain and the midbrain (13.5 and 15.5) to the entire brain regions and specially in the cortex. The WGO rendering clearly showed this diagonal direction between $< +z >$ and $< -z >$ with few significant large gradients toward $< +x, +y, +z >$ orientation. Therefore, in Figure 5.5(c), the GA visualization of *Rreb*1 presented a fairly diffusional result with the GA=0.3681. Similar to *Rreb*1, in the Figure 5.6, the third example gene *Tubb*3 also showed a diffusional expression behavior. Although the expression in the cortex and thalamus showed consistent higher values in later stages, the global gradient did not present a large variance and hence both WGO and GA visualization showed this diffusional pattern in the 3D rendering (Figure 5.6(a) and (c)). Therefore, the the 3D rendering of the WGO and the cone-glyph GA visualization provided a precise and intuitive discovery of the gene gradient and the isotropy in the developing mouse brain.

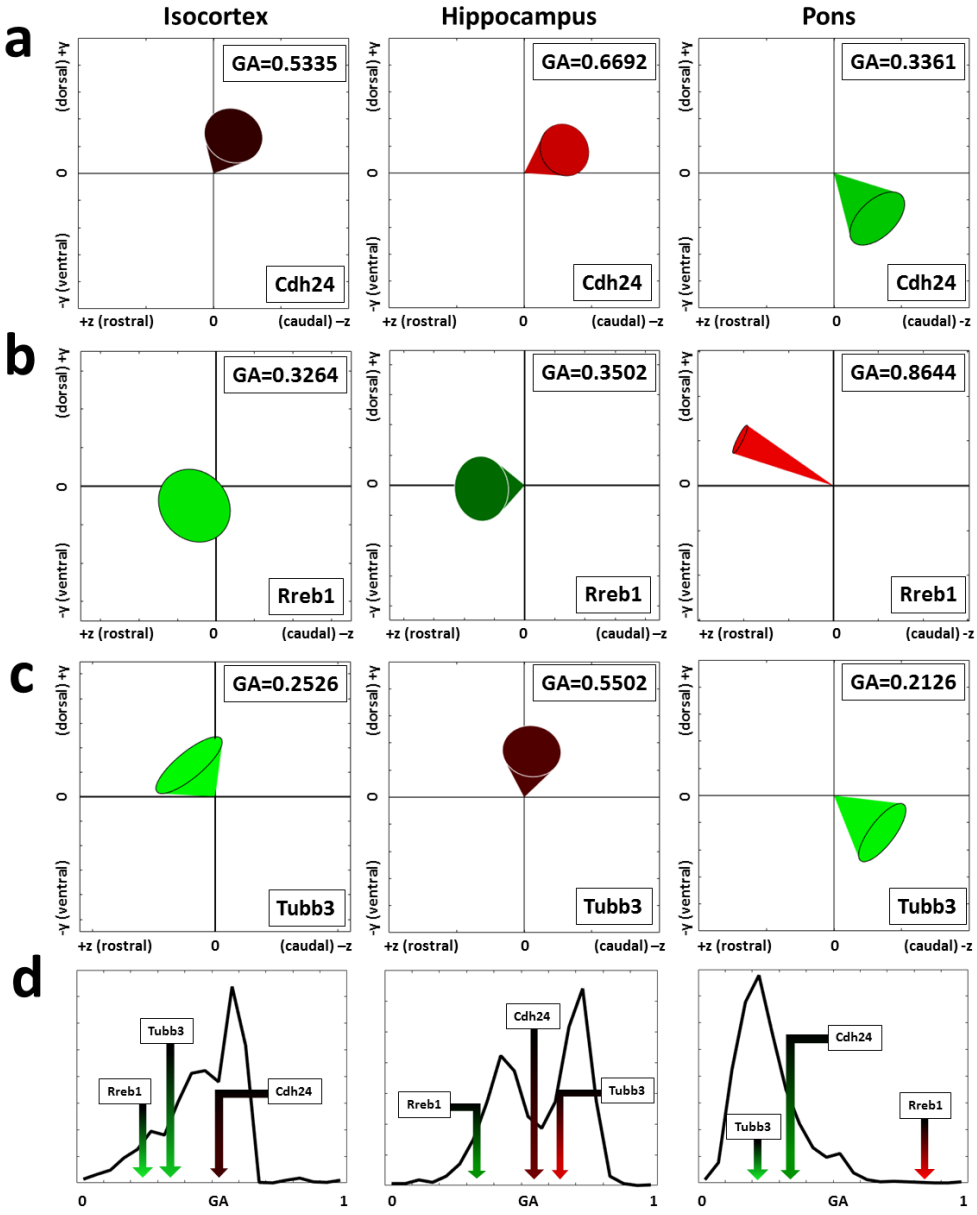

Figure 5.7: **Visualization of GA in various local regions.** Each column denotes a single brain region while each row denotes a gene: (a) *Cdh*24, (b) for *Rreb*1, and (c) for *Tubb*3. (d) The histogram of GA distribution of all genes in the corresponding region, with the example genes highlighted.

this phenomenon, we used GA values to separate the targeted genes-of-interest (GOI) and perform genomic enrichment analyses.

5.3.3 Genes with specific gradient and anisotropy properties enriched in specific biological functions.

In order to answer the question of how the gradient and the anisotropy reflect the biological process, we investigated the GOI with significant gradients using Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 [5]. Based on the WGO and GA, the investigation included two parts: analysis that focused on the GOI with the largest MGO in certain directions, and analysis that concerned the GOI with the smallest and largest GA values.

In the first part, we chose two groups of GOI: genes with the top 100 largest MGO in $< +x >$ direction, and genes with the top 100 largest MGO in $< -z, +z >$ direction. We used the entire 1753 genes as background, and Table 5.1 shows the p values of enrichment. From the result, clearly, the genes with the largest MGO in $< +x >$ direction have strong function enrichment in phosphorylation and metabolic process, while those with largest MGO in $< -z, +z >$ direction have strong function enrichment in the cell and nervous system development. The genes showing strong gradient in medial-lateral (largest MGO in $< +x >$) could be related to the development of cortex layers, where metabolic and phosphorylation played as the major biological processes. On the other hand, since the neuronal system is majorly located in the cortex which develops stronger in caudal-rostral, this observation could also be interpreted in biological accomplishment.

In the second part, we performed genomic enrichment analysis based on the genes in different GA properties. Figure 5.2 shows the enrichment analysis among the top 50 diffusional genes (with smallest GA) and Table 5.3 shows the top 50 directional genes

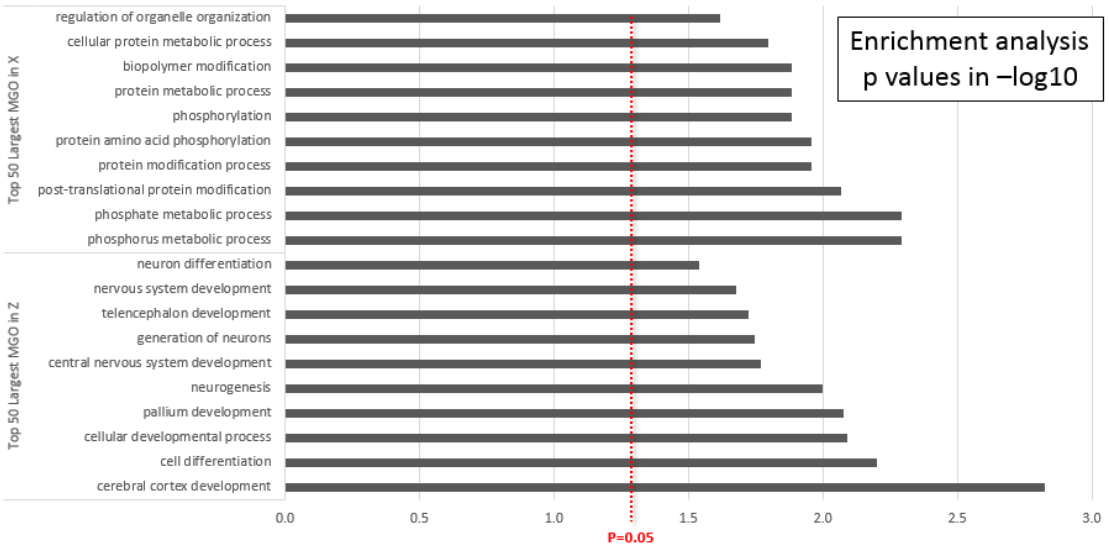

Table 5.1: **Enrichment analysis of the top GOI with significant MGO properties.** GOI with top 100 largest MGO in $< +x >$ direction, and with top 100 largest MGO in $< -z, +z >$ direction were chosen for the analysis (the entire 1753 genes are used as enrichment background). The genes with the largest MGO in $< +x >$ direction have strong function enrichment in phosphorylation and metabolic process, while those with largest MGO in $< -z, +z >$ direction have strong function enrichment in the cell and nervous system development.

a

	Term	Count	%	p value
Top 50 Diffusional Genes	nervous system development	11	24	5.90E-05
	tissue morphogenesis	6	13	4.00E-04
	positive regulation of cellular process	12	26	4.80E-04
	tube development	6	13	6.50E-04
	tube morphogenesis	5	11	1.10E-03
	morphogenesis of an epithelium	5	11	1.10E-03
	morphogenesis of embryonic epithelium	4	9	1.20E-03
	positive regulation of biological process	12	26	1.40E-03
	anatomical structure development	14	30	3.00E-03
	epithelial tube morphogenesis	4	9	3.20E-03

b

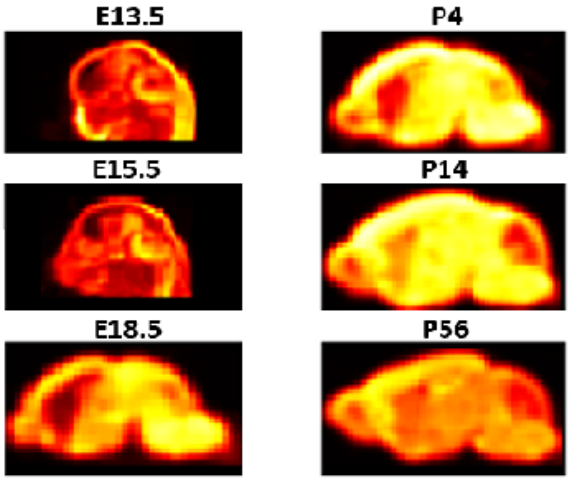

Table 5.2: **GOI analysis based on the 50 top smallest GA values.** (a) The genomic enrichment of the GOI of top the 50 diffusional genes (with the smallest GA values), which enriched consistently in nervous and tube development. (b) Averaged 3D expression data of the sample genes with the smallest GA values.

a

	Term	Count	%	p value
Top 50 Directional Genes	system development	25	54	2.20E-10
	organ development	23	50	2.50E-10
	developmental process	28	61	4.70E-10
	anatomical structure development	25	54	9.40E-10
	anatomical structure morphogenesis	19	41	1.00E-09
	multicellular organismal development	26	57	3.30E-09
	regulation of macromolecule biosynthetic process	25	54	6.40E-09
	regulation of gene expression	25	54	9.60E-09
	regulation of transcription	24	52	1.10E-08
	regulation of cellular biosynthetic process	25	54	1.30E-08

b

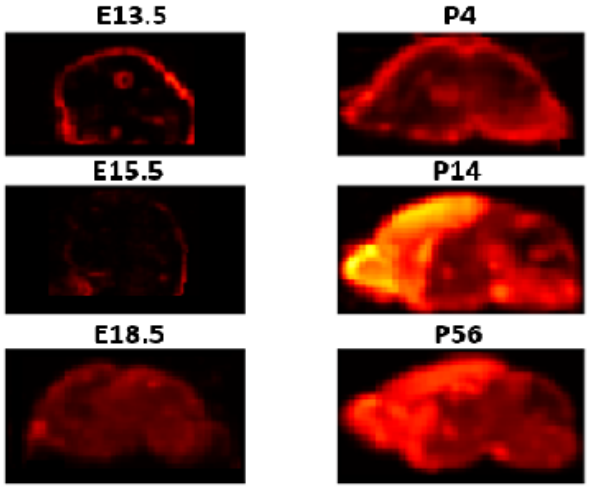

Table 5.3: **GOI analysis based on the 50 top largest GA values.** (a) The genomic enrichment of the GOI of the top 50 directional genes (with the largest GA values) which strongly enriched in structure development. (b) Averaged 3D expression data of the sample genes with the largest GA values.

