Computational Analysis of Gene Expression and Connectivity Patterns in the Convoluted Structures of Mouse Cerebellum

Tao Zeng
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/computerscience_etds

Part of the Computational Biology Commons, Genetics Commons, and the Numerical Analysis and Scientific Computing Commons

Recommended Citation
Zeng, Tao. "Computational Analysis of Gene Expression and Connectivity Patterns in the Convoluted Structures of Mouse Cerebellum" (2014). Master of Science (MS), Thesis, Computer Science, Old Dominion University, DOI: 10.25777/pc7t-c169
https://digitalcommons.odu.edu/computerscience_etds/167

This Thesis is brought to you for free and open access by the Computer Science at ODU Digital Commons. It has been accepted for inclusion in Computer Science Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
COMPUTATIONAL ANALYSIS OF GENE EXPRESSION AND CONNECTIVITY PATTERNS IN THE CONVOLUTED STRUCTURES OF MOUSE CEREBELLUM

by

Tao Zeng
M.S. July 2008, Chinese Academy of Sciences, China

A Thesis Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY
June 2014

Approved by:

Shuiwang Ji (Director)

Andrey Chernikov (Member)

Ravi Mikkamala (Member)
ABSTRACT

COMPUTATIONAL ANALYSIS OF GENE EXPRESSION AND CONNECTIVITY PATTERNS IN THE CONVOLUTED STRUCTURES OF MOUSE CEREBELLUM

Tao Zeng
Old Dominion University, 2014
Director: Dr. Shuiwang Ji

One significant difference between evolved mammalian brains and other species is that mammalian brains exhibit increasingly convoluted structures in the cerebral cortex. Groove and ridge shaped structures named gyri and sulci expand surface area of cerebral cortex, making more functions possible. Prior studies using neuroimaging techniques such as dMRI and DTI have revealed that neural fibers are heavily connected to gyri comparing to those connected to sulci, such macro-scale experiments indicates that gyri are involved in large scale information processing while sulci process information locally. However, molecular and cellular level evidences, namely, gene expression pattern and its resulting neuronal connectivity pertaining to such findings are still lacking. The Allen Mouse Brain Connectivity Atlas provides a comprehensive mouse brain neuronal connectivity map, which is build from brain-wide injection sites via anterograde tracers coupled with serial two-photon tomography. In addition, the Allen Mouse Brain Atlas offers a genome-wide gene expression database built upon a series of in situ hybridization images covering whole mouse brain. These concurrent and co-registered datasets provide an unparalleled opportunity for systematically analyzing and characterizing spatial neuronal connectivity and gene expression patterns. Inspired by recent macro-scale neuroimaging results showing that there are significantly different structural and functional connectivity patterns on the gyri and sulci of cerebral cortex in primate brains, this thesis research systematically examines the axonal connectivity and gene expression patterns on gyri and sulci of the cerebellum. The results demonstrate that the cerebellum gyri and sulci of rodent brains are significantly different in both axonal connectivity and gene expression patterns. This discovery enriches and extends prior findings in macro-scale neuroimaging studies in primates. Additionally, this work offers novel insights on the molecular and structural architectures of the cerebellum in particular and the brain in general.
ACKNOWLEDGEMENTS

This research work would not have been possible without the help and the guidance of following individuals who contributed their valuable assistance immensely in the preparation and completion of this thesis.

First and foremost, I would like to express my special gratitude to my advisor, Prof. Shuiwang Ji, for all his guidance, inspiration and continue support throughout my thesis research. He is an incredible advisor, an enthusiastic researcher, and an easygoing friend. The experience with him are my lifelong assets. I would also like to thank Prof. Tianming Liu and his Ph.D student Hanbo Chen for their inspirational ideas, informative discussions and hard works contributing to this thesis.

I would like to thank my thesis committee members, Prof. Andrey Chernikov and Prof. Ravi Mukkamala, for their precious time, guidances, valuable interactions and thoughtful comments.

I have been very fortunate to work on the mouse brain projects, which involve a group of incredible people to whom I deeply indebted. I would like to thank my collaborators: Ahmed Fakhry for his contribution to the voxel selecting tool, Rongjian Li and Wenlu Zhang for their insightful discussions and valuable interactions. I also would like to acknowledge Allen Institute for Brain Science for making the Allen Brain Atlas data available.

Last but not least, I want to thank my family for their support and understanding.