(with largest GA). Table 5.2(a) shows the enrichment analysis of the top 50 GOI with the most diffusional gradient, and clearly, they had stable enrichment in nervous and tube development. As described in the above observation, the neuronal system is located in the cortex where the developing orientation is diffusing in caudal-rostral, and hence proved this observation. Table 5.3(a) shows the enrichment analysis of the top 50 GOI with the most directional gradient. We found these genes strongly enriched in structure development, which is obviously interpreted by the brain's major developing orientation (caudal-rostral).

Finally, in order to provide an intuitive understanding of the difference between diffusional and directional expressed genes, we also collected several genes from each GOI group and rendered the average expression data in sagittal view. Table 5.2(b) shows the average raw expression data of the top five diffusional genes including *Grik*1, *Tubb*3, *Sema*4*g*, $Mpped1$, and $Nefh$; Table 5.3(b) shows the average raw expression data of the top five directional genes including *Bmp*4, *Cnr*2, *Foxo*1, *Zdhhc*2, and $Fbxw4$. From the rendering, the diffusion group shows an agnostic pattern in the entire brain while the directional group presents a special pattern depending on the location. Thus, the properties of both diffusional and directional expression can be easily observed. Therefore, based on the observations above, we believed most genes expressed in certain gradients during the brain development, and these gradients reflect the development progress in biological accomplishment. More importantly, the proposed visual analytics method provides a precise and intuitive discovery of these gradient as well as the isotropy, which potentially lead to future studies.

5.4 Summary

In this chapter, we proposed a visual analytics method which facilitated the comprehensive exploration of the spatiotemporal gradient and its anisotropy of the gene gradient in the developing mouse brain. Focusing on the ADMBA data, we first generated the 3D spatial developmental orientations of brain structures, and next we used fold changes of the expression values to weight them into the spatiotemporal fold changes of gene expression. Since the spatiotemporal fold change of gene expression was explored through the spatial developmental orientation of brain structures and the corresponding gene expression values, it contained both time-varying expression patterns and structure-development patterns. Therefore, we used such spatiotemporal fold changes of gene expression to capture the gene gradient during the brain development. Consequently, we defined MGO (main gradient orientation) and GA (gradient anisotropy) to measure the direction and anisotropy of the learned spatiotemporal gradient, correspondingly. The anisotropy property of the gene gradient, GA, is a scale value in the range from 0 to 1. For any given gene, based on the proposed method, a lower GA value (closer to 0) means it exhibits a diffusional expression while a higher GA value (closer to 1) indicates its expression has a strong directionality towards the direction of the corresponding MGO. Moreover, we designed a cone-glyph visualization approach to provide an intuitive presentation of the MGO and GA. In this visualization approach, the central axis of the cone represents the direction of the MGO, and the height and radius of the cone are calculated by the GA value.

We also investigated the spatiotemporal anisotropy and the gradient of several genes to examine the precision of the proposed method. Using the DAVID genomic enrichment analysis, several interesting patterns of MGO and GA can be interpreted in biology, and potentially lead to future studies. Therefore, we firmly believe that the proposed visualization

method facilitated the exploration of the spatiotemporal gene gradient and the anisotropy in the developing mouse brain. However, once again, although several observations above were interpretable, the verification of the accuracy of the gene gradient is not guaranteed since the ground truth in this topic is lacking and hence user studies are not provided. Last but not the least, although our work targeted the ADMBA data, it could also serve as an efficient and robust visualization solution for spatiotemporal pattern exploration of multi-dimensional data in a similar data format.

However, the spatiotemporal pattern exploration has its limitations. First, it is still not efficient enough to perform visual analytics on single structure or region. Since the WGO are calculated based on a bunch of structures together, changing from global to local exploration is costly, and the WGO of a single structure is entirely useless for the spatiotemporal pattern exploration. Another limitation is the definition of GO. Since it used the absolute value of the expression fold change, positive (increasing) and negative (decreasing) cannot be distinguished. Thus, we plan to look for a more robust definition of the gene gradient and spatiotemporal pattern measurement in the future.

Chapter 6: Contributions and Future Work

The brain is the most complex biological system in the mammalian body, and recently, the exploration of the mammal brain development has been put under a bright spotlight by neuroscientists. In developing neuroscience, patterns are the keys to describe these changes and hence to learn about brain development. Among various developing patterns, the comprehensive pattern that integrated both spatial and temporal pattern of gene expression becomes the key bridge that links genomics with phenomics in the brain development domain.

In this dissertation, I present an extensible visual analytics framework for exploring the spatiotemporal patterns of gene expression the developing mammalian brain. The data used were collected from the Allen Developing Mouse Brain Atlas (ADMBA) project and Allen Developing Human Brain Atlas (ADHBA) project. In order to overcome the challenges as mentioned earlier in this research field, this dissertation has achieved the following three tasks:

Aim 1. Develop an interactive visual analytics method to present the spatial information of structures based on the AIBS datasets.

Aim 2. Develop a robust method to learn and visualize the inherent temporal patterns of gene expression in developing brains based on the AIBS datasets.

Aim 3. Define an integrative method to provide an intuitive presentation of the spatiotemporal pattern of gene expression.

In Chapter 3, I presented the HOS-Tree, a visual analytics approach that allows the user to explore the spatial patterns of structural development. The HOS-Tree utilizes the prior knowledge of the hierarchical ontology to generate a tree-layout visualization for the brain structures. Focusing on the ADMBA data, this data-driven visual analytics not only learned the natural patterns of structures in the developing brain but also served as a visualization component that showed the gene expression among structures. For each brain structure, the HOS-Tree calculated its spatial development orientation (DO) in vector space and color-coded it in the RGB space. In order to provide precise and robust DO calculation, I also designed an ontology-based structure registration approach which assumes a structure's new location at a higher hierarchical level or a later stage. In this approach, I used the union of entire successor structures at the next level to predict the location of the predecessor structure. I next used two case studies to investigate the performance of the HOS-Tree. The first case study focused on the global spatial developing pattern of all structures, while the second one picked two example genes to analyze the spatial expression pattern at various stages. Both of them indicate the HOS-Tree provides efficient solutions for the parts of 2013 IEEE Data Visualization Contest tasks and, more importantly, is suitable for delivering efficient and intuitive spatial pattern exploration approaches to the users.

In Chapter 4, I presented BGEFM, a data-driven visual analytics approach that focuses on the association between gene and structure in a temporal manner. In order to provide a comprehensive pattern exploration, I first used a data structure $MV_{g,s}$ to store the entire gene-per-region expression data in the manner of temporal profiles, Next, I filtered this data structure using the variance and skewness properties, and applied K-Means clustering

to seek the significant temporal patterns. After color-coding the learned temporal pattern, I used the 2D matrix-layout visualization component — the BGEFM — to present the overall visual analytics. The BGEFM has provided several salient patterns and observations, and I investigated several of them using the neurodevelopmental events and gene enrichment analyses. Based on the result, several learned patterns could be well-interpreted by the neurodevelopmental events and reflect the developmental processes of the human brain. On the other hand, I also observed several interesting genomic characteristics in both developing mouse and human brains. First is that I found genes were expressing in certain but various temporal patterns in the brain. More importantly, some of the patterns were consistent in all brain regions while others were not, hence I named them region-agnostic and region-specific genes. Furthermore, I also found that the brain regions that are physically close contained similar temporal patterns of gene expression, while the similarity was correlated negatively with the physical distance among the structures. These patterns and observations clearly provide scientists valuable information to gain a deeper understanding of the brain development.

In Chapter 5, I presented the gene gradient-based spatiotemporal pattern. The gene gradient measures how a gene expression changes across the developing structures, and hence represents the spatiotemporal pattern of gene expression. In order to achieve this goal, I first calculated the spatiotemporal fold change of gene expression using the DO and the corresponding fold changes between each stage for any give gene. Since the spatiotemporal fold change of gene expression was explored through the spatial developmental orientation of brain structures and the corresponding gene expression values, it contained both time-varying expression patterns and structure-development patterns. Therefore, such

spatiotemporal fold changes of gene expression could be described as the spatiotemporal pattern (so called gene gradient) during the brain development. After converting these spatiotemporal fold changes into spatiotemporal pattern in the format of collection of 3D vectors, I use 3D rendering to present them. Consequently, I defined MGO (main gradient orientation) and GA (gradient anisotropy) to measure the directionality and anisotropy of these learned spatiotemporal gradient, correspondingly. The anisotropy property of the gene gradient, GA, is a scale value in the range from 0 to 1. For any given gene, based on the proposed method, a lower GA value (closer to 0) means it exhibits a diffusional expression while a higher GA value (closer to 1) indicates its expression has a strong directionality towards the direction of the corresponding MGO. Moreover, I designed a cone-glyph visualization approach to provide an intuitive presentation of the MGO and GA. In the 3D visualization approach, the central axis of the cone represents the direction of the MGO, and the height and radius of the cone are calculated by the GA value. I investigate the functional enrichment of genes that have significant and distinct gradient properties and demonstrate how the spatiotemporal pattern reflects brain development.

However, the work described in this dissertation also has some limitations as well as leaves considerable room for future research. Here I list some of the drawbacks and future directions for improvement:

- **The overlying problem in the HOS-Tree.** The major issue is that of the overlay of the structure nodes on the interface due to the large amount of structure nodes. It is a limitation which results from the iteration when calculating the position of the nodes on the tree. Since each node was placed on a certain circle to indicate the developing hierarchy level, the overlay problem in the HOS-Tree cannot be simply solved until

another force-based algorithm was used. Thus, in the future, I plan to look for a robust force-directed method to revise the node overlay problem in the HOS-Tree.

- **The temporal patterns are sensitive to K.** Since the temporal patterns were presented by each of the learned clusters, choosing the K for the clustering process would influence the pattern-learning result. In our cases of ADMBA and ADHBA data, I chose the best K based on the averaged silhouette values. However, this could lead to bias result since not all available patterns were observed, and even worse, user could control their final result of the temporal patterns by choosing appropriate K. Thus, in the future, I plan to involve non-parameter cluster approaches into the pattern-learning module in order to solve this problem.

- **The gene gradient-based spatiotemporal pattern is not robust in small regions.** The gene gradient is not precise enough to perform visual analytics on single structures or regions. Since the WGO are calculated based on several structures together, changing from global to local exploration is costly. For a single structure or small regions, WGO cannot provide the precise gradient and the calculation of anisotropy is entirely useless. This limitation is caused by the definition of gene gradient as well as the data detail and I thereby believe this cannot be overcome until AIBS publish more detailed data sources.

- **Positive (increasing) and negative (decreasing) fold change cannot be distinguished.** The since the calculation of GO used the absolute value of the expression fold change, positive (increasing) and negative (decreasing) cannot be distinguished. Especially, for the genes with lower GA values, it is agnostic to detect whether this gene has strongly changed expression but in large variance, or has weakly changed

expression but in small variance. Thus, I plan to look for a more robust definition of the gene gradient and spatiotemporal pattern measurement in the future.

- **Unable to explore either spatial or spatiotemporal patterns for human data since lacking data.** The spatial pattern exploration (HOS-Tree) and the spatiotemporal pattern exploration (gene gradient, MGO, and GA) were only applied for the ADMBA data. Since AIBS hasn't provided the spatial information of the structures in the of human brain, either type of pattern cannot be explored in the proposed platform. However, this potential work can be finished as soon as AIBS publish the new 3D data for the human structures in the future.

- **Lacking ground-truth validation.** In the spatiotemporal pattern exploration, as mentioned in Chapter 5, the accuracy of either the gene gradient or the anisotropy can not be guaranteed since the ground truth in this topic is lacking. As described in Chapter 2, no reported work has provided similar visualization approaches in this topic and hence no gradient-relative results have been published in developmental neuroscience which can verify the observations. However, the proposed visualization method can serve as a platform for exploring the gene gradient for scientists in related areas, and the future studies of the potential biological hypotheses can be used for the user studies and validating the proposed platform.

Finally, the contribution of this dissertation is providing an extensible visual analytics framework for the exploration of the comprehensive patterns of gene expression in the developing mammalian brain, and the future directions I have listed can further improve the proposed framework. Although this work is only a beginning of the emerging research on the spatiotemporal pattern exploration, in the future, this proposed platform could serve as

a pattern-database for providing reference pattern as normal or control group for related studies, which can lead to more insights into the biological knowledge of brain development.

Bibliography

[1] Genes for autism and schizophrenia switch 'on' in developing brain. http://www.visembryo.com/.

[2] Documentation: Allen mouse brain atlas. http://help.brain-map.org/display/mousebrain/Documentation.

[3] Pavlopoulos GA, Malliarakis D, Papanikolaou N, Theodosiou T, Enright AJ, and Iliopoulos I. Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future. *IEEE Computer Graphics and Applications*, 34(2):24–25, Mar. 2014.

[4] Qihang Li, Zachmann Gabriel, Feng David, Huang Kun, and Machiraju Raghu. 2013 ieee scientific visualization contest winner: Observing genomics and phenotypical patterns in the developing mouse brain. *IEEE Computer Graphics and Application*.

[5] Database for annotation, visualization and integrated discovery (david) v6.8. https://david.ncifcrf.gov/.

[6] Translating time. http://www.translatingtime.net/.

[7] Brain anatomy. http://www.atlantabrainandspine.com/subject.php?pn=brain-anatomy-066/.

[8] Mark H. Johnson. Functional brain development in humans. *Nature Reviews Neuroscience*, 2:475–483, 2001.

[9] Joan Stiles and Terry L. Jernigan. The basics of brain development. *Neuropsychol Rev*, 20(4):327–48, 2010.

[10] Brain development. http://neurosciencelibrary.org/.

[11] De Montfort, Gregory J., and Rosemary Boon. Stages of brain development. 2011. http://www.lispectrum.com/pdfarticles/06_Stages_of_Brain_Development.pdf.

[12] Ackerman S. *The Development and Shaping of the Brain*. National Academies Press, 1992.

[13] Neuron. https://en.wikipedia.org/wiki/Neuron/.

[14] Central nervous system. https://en.wikipedia.org/wiki/Central_nervous_system.

[15] Development of the nervous system in humans. https://en.wikipedia.org/wiki/Development_of_the_nervous_system_in_humans.

[16] Neural development. https://en.wikipedia.org/wiki/Neural_development.

[17] Damir Janigro, editor. *Mammalian Brain Development*. Springer, 2009.

[18] J. Upledger, editor. *A Brain Is Born*. Springer, 1999.

[19] Regulation. https://en.wikipedia.org/wiki/Regulation/.

[20] Laura N. Borodinsky, Yesser H. Belgacem, Olesya Visina Immani Swapna, Olga A. Balashova, Eduardo B. Sequerra, and Sangwoo Shim. Spatiotemporal integration of developmental cues in neural development. *Developmental Neurobiology*, 75(4):349–359.

[21] I. Su Andrew, P. Cooke Michael, A. Ching Keith, Hakak Yaron, R. Walker John, Wiltshire Tim, and P. Orth Anthony. Large-scale analysis of the human and mouse transcriptomes. *PNAS*, 99(7):4465–70.

[22] Tebbenkamp AT, Willsey AJ, State MW, and Sestan N. The developmental transcriptome of the human brain: implications for neurodevelopmental disorders. *Curr Opin Neurol*, 27(2):149–56.

[23] Hyo Jung Kang, Yuka Imamura Kawasawa, Feng Cheng, Ying Zhu, Xuming Xu, Mingfeng Li, Andre M. M. Sousa, Mihovil Pletikos, and Nenad estan. Spatiotemporal transcriptome of the human brain. *Nature*, 478:483–9, Dec. 2010.

[24] RHOSHEL K. LENROOT and JAY N. GIEDD. The changing impact of genes and environment on brain development during childhood and adolescence: Initial findings from a neuroimaging study of pediatric twins. *Dev Psychopathol*, 2010.

[25] AT Tebbenkamp, AJ Willsey, MW State, and N. Sestan. The developmental transcriptome of the human brain: implecations for neurodevelopmental disorders. *Curr Opin Neurol.*, 27(2):149–56, Apr 2014.

[26] F Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–3, 1970.

[27] Gene expression. https://en.wikipedia.org/wiki/Gene_expression/.

[28] Ming an Sun, Zhixiong Sun, Xiaowei Wu, Veena Rajaram, David Keimig, Jessica Lim, Hongxiao Zhu, and Hehuang Xie. Mammalian brain development is accompanied by a dramatic increase in bipolar dna methylation. *Scientific Reports 6*, 32298, 2016.

[29] Epstein HT. Stages in human brain development. *Brain Res*, 395(1):114–9, 1986.

[30] 2013 ieee scientific visualization contest. `sciviscontest.ieeevis.org/2013/VisContest/index.html`.

[31] Yu-Zhong Chen, Zi-Gang Huang, Shouhuai Xu, and Ying-Cheng Lai. Spatiotemporal patterns and predictability of cyberattacks. *PLoS One*, 10(5), 2015.

[32] Andrea E. Gaughan, Forrest R. Stevens, Zhuojie Huang, Jeremiah J. Nieves, Alessandro Sorichetta, Shengjie Lai, Xinyue Ye, Catherine Linard, Graeme M. Hornby, Simon I. Hay, Hongjie Yu, and Andrew J. Tatem. Spatiotemporal patterns of population in mainland china, 1990 to 2010. *Scientific Data*, 3, 2016.

[33] Miriam Grace and Marc-Thorsten Hutt. Regulation of spatiotemporal patterns by biological variability: General principles and applications to dictyostelium discoideum. *PLOS*, 12, 2015.

[34] Samiul Hasan, Christian M. Schneider, Satish V. Ukkusuri, and Marta C. Gonzalez. Spatiotemporal patterns of urban human mobility. *Journal of Statistical Physics*, 151(1):304–18, 2013.

[35] Menno-Jan Kraak and Daniel E. van de Vlag. Understanding spatio-temporal patterns: Visual ordering of space and time. *Cartographica*, 42(2):153–161, 2007.

[36] R. Machiraju, S. Parthasarathy, J. Wilkins, D. Thompson, B. Gatlin, D. Richie, T. Choy, M. Jiang, S. Mehta, M. Coatney, S. Barr, and K. Hazzard. AAAI/MIT Press, 2004.

[37] Sreedevi Varier and Marcus Kaiser. Neural development features: Spatio-temporal development of the caenorhabditis elegans neuronal network. *PLOS*, 1, 2011.

[38] Tricoire L, Pelkey KA, Erkkila BE, Jeffries BW, Yuan X, and McBain CJ. A blueprint for the spatiotemporal origins of mouse hippocampal interneuron diversity. *Journal of Neuroscience*, 31(30):10948–70, 2011.

[39] Tando Maduna and Vincent Lelievre. Neuropeptides shaping the central nervous system development: Spatiotemporal actions of vip and pacap through complementary signaling pathways. *Neuroscience Research*, 94(12):1472–87, 2016.

[40] Oswald Steward. *Principles of Cellular, Molecular, and Developmental Neuroscience*. Springer, 1989.

[41] Yusuke Takeda, Nobuo Hiroe, Okito Yamashita, and Masa aki Sato. Estimating repetitive spatiotemporal patterns from resting-state brain activity data. *NeuroImage*, 133:251–265, 2016.

[42] Huda Akil, Maryann E. Martone, and David C. Van Essen. Challenges and opportunities in mining neuroscience data. *Science*, 331(6018):708–21, Feb 2011.

[43] Lynette Hirschman, Jong C. Park, Junichi Tsujii, Limsoon Wong, and Cathy H. Wu. Accomplishments and challenges in literature data mining for biology. *BIOINFORMATICS REVIEW*, 18(12):1553–61, May 2002.

[44] Data mining. `https://en.wikipedia.org/wiki/Data_mining/`.

[45] Usama M. Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. *From data mining to knowledge discovery: an overview*. American Association for Artificial Intelligence, 1996.

[46] James J. Thomas and Kristin A. Cook, editors. National Visualization and Analytics Center, 2005.

[47] Visual analytics. `https://en.wikipedia.org/wiki/Visual_analytics`.

[48] Raghu Machiraju, Carsten Gorg, and Arthur Olson. Visual analytics for biological data. *IEEE Computer Graphics and Applications*, 34(2):24–25, Mar. 2014.

[49] Allen institute. `http://www.alleninstitute.org/`.

[50] Allen brain atlas data portal. `http://www.brain-map.org/`.

[51] L Zhang, A Zhang, and M Ramanathan. Fourier harmonic approach for visualizing temporal patterns of gene expression data. *Bioinformatics Conference*, Sep 2003.

[52] CC Fowlkes, CL Hendriks, SV Keranen, GH Weber, O Rubel, MY Huang, S Chatoor, AH DePace, and L Simirenko. A quantitative spatiotemporal atlas of gene expression in the drosophila blastoderm. *Cell*, 133(2):364–74, Apr 2008.

[53] M Meyer, T Munzner, A DePace, and H. Pfister. Multeesum: A tool for comparative spatial and temporal gene expression data. *IEEE Trans Vis Comput Graph*, 16(6):908–17, 2010.

[54] VR Iyer, MB Eisen, DT Ross, G Schuler, T Moore, JC Lee, JM Trent, LM Staudt, and J Jr Hudson. The transcriptional program in the response of human fibroblasts to serum. *Science*, 283(5398):83–7, Jan 1999.

[55] G. Parmigiani, editor. *The Analysis of Gene Expression Data*. Springer, 2003.

[56] Hautaniemi Sampsa, Yli-Harja Olli, Astola Jaakko, Kauraniemi Paivikki, Kallioniemi Anne, Wolf Maija, Ruiz Jimmy, Mousses Spyro, and Kallioniemi Olli. Analysis and visualization of gene expression microarray data in human cancer using self-organizing maps. *Machine Learning*, 52(1):45–66, Jul 2003.

[57] M.O. Ward. Xmdvtool: integrating multiple methods for visualizing multivariate data. *Visualization*, 1994.

[58] S. Holter Neal, Mitra Madhusmita, Maritan Amos, Cieplak Marek, R. Banavar Jayanth, and V. Fedoroff Nina. Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *PNAS*, 97(15):8409–14.

[59] MB Eisen, PT Spellman, PO Brown, and D. Botstein. Cluster analysis and display of genomewide expression pattern. *PNAS*, 95(25):14863–68.

[60] DJ Lockhart and C. Barlow. Expressing what's on your mind: Dna arrays and the brain. *Nat Rev Neurosci*, 2(1):63–8, Jan. 2001.

[61] P Bonaventure, H Guo, B Tian, X Liu, A Bittner, and B Roland. Nuclei and subnuclei gene expression profiling in mammalian brain. *Brain Res*, 943(1):38–47, Jul. 2002.

[62] P Carninci, T Kasukawa, S Katayama, J Gough, MC Frith, N Maeda, and R Oyama. The transcriptional landscape of the mammalian genome. *Science*, 309(5704):1559–63, Sep. 2005.

[63] S Gong, C Zheng, ML Doughty, K Losos, N Didkovsky, UB Schambra, and NJ Nowak. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*, 425(6961):917–25, Oct 2003.

[64] Sandberg Rickard, Yasuda Rie, G. Pankratz Daniel, A. Carter Todd, A. Del Rio Jo, Wodicka Lisa, and Mayford Mark. Regional and strain-specific gene expression mapping in the adult mouse brain. *PNAS*, 97(20):11038–43.

[65] ES Lein, MJ Hawrylycz, N Ao, M Ayres, A Bensinger, A Bernard, AF Boe, MS Boguski, and KS Brockway. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445(7124):168–76, Jan 2007.

[66] R. Jones Allan, C. Overly Caroline, and M. Sunkin Susan. The allen brain atlas: 5 years and beyond. *Nature Reviews Neuroscience*, 10:821–28, 2009.

[67] SM Sunkin and JG. Hohmann. Insights from spatially mapped gene expression in the mouse brain. *Hum Mol Genet*, Oct 2007.

[68] M Hawrylycz, L Ng, D Page, J Morris, C Lau, S Faber, V Faber, S Sunkin, and V Menon. Multi-scale correlation structure of gene expression in the brain. *Neural Netw*, 24(9):933–42, 2011.

[69] A. Miller Jeremy, Ding Song-Lin, M. Sunkin Susan, A. Smith Kimberly, Ng Lydia, Szafer Aaron, Ebbert Amanda, L. Riley Zackery, and J. Royall Joshua. Transcriptional landscape of the prenatal human brain. *NATURE*, 508:199–206, Apr 2004.

[70] CL Thompson, L Ng, V Menon, S Martinez, CK Lee, K Glattfelder, SM Sunkin, A Henry, and C Lau. A high resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron*, 83(2):309–23, Jul 2014.

[71] Zeng Tao, Li Rongjian, Mukkamala Ravi, Ye Jieping, and Ji Shuiwang. Deep convolutional neural networks for annotating gene expression patterns in the mouse brain. *BMC Bioinformatics*, 16(147), 2015.

[72] ME Garrett, I Nauhaus, JH Marshel, and EM Callaway. Topography and areal organization of mouse visual cortex. *J Neurosci*, 34(37):12587–600, Sep 2014.

[73] Wook Oh Seung, A. Harris Julie, Ng Lydia, Winslow Brent, Cain Nicholas, Mihalas Stefan, Wang Quanxin, Lau Chris, Kuan Leonard, M. Henry Alex, T. Mortrud Marty, Ouellette Benjamin, and Nghi Nguyen Thuc. A mesoscale connectome of the mouse brain. *NATURE*, 508:207–14, Apr 2014.

[74] JM Stafford, BR Jarrett, O Miranda-Dominguez, BD Mills, N Cain, S Mihalas, GP Lahvis, KM Lattal, and SH Mitchell. Large-scale topology and the default mode network in the mouse connectome. *Proc Natl Acad Sci USA*, 111(52):18745–50, Dec 2014.

[75] L Kuan, Y Li, C Lau, D Feng, A Bernard, SM Sunkin, H Zeng, C Dang, M Hawrylycz, and L Ng. Neuroinformatics of the allen mouse brain connectivity atlas. *Methods*, 73:4–17, Feb 2015.

[76] Feng David, Lau Chris, Ng Lydia, Li Yang, Kuan Leonard, M. Sunkin Susan, Dang Chinh, and Hawrylycz Michael. Exploration and visualization of connectivity in the adult mouse brain. *Methods*, 73:90–97, Feb 2015.

[77] Fakhry Ahmed, Zeng Tao, Hanchuand Peng, and Shuiwangd Ji. Global analysis of gene expression and projection target correlations in the mouse brain. *Brain Informatics*, 2(107), Mar 2015.

[78] Bopp Rita, Macarico da Costa Nuno, M. Kampa Bjorn, A. C. Martin Kevan, and M. Roth Morgane. Pyramidal cells make specific connections onto smooth (gabaergic) neurons in mouse visual cortex. *PLOS*, Aug 2014.

[79] L Glickfeld Lindsey, Clay Reid R, and L Andermann Mark. A mouse model of higher visual cortical function. *Current Opinion in Neurobiology*, 24:28–33, Feb 2014.

[80] GX Wang, SJ Smith, and P Mourrain. Fmr1 ko and fenobam treatment differentially impact distinct synapse populations of mouse neocortex. *Neuron*, 84(6):1273–86, Dec 2014.

[81] Panser Karin, Tirian Laszlo, Schulze Florian, Villalba Santiago, S.X.E. Jefferis Gregory, Buhler Katja, and D. Straw Andrew. Automatic segmentation of drosophila neural compartments using gal4 expression data reveals novel visual pathways. *Curr Biol.*, 26(15):1943–54, Aug 2016.

[82] Y Kim, KU Venkataraju, K Pradhan, C Mende, J Taranda, SC Turaga, I Arganda-Carreras, L Ng, MJ Hawrylycz, KS Rockland, HS Seung, and P Osten. Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Rep.*, 10(2):292–305, Jan 2015.

[83] Grange Pascal, Menashe Idan, and Hawrylycz Michael. Cell-type-specific neuroanatomy of cliques of autism-related genes in the mouse brain. *Front Comput Neurosci.*, 9(55), May 2015.

[84] LA Quina, L Tempest, L Ng, JA Harris, S Ferguson, TC Jhou, and EE Turner. Efferent pathways of the mouse lateral habenula. *J Comp Neurol*, 523(1):32–60, Jan 2005.

[85] AS Shai, CA Anastassiou, ME Larkum, and C Koch. Physiology of layer 5 pyramidal neurons in mouse primary visual cortex: Coincidence detection through bursting. *PLoS Comput Biol.*, 11(3), Mar 2015.

[86] RM Empson, C Goulton, D Scholtz, Y Gallero-Salas, and Hl Zeng. Validation of optical voltage reporting by the genetically encoded voltage indicator vsfp-butterfly from cortical layer 2/3 pyramidal neurons in mouse brain slices. *Physiol Rep.*, 3(7), Jul 2015.

[87] Ji W, R Gamanut, P Bista, RD D'Souza, Q Wang, and A Burkhalter. Modularity in the organization of mouse primary visual cortex. *Neuron*, 87(3):632–43, Aug 2015.

[88] DJ Denman and D Contreras. On parallel streams through the mouse dorsal lateral geniculate nucleus. *Front Neural Circuits.*, 10(20), Mar 2016.

[89] Fakhry Ahmed and Ji Shuiwang. High-resolution prediction of mouse brain connectivity using gene expression patterns. *Methods*, 73:71–78, Feb 2015.

[90] Shuiwang Ji. Computational genetic neuroanatomy of the developing mouse brain: dimensionality reduction, visualization, and clustering. *BMC Bioinformatics*, 14(222), 2013.

[91] Shuiwang Ji. Computational network analysis of the anatomical and genetic organizations in the mouse brain. *BMC Bioinformatics*, 27(23):3293–9, Dec 2011.

[92] S Ji, A Fakhry, and H Deng. Integrative analysis of the connectivity and gene expression atlases in the mouse brain. *Neuroimage*, 84:245–53, Jan 2014.

[93] YANG TAO, ZHAO XINLIN, LIN BINBIN, ZENG TAO, JI SHUIWANG, and Y JIEPING. Automated gene expression pattern annotation in the mouse brain. *Engineering in Medicine and Biology Society*, 2004.

[94] Bergqvist Goran and G. Larsson Erik. Higher-order singular value decomposition: Theory and an application. *IEEE SIGNAL PROCESSING MAGAZINE*, 27(3):151–54, 2010.

[95] L.D Lathauwer, B.D Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J.Matrix Anal. Appl.*, 21:1253–78, 2000.

[96] B Pascual, JC Masdeu, M Hollenbeck, and N Makris. Large-scale brain networks of the human left temporal pole: a functional connectivity mri study. *Cereb Cortex*, 25(3):680–702, Mar 2015.

[97] JA Miller, SL Ding, SM Sunkin, KA Smith, L Ng, A Szafer, A Ebbert, ZL Riley, JJ Royall, and K Aiona. Transcriptional landscape of the prenatal human brain. *Nature*, 508(7495):199–206, Apr 2014.

[98] K Amunts, MJ Hawrylycz, Essen DC Van, Horn JD Van, N Harel, and JB Poline. Interoperable atlases of the human brain. *Neuroimage*, 1(99):525–32, Oct 2014.

[99] Ouyang A. Spatial mapping of structural and connectional imaging data for the developing human brain with diffusion tensor imaging. *Methods*, 73:27–37, Feb 2015.

[100] Bakken TE, Miller JA, Luo R, Bernard A, Bennett JL, Lee CK, Bertagnolli D, Parikshak NN, Smith KA, Sunkin SM, Amaral DG, Geschwind DH, and Lein ES. Spatiotemporal dynamics of the postnatal developing primate brain transcriptome. *Hum Mol Genet*, 24(15):4327–39, Aug 2015.

[101] J Richiardi, A Altmann, AC Milazzo, C Chang, MM Chakravarty, T Banaschewski, GJ Barker, and AL Bokde. Brain networks. correlated gene expression supports synchronous activity in brain networks. *Science*, 348(6240):1241–4, Jun 2015.

[102] SL Ding, JJ Royall, SM Sunkin, L Ng, BA Facer, P Lesnar, A Guillozet-Bongaarts, and B McMurray. Comprehensive cellular-resolution atlas of the adult human brain. *J Comp Neurol.*, 524(16):3127–481, Nov 2016.

[103] L Xie, JB Pluta, SR Das, LE Wisse, H Wang, and L Mancuso. Multi-template analysis of human perirhinal cortex in brain mri: Explicitly accounting for anatomical variability. *Neuroimage*, Oct 2016.

[104] MS Goyal, M Hawrylycz, JA Miller, AZ Snyder, and ME Raichle. Aerobic glycolysis in the human brain is associated with development and neotenous gene expression. *Cell Metab*, 19(1):49–57, Jan 2014.

[105] de Leemput J Van, NC Boles, TR Kiehl, B Corneo, P Lederman, V Menon, and C Lee. Cortecon: A temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. *Neuron*, 83(1):51–68, Jul 2014.

[106] PA Yushkevich, JB Pluta, H Wang, L Xie, SL Ding, EC Gertje, and L Mancuso. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum Brain Mapp.*, 36(1):258–87, Jan 2015.

[107] Quiroga R Quian, A Kraskov, F Mormann, I Fried, and C Koch. Single-cell responses to face adaptation in the human medial temporal lobe. *Neuron*, 84(2):363–9, Oct 2014.

[108] SL Ding and Hoesen GW Van. Organization and detailed parcellation of human hippocampal head and body regions based on a combined analysis of cyto- and chemoarchitecture. *J Comp Neurol*, 523(15):2233–53, Oct 2015.

[109] AM Pasca, SA Sloan, LE Clarke, and Y Tian. Functional cortical neurons and astrocytes from human pluripotent stem cells in 3d culture. *Nat Methods*, 12(7):671–8, Jul 2015.

[110] Song-Lin Ding. Detailed segmentation of human hippocampal and subicular subfields using a combined approach. *Neuroscience Communications*, 1, 2015.

[111] Song-Lin Ding. Comparative anatomy of the prosubiculum, subiculum, presubiculum, postsubiculum, and parasubiculum in human, monkey, and rodent. *J Comp Neurol*, 521(18):4145–62, Dec 2013.

[112] AR Pfenning, E Hara, O Whitney, MV Rivas, R Wang, PL Roulhac, JT Howard, and M Wirthlin. Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, 346(6215), Dec 2014.

[113] Mohan K. Namboodiri Vijay, M. Levy Joshua, Mihalas Stefan, W. Sims David, and G. Hussain Shuler Marshall. Rationalizing spatial exploration patterns of wild animals and humans through a temporal discounting framework. *PNAS*, 113(31):8747–52.

[114] AJ Willsey, SJ Sanders, M Li, S Dong, AT Tebbenkamp, RA Muhle, SK Reilly, L Lin, and S Fertuzinhos. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*, 155(5):997–1007, Nov 2013.

[115] RA Martinez, JL Stein, AR Krostag, AM Nelson, JS Marken, V Menon, and RC May. Genome engineering of isogenic human es cells to model autism disorders. *Nucleic Acids Res*, 43(10):e65, may 2015.

[116] M. Ursa Nikhil, Bidob Simone, M. Petersona Sean, L. Daiglea Tanya, E. Bassd Caroline, R. Gainetdinov Raul, Bezard Erwan, and G. Caron Marc. Targeting beta-arrestin2 in the treatment of l-dopa-induced dyskinesia in parkinson's disease. *Proc Natl Acad Sci USA*, 112(19):e2517–26, May 2015.

[117] KD Dave, S De Silva, NP Sheth, S Ramboz, MJ Beck, C Quang, and RC 3rd Switzer. Phenotypic characterization of recessive gene knockout rat models of parkinson's disease. *Neurobiol Dis*, 70:190–203, Oct 2014.

[118] Grange Pascal, Hawrylycz Michael, and P. Mitra Partha. Computational neuroanatomy and co-expression of genes in the adult mouse brain, analysis tools for the allen brain atlas. *Quantitative Biology*, 1(1):91–100, Mar 2013.

[119] Allen atlas brain toolbox. `http://www.brainarchitecture.org/allen-atlas-brain-toolbox/`.

[120] Ng Lydia, Lau Chris, Young Rob, Pathak Sayan, Kuan Leonard, Sodt Andrew, Sutram Madhavi, Lee Chang-Kyu, Dang Chinh, and Hawrylycz Michael. Neuroblast: a 3d spatial homology search tool for gene expression. *BMC Neuroscience*, 8(suppl 2):11, Jul 2007.

[121] Brain explorer 2. `http://mouse.brain-map.org/static/brainexplorer`.

[122] Ng Lydia, Bernard Amy, Lau Chris, C Overly Caroline, Dong Hong-Wei, and Kuan Chihchau. An anatomic gene expression atlas of the adult mouse brain. *Nature Neuroscience*, 12:356–62, 2009.

[123] FP Davis and SR Eddy. A tool for identication of genes expressed in patterns of interest using the allen brain atlas. *Bioinformatics*, 25(13):1647–54, 2009.

[124] Noa Liscovitch and Gal Chechik. Specialization of gene expression during mouse brain development. *PLoS Comput Biol*, 9(9), Sept 2013.

[125] Cornelia M. Hooper, Susan M. Hawes, Ursula R. Kees, Nicholas G. Gottardo, and Peter B. Dallas. Gene expression analyses of the spatio-temporal relationships of human medulloblastoma subgroups during early human neurogenesis. *PLOS One*, 9(11), 2014.

[126] Guang-Hui Liu, Cheng-Zhou Mao, Hai-Yan Wu, Deng-Cheng Zhou, Jing-Bo Xia, Soo-Ki Kim, Dong-Qing Cai, Hui Zhao, and Xu-Feng Qi. Expression profile of

rrbp1 genes during embryonic development and in adult tissues of xenopus laevis. *Gene Expression Patterns*, 23-24:1–6, 2017.

[127] Saadettin Sel, Eva Patzel, Lucia Poggi, Delia Kaiser, Thomas Kalinski, Martin Schicht, Friedrich Paulsen, and Norbert Nass. Temporal and spatial expression pattern of nnat during mouse eye development. *Gene Expression Patterns*, 23-24:7–12, 2017.

[128] Qihang Li, Huang Kun, and Machiraju Raghu. Spatiotemporal visualization of gene expression in the developing mouse brain. *Eurographics Conference on Visualization (EuroVis 2017)*, 2017. (forthcoming).

[129] Pavel Tomancak, Benjamin P Berman, Amy Beaton, Richard Weiszmann, Elaine Kwan, Volker Hartenstein, Susan E Celniker, and Gerald M Rubin. Global analysis of patterns of gene expression during drosophilaembryogenesis. *Genome Biology*, 8(7), 2007.

[130] Ying Tao, Yang Liu, Carol Friedman, and Yves A. Lussier. Information visualization techniques in bioinformatics during the postgenomic era. *Drug Discov Today Biosilico.*, 2(6):237–45, Nov. 2004.

[131] Colorbrewer 2.0. `http://colorbrewer2.org/`.

[132] Oghabian A, Kilpinen S, Hautaniemi S, and Czeizler E. Biclustering methods: Biological relevance and application in gene expression analysis. *PLoS ONE*, 9(3).

[133] Kemal Eren, Mehmet Deveci, Onur Kuuktun, and Umit V. Catalyurek. A comparative analysis of biclustering algorithms for gene expression data. *Brief Bioinform.*, 14(3):279–92, 2012.

[134] Irene Gallego Romero, Ilya Ruvinsky, and Yoav Gilad. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, 13:505–516, 2012.

[135] J. Basser, Peter. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR Biomed*, 8(7-8):333–44, 1995.

[136] Evren Ozarslan, Baba C. Vemuri, and Thomas H. Mareci. Generalized scalar measures for diffusion mri using trace, variance, and entropy. *Magnetic Resonance in Medicine*, 53(4):1522–2594, 2005.

[137] E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. A user study of visualization effectiveness using eeg and cognitive load. *Computer Graphics Forum*, 30(3):791–800, 2011.

[138] A. Seriai, O. Benomar, B. Cerat, and H. Sahraoui. Validation of software visualization tools: A systematic mapping study. pages 60–69, Sept 2014.

Appendix A: Supplementary Material for the Developing Mouse Brain Dataset

Cluster	Term	Count	%	P-Value
1	nervous system development	22	36.1	2.60E-02
	neural tube development	5	8.2	4.00E-02
2	developmental process	150	64.1	2.00E-05
	multicellular organismal development	146	62.4	2.30E-05
	generation of neurons	58	24.8	5.50E-05
	neurogenesis	60	25.6	5.90E-05
	cell motion	40	17.1	1.00E-04
	multicellular organismal process	157	67.1	1.40E-04
	cell morphogenesis involved in differentiation	33	14.1	1.60E-04
	neuron differentiation	48	20.5	2.20E-04
	nervous system development	78	33.3	2.30E-04
	organ development	111	47.4	2.60E-04
	system development	129	55.1	2.70E-04
	cell differentiation	100	42.7	3.00E-04
	axon guidance	22	9.4	3.50E-04
	anatomical structure development	131	56	4.50E-04
	cellular developmental process	100	42.7	5.90E-04
	cell morphogenesis	35	15	6.80E-04
	cell surface receptor linked signal transduction	81	34.6	8.10E-04
	cellular component morphogenesis	35	15	1.00E-03

Table A.1: Biological enrichment analysis of the 1^{st} and 2^{nd} gene clusters in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
3	metabolic process	66	65.3	1.50E-03
	macromolecule metabolic process	60	59.4	1.70E-03
	primary metabolic process	63	62.4	2.30E-03
	cellular macromolecule metabolic process	58	57.4	2.40E-03
	cellular metabolic process	62	61.4	2.90E-03
	biosynthetic process	47	46.5	2.10E-02
	gene expression	44	43.6	2.50E-02
	nitrogen compound metabolic process	46	45.5	2.80E-02
	peptidyl-amino acid modification	5	5	2.90E-02
	cell proliferation	10	9.9	3.00E-02
	cellular macromolecule biosynthetic process	43	42.6	3.20E-02
	cellular biosynthetic process	46	45.5	3.20E-02
	regulation of macromolecule metabolic process	54	53.5	3.30E-02
	macromolecule biosynthetic process	43	42.6	3.30E-02
	regulation of gene expression	52	51.5	3.50E-02
	tissue homeostasis	4	4	3.50E-02
	cellular nitrogen compound metabolic process	45	44.6	3.70E-02
	reproductive developmental process	11	10.9	4.00E-02
	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	43	42.6	4.10E-02
	chromatin remodeling	3	3	4.30E-02
	regulation of cellular metabolic process	55	54.5	4.40E-02

Table A.2: Biological enrichment analysis of the 3^{rd} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	cellular biosynthetic process	261	39.1	9.90E-03
	biosynthetic process	262	39.3	1.00E-02
	metabolic process	358	53.7	1.20E-02
	cellular metabolic process	336	50.4	1.60E-02
	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	243	36.4	1.70E-02
	organophosphate metabolic process	8	1.2	1.90E-02
	phospholipid metabolic process	8	1.2	1.90E-02
	cellular nitrogen compound metabolic process	254	38.1	1.90E-02
	cellular macromolecule biosynthetic process	239	35.8	1.90E-02
	transcription	232	34.8	2.20E-02
	macromolecule biosynthetic process	239	35.8	2.20E-02
	nitrogen compound metabolic process	256	38.4	2.20E-02
	gene expression	242	36.3	2.20E-02
	primary metabolic process	338	50.7	3.10E-02
4	negative regulation of cell proliferation	40	6	3.10E-02
	organelle organization	43	6.4	3.80E-02
	regulation of cell-substrate adhesion	11	1.6	4.10E-02
	regulation of gene expression	295	44.2	4.50E-02
	transport	108	16.2	4.60E-02
	cell cycle phase	16	2.4	4.80E-02
	establishment of localization	109	16.3	4.90E-02
	cellular process	513	76.9	5.50E-02
	cellular macromolecule metabolic process	302	45.3	5.50E-02
	positive regulation of RNA metabolic process	99	14.8	6.10E-02
	positive regulation of transcription, DNA-dependent	99	14.8	6.10E-02
	regulation of transcription	285	42.7	6.30E-02
	synapse organization	11	1.6	6.70E-02
	regulation of macromolecule biosynthetic process	287	43	6.90E-02

Table A.3: Biological enrichment analysis of the 4^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	embryonic pattern specification	6	10.9	4.70E-04
	anatomical structure formation involved in morphogenesis	12	21.8	9.00E-03
	taxis	4	7.3	1.30E-02
	chemotaxis	4	7.3	1.30E-02
	BMP signaling pathway	4	7.3	1.30E-02
	fibroblast growth factor receptor signaling pathway	4	7.3	1.80E-02
	negative regulation of cell migration	3	5.5	2.30E-02
	negative regulation of cell motion	3	5.5	2.30E-02
	positive regulation of osteoblast differentiation	3	5.5	2.80E-02
	enzyme linked receptor protein signaling pathway	10	18.2	2.90E-02
	positive regulation of developmental process	8	14.5	3.20E-02
	developmental process	35	63.6	3.20E-02
	negative regulation of locomotion	3	5.5	3.50E-02
5	angiogenesis	6	10.9	3.50E-02
	cell fate commitment	8	14.5	3.60E-02
	regulation of ossification	4	7.3	3.80E-02
	induction of an organ	3	5.5	4.20E-02
	transmembrane receptor protein serine/threonine kinase signaling pathway	5	9.1	4.20E-02
	positive regulation of cell differentiation	7	12.7	4.30E-02
	blood vessel morphogenesis	7	12.7	4.50E-02
	respiratory system development	6	10.9	4.70E-02
	anatomical structure morphogenesis	22	40	4.90E-02
	vasculature development	8	14.5	5.30E-02
	transforming growth factor beta receptor signaling pathway	4	7.3	5.60E-02
	embryonic development	14	25.5	6.30E-02
	multicellular organismal development	33	60	6.50E-02
	regulation of protein metabolic process	5	9.1	6.60E-02

Table A.4: Biological enrichment analysis of the 5^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
6	regulation of cytokine biosynthetic process	5	3.4	9.30E-03
	proteolysis involved in cellular protein catabolic process	6	4.1	1.30E-02
	cellular protein catabolic process	6	4.1	1.30E-02
	regulation of primary metabolic process	77	52.7	1.50E-02
	regulation of cellular biosynthetic process	74	50.7	1.70E-02
	regulation of biosynthetic process	74	50.7	1.70E-02
	regulation of macromolecule biosynthetic process	72	49.3	1.80E-02
	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	73	50	1.80E-02
	regulation of cytokine production	6	4.1	2.00E-02
	regulation of transcription	71	48.6	2.10E-02
	regulation of nitrogen compound metabolic process	73	50	2.10E-02
	cellular macromolecule catabolic process	6	4.1	2.50E-02
	regulation of macromolecule metabolic process	75	51.4	2.60E-02
	aging	5	3.4	3.00E-02
	regulation of metabolic process	79	54.1	3.00E-02
	protein catabolic process	6	4.1	3.00E-02
	regulation of cellular metabolic process	77	52.7	3.00E-02
	regulation of gene expression	72	49.3	3.00E-02
	tube development	19	13	3.60E-02
	regulation of developmental process	28	19.2	4.80E-02
	macromolecule catabolic process	6	4.1	5.70E-02
	positive regulation of cytokine biosynthetic process	3	2.1	5.80E-02
	positive regulation of DNA metabolic process	3	2.1	5.80E-02
	negative regulation of cellular metabolic process	22	15.1	5.80E-02
	cellular catabolic process	8	5.5	5.90E-02
	negative regulation of macromolecule biosynthetic process	21	14.4	6.00E-02
	negative regulation of macromolecule metabolic process	22	15.1	6.10E-02
	mesoderm development	7	4.8	6.30E-02
	tissue morphogenesis	17	11.6	6.30E-02

Table A.5: Biological enrichment analysis of the 6^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
7	regulation of epithelial cell proliferation involved in prostate gland development	3	2.9	1.80E-02
	regulation of signal transduction	14	13.3	4.00E-02
	G-protein coupled receptor protein signaling pathway	13	12.4	7.30E-02
	regulation of cell communication	16	15.2	7.50E-02
	hair follicle maturation	3	2.9	8.80E-02
	hair cycle	4	3.8	9.90E-02
	hair cycle process	4	3.8	9.90E-02
	molting cycle process	4	3.8	9.90E-02
	molting cycle	4	3.8	9.90E-02
	hair follicle development	4	3.8	9.90E-02

Table A.6: Biological enrichment analysis of the 7^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	prostate epithelial cord arborization involved in prostate glandular acinus morphogenesis	3	7.1	5.80E-03
	prostate glandular acinus morphogenesis	3	7.1	5.80E-03
	regulation of cell proliferation	12	28.6	6.70E-03
	urogenital system development	7	16.7	1.20E-02
	gland morphogenesis	6	14.3	1.30E-02
	prostate gland epithelium morphogenesis	4	9.5	1.40E-02
	prostate gland morphogenesis	4	9.5	1.40E-02
	gland development	8	19	1.50E-02
	branching involved in prostate gland morphogenesis	3	7.1	1.50E-02
	regulation of cellular process	36	85.7	1.70E-02
	prostate glandular acinus development	3	7.1	2.00E-02
	system development	27	64.3	2.20E-02
8	osteoblast differentiation	4	9.5	2.30E-02
	forebrain development	7	16.7	2.50E-02
	mammary gland morphogenesis	4	9.5	2.50E-02
	tube morphogenesis	7	16.7	2.60E-02
	mammary gland development	5	11.9	2.70E-02
	prostate gland development	4	9.5	3.10E-02
	morphogenesis of a branching structure	6	14.3	3.20E-02
	anatomical structure development	27	64.3	3.20E-02
	smoothened signaling pathway	3	7.1	3.40E-02
	regulation of biological process	36	85.7	3.40E-02
	cell differentiation	21	50	3.50E-02
	lung morphogenesis	3	7.1	4.00E-02
	cellular developmental process	21	50	4.10E-02
	branching morphogenesis of a tube	5	11.9	4.50E-02
	developmental process	29	69	4.60E-02

Table A.7: Biological enrichment analysis of the 8^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	DNA metabolic process	6	6.1	9.00E-03
	regulation of transcription from RNA polymerase II promoter	28	28.3	1.10E-02
	negative regulation of gene-specific transcription	4	4	1.50E-02
	negative regulation of specific transcription from RNA polymerase II promoter	4	4	1.50E-02
	response to extracellular stimulus	7	7.1	1.70E-02
	gene expression	43	43.4	2.30E-02
	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	43	43.4	2.30E-02
	macromolecule metabolic process	54	54.5	2.40E-02
9	regulation of specific transcription from RNA polymerase II promoter	6	6.1	2.60E-02
	transcription	41	41.4	2.90E-02
	cellular macromolecule biosynthetic process	42	42.4	2.90E-02
	protein complex assembly	6	6.1	3.00E-02
	protein complex biogenesis	6	6.1	3.00E-02
	macromolecule biosynthetic process	42	42.4	3.00E-02
	macromolecular complex assembly	6	6.1	3.40E-02
	nitrogen compound metabolic process	44	44.4	4.00E-02
	lymphocyte differentiation	6	6.1	4.70E-02
	T cell differentiation	5	5.1	4.70E-02
	regulation of gene-specific transcription	6	6.1	5.20E-02
	cellular nitrogen compound metabolic process	43	43.4	5.20E-02
	macromolecular complex subunit organization	6	6.1	5.70E-02

Table A.8: Biological enrichment analysis of the 9^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
10	regulation of biological quality	49	23.3	4.60E-04
	homeostatic process	30	14.3	1.40E-03
	membrane depolarization	9	4.3	1.80E-03
	regulation of system process	19	9	2.90E-03
	synaptic transmission, glutamatergic	6	2.9	3.50E-03
	regulation of excitatory postsynaptic membrane potential	7	3.3	3.50E-03
	chemical homeostasis	23	11	5.00E-03
	cellular homeostasis	20	9.5	5.00E-03
	cellular chemical homeostasis	19	9	5.20E-03
	ion homeostasis	19	9	5.20E-03
	regulation of acetylcholine secretion	4	1.9	6.10E-03
	regulation of synaptic plasticity	8	3.8	6.10E-03
	regulation of synaptic transmission	12	5.7	6.50E-03
	regulation of neurological system process	13	6.2	6.90E-03
	regulation of amine transport	7	3.3	7.60E-03
	nerve-nerve synaptic transmission	7	3.3	7.60E-03
	ion transport	23	11	7.90E-03
	establishment of localization	43	20.5	8.90E-03
	cellular ion homeostasis	18	8.6	9.20E-03
	regulation of body fluid levels	5	2.4	9.30E-03
	regulation of transmission of nerve impulse	12	5.7	9.50E-03
	regulation of postsynaptic membrane potential	7	3.3	1.10E-02
	transmission of nerve impulse	20	9.5	1.20E-02
	transport	42	20	1.20E-02
	synaptic transmission	17	8.1	1.20E-02
	regulation of membrane potential	13	6.2	1.40E-02
	localization	55	26.2	1.40E-02

Table A.9: Biological enrichment analysis of the 10^{th} gene cluster in the BGEFM (Figure 4.5) [5]. Used entire 1753 genes as background. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Neurodevelopment Events	Equivalent PC days
Cranial motor nuclei peak	9
Inferior olive generation onset	9
Locus coeruleus onset	9
Locus coeruleus peak	9
Occulomotor nucleus cell generation onset	9
Trigeminal ganglion cell generation onset	9
Trigeminal mesencephalon nucleus offset	9
Trigeminal mesencephalon nucleus onset	9
Trigeminal motor nucleus peak	9
Medial longitudinal fasciculus appears	9.5
Post-proliferative zone in the medulla appears (caudal to the trigeminal nerve)	9.8
Entorhinal cortex neurogenesis onset	10
Granule cell layer fascia dentata neurogenesis onset	10
Granule cells in the dentate gyrus onset	10
Hippocampal CA1 neurogenesis onset	10
Hippocampal CA3 neurogenesis onset	10
Inferior olivary nucleus peak	10
Locus coeruleus offset	10
Medial geniculate nucleus onset	10
Olfactory tubercle generation onset	10
Photoreceptor generation onset	10
Purkinje cell generation onset	10
Septal nuclei onset	10
Subicular cortex neurogenesis onset	10
Subplate onset of neurogenesis	10
Trigeminal motor nucleus offset	10
dLGN start	10.5
Mitral cells onset	10.5
Post-proliferative zone appears in the tegmentum	10.5
Purkinje cells peak	10.5
Retinal ganglion cell generation start	10.5
Superficial SC laminae start	10.5
Ganglionic eminence post-proliferative zone appears	10.8
Cranial sensory nuclei peak	11
Globus pallidus peak	11
Inferior olive generation offset	11
Medial geniculate nucleus peak	11
Neurogenesis cortical layer VI start (VC)	11
Reticular nuclei peak	11
Subplate peak	11
Superior colliculus generation onset	11
Post-proliferative zone appears in the bulging thalamus	11.5
Post-proliferative zone appears in the cerebellum	11.5
Post-proliferative zone appears in the pre-tectum	11.5
vLGN peak	11.5
Post-proliferative/white matter appears in the hypothalamus	11.6

Table A.10: List of the neurodevelopment events of mouse brain [6]. Sorted in the increasing PC days order.

Neurodevelopment Events	Equivalent PC days
Post-proliferative zone appears in the medial pallium	11.8
Amygdala peak	12
Anteroventral cochlear nucleus peak	12
Caudoputamen generation onset	12
Cochlear nuclei peak	12
Cortical plate first observed/visible, first cortical plate neurons born	12
dLGN peak	12
GABA cells in subplate start	12
GABAir cells in lower intermediate zone/subventricular border (no synapses)	12
Mitral cells peak	12
Neurogenesis cortical layer V start (VC)	12
Post-proliferative zone appears in the superior colliculus	12
Purkinje cell generation offset	12
Trigeminal ganglion cell generation offset	12
Axons in optic stalk	12.3
Claustrum peak	12.5
dLGN end	12.5
Neurogenesis cortical layer VI peak (VC)	12.5
Nucleus of lateral olfactory tract peak	12.5
Preoptic nucleus peak	12.5
VP (Ventroposterior) and VB (ventrobasal complex) nuclei peak (thalamus)	12.5
Entorhinal cortex peak	13
GABA cells in subplate end	13
Medial forebrain bundle appears	13
Medial geniculate nucleus offset	13
Neurogenesis cortical lamina VI end (VC)	13
Neurogenesis cortical layer V peak (VC)	13
Optic axons at chiasm of optic tract	13
Primary somatosensory cortex layer 2/3 offset	13
Retinal ganglion cells peak	13
Rod generation onset	13
Septal nuclei peak (neuron)	13
Subiculum peak (hippocampus)	13
Superior colliculus peak	13
Suprachisamatic nucleus peak (hypothalamus)	13
Anterior olfactory nucleus peak (smell func)	13.5
AV, AM, AD nuclei	13.5
Cortical subventricular zone (abventricular cells) onset	13.5
Lhx6 first in cortex in GABAergic cells	13.5
Olfactory tubercle generation peak	13.5
Parasubiculum peak	13.5
Pontine nuclei peak	13.5
Raphe complex peak	13.5
Stria terminalis appears	13.5

Table A.11: List of the neurodevelopment events of mouse brain (continued) [6].

Neurodevelopment Events	Equivalent PC days
Caudoputamen peak	14
Fasciculus retroflexus appears	14
Fornix appears	14
Horizontal cell generation offset	14
LGN axons in subplate	14
Neurogenesis cortical layer V end (VC)	14
Primary somatosensory cortex layer IV offset	14
Primary somatosensory cortex layer V offset	14
Superficial SC laminae end	14
Anterior commisure appears	14.5
Optic axons reach dLGN and SC	14.5
CA 1, CA 2 peak	15
Hippocampal commisure appears	15
Neurogenesis cortical lamina IV start (VC)	15
Neurogenesis cortical layer II/III peak (VC)	15
Retinal amacrine cells peak	15
Subicular cortex neurogenesis offset	15
Superior colliculus generation offset	15
Mitral cells offset	15.5
Cortical axons reach thalamus	16
Entorhinal cortex neurogenesis offset	16
Nucleus accumbens peak	16
Onset of retinal waves stage 1	16
Primary somatosensory cortex layer VI offset	16
Septal nuclei offset	16
Tufted cells peak (olfactory bulb)	16
Corpus callosum appears	17
Hippocampal CA1 neurogenesis offset	17
Hippocampal CA3 neurogenesis offset	17
Neurogenesis cortical layer IV end (VC)	17
Neurogenesis cortical layer IV peak (VC)	17
End of retinal waves stage 1	18.5
Onset of retinal waves stage 2	18.5
Retinal ganglion cell generation end	18.5
Initial differentiation of layer V (S1)	19
Rods peak (neuron)	19
Onset of decrease in fractional anisotropy	20
Surface righting onset	21.5
Onset of barrels (S1)	22
Onset of sublayers in layer V (S1)	22
Onset of trilaminar plate (S1)	22
Rod generation offset	23.5
Onset of barrel field septa (S1)	24
Rooting reflex offset	24.5
Ipsi/contra segregation in LGN	25.5
Ipsi/contra segregation in SC	25.5
Inferior cerebellar peduncle myelination onset	27

Table A.12: List of the neurodevelopment events of mouse brain (continued) [6]. Sorted in the increasing PC days order.

Neurodevelopment Events	Equivalent PC days
Lateral lemniscus myelination onset	27
Medial lemniscus myelination onset	27
Middle cerebellar peduncle myelination onset	27
Corticospinal projections reach lumbar levels	27.5
Internal capsule myelination onset	28
Optic tract myelination onset	28
Posterior commisure myelination onset	28
Superior cerebellar peduncle myelination onset	28
Fasciculus retroflexus myelination onset	29
End of retinal waves stage 2	29.5
Onset of retinal waves stage 3	29.5
Postconception walking onset	29.5
Cingulum myelination onset	30
Eye opening	30
Preyer reflex	30
Stria medullaris myelination onset	30
Striatum myelination onset	30
air righting reflex	31
Brachium inferior colliculus myelination onset	31
Ears open (auditory canals fully open)	31
Olfactory tract myelination onset	31
Vibrissa placing adult like pattern	31
Anterior commisure myelination onset	32
Fornix myelination onset	32
Hippocampus myelination onset	32
Mammillothalamic tract myelination onset	32
End of retinal waves stage 3	32.5
Corpus callosum body myelination onset	33
Lenticular fasciculus myelination onset	33
Splenium myelination onset	33
Plasticity/OD critical period start	37.5
Visual placing is mature	37.5
Plasticity/OD critical period end	50.5
Corpus callosum body myelination end	62
Middle cerebellar peduncle myelination end	78

Table A.13: List of the neurodevelopment events of mouse brain (continued) [6]. Sorted in the increasing PC days order.

Appendix B: Supplementary Material for the Developing Human Brain Dataset

Cluster	Term	Count	%	P-Value
	chemical synaptic transmission	44	17.3	6.20E-18
	anterograde trans-synaptic signaling	44	17.3	6.20E-18
	synaptic signaling	44	17.3	6.20E-18
	trans-synaptic signaling	44	17.3	6.20E-18
	inorganic cation transmembrane transport	37	14.6	2.20E-12
	regulation of transport	66	26	5.30E-12
	cation transmembrane transport	38	15	1.00E-11
	metal ion transport	41	16.1	1.60E-11
	inorganic ion transmembrane transport	38	15	2.30E-11
	ion transport	56	22	4.10E-11
	single-organism localization	98	38.6	5.60E-11
	single-organism transport	94	37	8.90E-11
1	ion transmembrane transport	43	16.9	1.30E-10
	cation transport	43	16.9	2.60E-10
	regulation of localization	76	29.9	2.60E-10
	transport	113	44.5	8.80E-10
	cell-cell signaling	54	21.3	1.80E-09
	establishment of localization	114	44.9	2.10E-09
	transmembrane transport	49	19.3	2.30E-09
	single-organism process	227	89.4	6.70E-09
	single-organism cellular process	215	84.6	8.60E-09
	modulation of synaptic transmission	21	8.3	9.80E-09
	localization	129	50.8	1.00E-08
	nervous system development	66	26	1.30E-08
	regulation of ion transmembrane transport	24	9.4	1.90E-08

Table B.1: Biological enrichment analysis of the 1^{st} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	purine ribonucleoside metabolic process	11	20.8	1.60E-08
	purine nucleoside metabolic process	11	20.8	1.80E-08
	purine ribonucleoside triphosphate metabolic process	10	18.9	1.80E-08
	ribonucleoside triphosphate metabolic process	10	18.9	2.20E-08
	purine nucleoside triphosphate metabolic process	10	18.9	2.30E-08
	ribonucleoside metabolic process	11	20.8	3.40E-08
	mitochondrion organization	14	26.4	3.80E-08
	nucleoside triphosphate metabolic process	10	18.9	4.70E-08
	nucleoside metabolic process	11	20.8	6.60E-08
	glycosyl compound metabolic process	11	20.8	1.10E-07
	ATP metabolic process	9	17	1.50E-07
2	purine ribonucleoside monophosphate metabolic process	9	17	4.90E-07
	purine nucleoside monophosphate metabolic process	9	17	5.00E-07
	ribonucleoside monophosphate metabolic process	9	17	6.90E-07
	nucleoside monophosphate metabolic process	9	17	9.50E-07
	purine-containing compound metabolic process	11	20.8	1.60E-06
	nucleobase-containing small molecule metabolic process	12	22.6	1.80E-06
	purine ribonucleotide metabolic process	10	18.9	5.70E-06
	nucleotide metabolic process	11	20.8	6.10E-06
	mitochondrial transport	8	15.1	6.40E-06
	nucleoside phosphate metabolic process	11	20.8	6.80E-06
	ribonucleotide metabolic process	10	18.9	7.20E-06
	oxidative phosphorylation	6	11.3	7.60E-06

Table B.2: Biological enrichment analysis of the 2^{nd} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	nervous system development	72	38.3	5.90E-20
	neurogenesis	56	29.8	8.30E-18
	neuron differentiation	50	26.6	1.60E-16
	generation of neurons	52	27.7	4.20E-16
	central nervous system development	41	21.8	9.60E-15
	multicellular organism development	100	53.2	2.80E-14
	neuron projection development	38	20.2	5.70E-14
	neuron development	41	21.8	6.80E-14
	system development	92	48.9	7.90E-14
	anatomical structure development	105	55.9	2.40E-13
	cell morphogenesis involved in differentiation	35	18.6	3.60E-13
	cell differentiation	82	43.6	4.30E-13
3	single-multicellular organism process	109	58	8.90E-13
	axon development	27	14.4	1.10E-12
	developmental process	105	55.9	1.80E-12
	cell morphogenesis involved in neuron differentiation	28	14.9	2.00E-12
	single-organism developmental process	103	54.8	2.50E-12
	cellular developmental process	84	44.7	6.00E-12
	axonogenesis	25	13.3	9.10E-12
	neuron projection morphogenesis	28	14.9	1.50E-11
	axon guidance	19	10.1	1.90E-11
	animal organ development	72	38.3	2.00E-11
	neuron projection guidance	19	10.1	2.00E-11
	brain development	31	16.5	5.00E-11

Table B.3: Biological enrichment analysis of the 3^{rd} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
4	cellular response to type I interferon	13	6.4	3.90E-11
	type I interferon signaling pathway	13	6.4	3.90E-11
	response to type I interferon	13	6.4	7.50E-11
	signaling	111	54.7	8.10E-09
	single-multicellular organism process	108	53.2	9.20E-09
	single organism signaling	110	54.2	1.10E-08
	developmental process	104	51.2	1.20E-08
	cellular response to chemical stimulus	63	31	1.50E-08
	cell communication	110	54.2	2.30E-08
	anatomical structure development	101	49.8	2.80E-08
	regulation of biological quality	75	36.9	2.80E-08
	single-organism process	180	88.7	5.00E-08
	single-organism cellular process	171	84.2	5.10E-08
	multicellular organism development	92	45.3	9.40E-08
	negative regulation of biological process	88	43.3	1.10E-07
	system development	84	41.4	1.30E-07
	single-organism developmental process	99	48.8	1.50E-07
	cellular response to organic substance	53	26.1	2.10E-07
	regulation of multicellular organismal process	60	29.6	2.20E-07
	cellular response to stimulus	116	57.1	2.40E-07
	response to external stimulus	53	26.1	2.70E-07
	cell differentiation	74	36.5	2.90E-07
	response to chemical	81	39.9	3.00E-07
	response to organic substance	63	31	3.20E-07

Table B.4: Biological enrichment analysis of the 4^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	biological regulation	32	88.9	2.20E-03
	regulation of cellular process	30	83.3	3.60E-03
	regulation of biological process	30	83.3	8.50E-03
	regulation of cellular biosynthetic process	16	44.4	1.40E-02
	negative regulation of cAMP-dependent protein kinase activity	2	5.6	1.60E-02
	regulation of biosynthetic process	16	44.4	1.60E-02
	negative regulation of transcription from RNA polymerase II promoter	6	16.7	1.80E-02
	regulation of nitrogen compound metabolic process	16	44.4	1.80E-02
	regulation of transcription, DNA-templated	14	38.9	1.90E-02
	regulation of nucleic acid-templated transcription	14	38.9	2.00E-02
	regulation of RNA biosynthetic process	14	38.9	2.10E-02
	renal system process	3	8.3	2.30E-02
5	multicellular organism development	17	47.2	2.30E-02
	regulation of macromolecule biosynthetic process	15	41.7	2.30E-02
	regulation of nucleobase-containing compound metabolic process	15	41.7	2.40E-02
	organic cyclic compound biosynthetic process	16	44.4	2.40E-02
	regulation of cellular metabolic process	19	52.8	2.40E-02
	clathrin coat assembly	2	5.6	2.60E-02
	regulation of RNA metabolic process	14	38.9	2.70E-02
	nucleic acid-templated transcription	14	38.9	2.80E-02
	regulation of cAMP-dependent protein kinase activity	2	5.6	2.80E-02
	nervous system development	10	27.8	2.90E-02
	anatomical structure development	18	50	3.00E-02
	single-multicellular organism process	19	52.8	3.10E-02
	single-organism developmental process	18	50	3.10E-02
	RNA biosynthetic process	14	38.9	3.60E-02

Table B.5: Biological enrichment analysis of the 5^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	trans-synaptic signaling	35	28.7	2.90E-22
	synaptic signaling	35	28.7	2.90E-22
	anterograde trans-synaptic signaling	35	28.7	2.90E-22
	chemical synaptic transmission	35	28.7	2.90E-22
	neurotransmitter transport	18	14.8	2.10E-14
	regulation of neurotransmitter levels	18	14.8	2.30E-14
	signal release from synapse	16	13.1	2.70E-14
	neurotransmitter secretion	16	13.1	2.70E-14
	presynaptic process involved in chemical synaptic transmission	16	13.1	5.00E-14
	modulation of synaptic transmission	20	16.4	6.80E-14
	synaptic vesicle cycle	14	11.5	9.90E-13
	behavior	24	19.7	3.60E-12
	vesicle mediated transport in synapse	13	10.7	5.00E-12
	synaptic vesicle localization	14	11.5	6.80E-12
	cell-cell signaling	37	30.3	8.40E-12
6	nervous system development	44	36.1	2.00E-11
	locomotory behavior	15	12.3	3.40E-11
	synaptic vesicle transport	13	10.7	6.40E-11
	establishment of synaptic vesicle localization	13	10.7	6.40E-11
	establishment of localization	66	54.1	7.20E-11
	transport	65	53.3	7.30E-11
	vesicle localization	16	13.1	1.70E-10
	adult behavior	13	10.7	2.00E-10
	single-organism behavior	19	15.6	2.60E-10
	regulation of synapse structure or activity	15	12.3	7.10E-10
	establishment of vesicle localization	15	12.3	9.30E-10
	single-organism localization	54	44.3	1.20E-09
	single-organism transport	52	42.6	1.70E-09
	neuron-neuron synaptic transmission	11	9	1.80E-09
	localization	71	58.2	2.60E-09
	signal release	18	14.8	4.40E-09

Table B.6: Biological enrichment analysis of the 6^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	generation of neurons	25	39.7	1.30E-11
	neuron differentiation	24	38.1	1.40E-11
	nervous system development	30	47.6	3.10E-11
	neurogenesis	25	39.7	4.80E-11
	neuron development	21	33.3	7.90E-11
	neuron projection development	19	30.2	3.70E-10
	cell development	27	42.9	6.60E-10
	cell projection organization	21	33.3	1.20E-08
	regulation of neuron differentiation	15	23.8	1.30E-08
	cell morphogenesis involved in neuron differentiation	14	22.2	1.60E-08
	regulation of neuron projection development	13	20.6	3.50E-08
	neuron projection morphogenesis	14	22.2	4.50E-08
	regulation of nervous system development	16	25.4	7.20E-08
	regulation of neurogenesis	15	23.8	1.20E-07
	regulation of cell projection organization	14	22.2	1.30E-07
	brain development	15	23.8	1.30E-07
7	head development	15	23.8	2.50E-07
	cell morphogenesis involved in differentiation	15	23.8	2.80E-07
	axon development	12	19	4.30E-07
	cell projection morphogenesis	15	23.8	5.20E-07
	central nervous system development	16	25.4	6.00E-07
	cell part morphogenesis	15	23.8	7.00E-07
	regulation of multicellular organismal development	21	33.3	1.10E-06
	regulation of cell development	15	23.8	1.60E-06
	axonogenesis	11	17.5	1.80E-06
	cell morphogenesis	17	27	4.00E-06
	cell differentiation	30	47.6	4.20E-06
	behavior	12	19	4.40E-06
	developmental growth	12	19	6.00E-06
	forebrain development	10	15.9	6.20E-06
	developmental growth involved in morphogenesis	8	12.7	7.30E-06
	cellular developmental process	31	49.2	7.70E-06

Table B.7: Biological enrichment analysis of the 7^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
8	trans-synaptic signaling	13	17.8	8.70E-06
	anterograde trans-synaptic signaling	13	17.8	8.70E-06
	chemical synaptic transmission	13	17.8	8.70E-06
	synaptic signaling	13	17.8	8.70E-06
	associative learning	6	8.2	1.80E-05
	behavior	11	15.1	1.30E-04
	single-organism process	68	93.2	1.40E-04
	single organism signaling	42	57.5	1.90E-04
	signaling	42	57.5	2.30E-04
	learning	6	8.2	3.00E-04
	learning or memory	7	9.6	4.10E-04
	cell communication	41	56.2	6.00E-04
	cognition	7	9.6	8.20E-04
	single-organism localization	28	38.4	8.90E-04
	single-organism transport	27	37	9.00E-04
	single-multicellular organism process	39	53.4	1.20E-03
	single-organism behavior	8	11	1.60E-03
	regulation of biological process	58	79.5	2.40E-03
	secretion by cell	12	16.4	2.50E-03
	biological regulation	60	82.2	2.70E-03
	phosphate-containing compound metabolic process	24	32.9	2.90E-03
	phosphorus metabolic process	24	32.9	3.00E-03
	regulation of cellular process	56	76.7	3.40E-03
	positive regulation of cellular process	32	43.8	3.50E-03
	response to stimulus	48	65.8	3.50E-03
	single-organism cellular process	62	84.9	3.80E-03
	cell-cell signaling	15	20.5	4.10E-03
	phosphorylation	19	26	4.10E-03
	positive regulation of biological process	34	46.6	4.70E-03
	regulation of molecular function	22	30.1	5.50E-03
	response to abiotic stimulus	12	16.4	5.80E-03

Table B.8: Biological enrichment analysis of the 8^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
	SRP-dependent cotranslational protein targeting to membrane	15	10.9	2.00E-14
	cotranslational protein targeting to membrane	15	10.9	5.40E-14
	protein localization to endoplasmic reticulum	16	11.7	6.00E-14
	protein targeting to ER	15	10.9	6.20E-14
	establishment of protein localization to endoplasmic reticulum	15	10.9	1.10E-13
	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	15	10.9	6.50E-13
	translational initiation	17	12.4	1.50E-12
	viral transcription	15	10.9	1.30E-10
	protein targeting to membrane	15	10.9	2.50E-10
	rRNA processing	17	12.4	2.60E-10
	viral gene expression	15	10.9	2.90E-10
	rRNA metabolic process	17	12.4	3.50E-10
	nuclear-transcribed mRNA catabolic process	15	10.9	5.60E-10
9	multi-organism metabolic process	15	10.9	1.30E-09
	mRNA catabolic process	15	10.9	1.40E-09
	establishment of protein localization to membrane	18	13.1	3.40E-09
	extracellular matrix organization	17	12.4	5.90E-09
	ribosome biogenesis	17	12.4	5.90E-09
	extracellular structure organization	17	12.4	6.10E-09
	RNA catabolic process	15	10.9	7.10E-09
	protein localization to organelle	25	18.2	9.40E-08
	ncRNA processing	17	12.4	1.10E-07
	ribonucleoprotein complex biogenesis	18	13.1	1.60E-07
	establishment of protein localization to organelle	21	15.3	1.90E-07
	protein localization to membrane	18	13.1	2.00E-07
	translation	20	14.6	4.60E-07
	collagen catabolic process	8	5.8	6.50E-07
	peptide biosynthetic process	20	14.6	8.40E-07
	amide biosynthetic process	21	15.3	8.70E-07

Table B.9: Biological enrichment analysis of the 9^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Cluster	Term	Count	%	P-Value
10	single-multicellular organism process	108	67.1	6.30E-18
	anatomical structure development	100	62.1	9.80E-16
	developmental process	101	62.7	2.20E-15
	single-organism developmental process	99	61.5	4.20E-15
	multicellular organism development	92	57.1	1.20E-14
	response to growth factor	32	19.9	2.60E-14
	circulatory system development	38	23.6	3.50E-14
	cardiovascular system development	38	23.6	3.50E-14
	system development	84	52.2	1.00E-13
	multicellular organismal process	110	68.3	3.00E-13
	cell proliferation	52	32.3	3.20E-13
	cellular response to growth factor stimulus	30	18.6	3.60E-13
	tissue development	50	31.1	4.50E-13
	regulation of cell proliferation	47	29.2	4.80E-13
	response to abiotic stimulus	39	24.2	5.80E-13
	vasculature development	29	18	1.70E-12
	animal organ development	68	42.2	2.10E-12
	blood vessel development	28	17.4	2.80E-12
	response to oxygen levels	22	13.7	8.40E-12
	regulation of multicellular organismal process	60	37.3	1.50E-11
	anatomical structure morphogenesis	59	36.6	2.10E-11
	response to decreased oxygen levels	21	13	2.30E-11
	negative regulation of multicellular organismal process	35	21.7	4.60E-11
	epithelial cell proliferation	21	13	7.30E-11
	cell adhesion	45	28	1.00E-10
	response to hypoxia	20	12.4	1.00E-10
	cell migration	37	23	1.10E-10
	biological adhesion	45	28	1.10E-10
	negative regulation of biological process	81	50.3	1.30E-10

Table B.10: Biological enrichment analysis of the 10^{th} gene cluster in the BGEFM (Figure 4.9) [5]. Count indicates the number of genes which enriched in the corresponding term, and % indicates the percentage.

Neurodevelopment Events	Equivalent PC days
Medial forebrain bundle appears	33
Posterior commisure appears	33
Subplate onset of neurogenesis	35
Mammillo-thalamic tract appears	44
Stria medullaris thalami appears	44
Cortical plate first observed/visible, first cortical plate neurons born	50
LGN axons in subplate	50
Axons in optic stalk	51
Neocortical layer 1 emerges	52.5
Cortical subventricular zone (abventricular cells) onset	54.3
External capsule appears	56
Stria terminalis appears	56
Fornix appears	63
Internal capsule appears	63
Anterior commisure appears	70
Cortical plate apoptosis onset	77
GABA positive cells appear in lower intermediate zone	77
GABAir cells in lower intermediate zone/subventricular border (no synapses)	77
Hippocampal commisure appears	77
Subplate and intermediate zone apoptosis onset	77
Corpus callosum appears	81
Dopaminergic axons from the midbrain reach the subplate	91
Dopaminergic axons from the midbrain reach the cortical plate	105
Pulvinar projections in the intermediate zone of the developing pre striate iso cortex are present	105
Range of rapid synaptogenesis (VC) start	120
Pulvinar projections in the sub plate of the developing prestriate iso cortex are sparse	129.5
Onset of retinal waves stage 1	154
Eye opening	157.5
Pulvinar projections in the cortical plate of the developing prestriate isocortex are present	161
Dopaminergic axons from the midbrain are found throughout the cerebral cortex	168
Inferior cerebellar peduncle myelination onset	168
Medial lemniscus myelination onset	168
Onset of decrease in fractional anisotropy	173
Ipsi/contra segregation in LGN	175
Ipsi/contra segregation in SC	175
Superior cerebellar peduncle myelination onset	182
Lenticular fasciculus myelination onset	196
Corticospinal projections reach lumbar levels	203
End of retinal waves stage 1	210
Optic tract myelination onset	224

Table B.11: List of the neurodevelopment events of human brain [6]. Sorted in the increasing PC days order.

Neurodevelopment Events	Equivalent PC days
Anterior commisure myelination onset	266
Brachium inferior colliculus myelination onset	266
Optic radiation myelination onset	266
Fornix myelination onset	273
Stria medullaris myelination onset	273
Middle cerebellar peduncle myelination onset	280
Auditory radiation myelination onset	287
External capsule myelination onset	294
Lateral Geniculate Nucleus Myelination Onset	294
Mammillothalamic tract myelination onset	308
Cingulum myelination onset	329
Corpus callosum body myelination onset	350
Internal capsule myelination onset	350
Semi-adult like sleep cycle	361
Hippocampus myelination onset	371
Fasciculus retroflexus myelination onset	378
Splenium myelination onset	378
Stria terminalis myelination onset	378
Optic tract myelination end	483
Visual cortex peak synaptic density (area 17)	510
Range of rapid synaptogenesis (VC) end	580
Postconception walking onset	627
Corpus callosum body myelination end	735
Middle cerebellar peduncle myelination end	833

Table B.12: List of the neurodevelopment events of human brain (continued) [6]. Sorted in the increasing PC days order.