My thesis research work is supported in part by the National Science Foundation grant DBI-1147134 and DBI-1356621.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................. vii

LIST OF FIGURES ................................................................................. viii

Chapter

1. INTRODUCTION ............................................................................. 1
   1.1 MOTIVATION ............................................................................. 3
   1.2 OBJECTIVE .............................................................................. 4
   1.3 THESIS STATEMENT .................................................................. 5
   1.4 CONTRIBUTIONS ....................................................................... 6

2. TERMINOLOGY ............................................................................... 8
   2.1 GYRUS AND SULCUS .............................................................. 8
   2.2 IN SITU HYBRIDIZATION ......................................................... 8
   2.3 VIRAL NEURONAL TRACING ................................................... 9
   2.4 DIFFUSION MAGNETIC RESONANCE IMAGING ....................... 10
   2.5 FUNCTIONAL MAGNETIC RESONANCE IMAGING .................... 11

3. MATERIAL AND METHODS ............................................................ 12
   3.1 ALLEN MOUSE BRAIN ATLAS ................................................. 12
   3.2 ALLEN MOUSE BRAIN CONNECTIVITY ATLAS ....................... 14
   3.3 ALLEN REFERENCE ATLAS ..................................................... 15
   3.4 MANUAL IDENTIFICATION OF VOXELS IN GYRI AND SULCI . 15
   3.5 PROCESSING OF ACA DATA ..................................................... 18
   3.6 PROCESSING OF ABA DATA ..................................................... 19
   3.7 SPARSE MODELS ..................................................................... 19
   3.8 STABILITY SELECTION ............................................................ 21

4. RESULTS AND DISCUSSION .......................................................... 22
   4.1 GYRI CONTAIN STRONGER CONNECTIONS THAN THOSE OF SULCI AND CONNECT TO BROADER REGIONS OF THE BRAIN . 22
   4.2 GYRI AND SULCI HAVE DIFFERENT CONNECTIVITY PATTERNS .......................................................... 27
   4.3 GYRI AND SULCI EXHIBIT DIFFERENT GENE EXPRESSION PATTERNS .................................................. 29
   4.4 GYRUS, SULCUS AND NEGATIVE VOXELS CAN BE ACCURATELY CLASSIFIED BASED ON THEIR GENE EXPRESSION PATTERNS ................................................................................. 31

5. CONCLUSION AND FUTURE WORK .............................................. 33
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gyri Sulci connection statistics</td>
<td>22</td>
</tr>
<tr>
<td>2. Pairwise comparison of connectivity strength between gyri and sulci</td>
<td>26</td>
</tr>
<tr>
<td>3. The prediction AUC and accuracy for the three classification problems: gyrus against sulcus, gyrus against negative, and sulcus against negative</td>
<td>32</td>
</tr>
<tr>
<td>4. The prediction AUC and accuracy for the three classification problems: gyrus against sulcus, gyrus against negative, and sulcus against negative</td>
<td>32</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.</td>
<td>Illustration of gyrus and sulcus in the cerebral cortex.</td>
</tr>
<tr>
<td>2.</td>
<td>Illustration of In Situ Hybridization</td>
</tr>
<tr>
<td>3.</td>
<td>Sample brain section image of viral neuronal tracing from allen connectivity atlas</td>
</tr>
<tr>
<td>4.</td>
<td>Flowchart illustrating the data processing steps for the three data sets</td>
</tr>
<tr>
<td>5.</td>
<td>Sample ISH images of allen developing mouse brain atlas</td>
</tr>
<tr>
<td>6.</td>
<td>Visualization of the injection sites for the 1019 viral tracers in the mouse brain</td>
</tr>
<tr>
<td>7.</td>
<td>Sample brain section images of allen reference atlas</td>
</tr>
<tr>
<td>8.</td>
<td>A snapshot of the in-house software tool for identifying the gyrus and sulcus voxels</td>
</tr>
<tr>
<td>9.</td>
<td>The numbers of injection sites from which gyrus and sulcus connect</td>
</tr>
<tr>
<td>10.</td>
<td>3D visualization of connectivity</td>
</tr>
<tr>
<td>11.</td>
<td>Heat maps for the connectivity of gyri and sulci from the 1019 injection sites</td>
</tr>
<tr>
<td>12.</td>
<td>The correlations of connectivity for gyri, sulci and negative regions</td>
</tr>
<tr>
<td>13.</td>
<td>The correlations of gene expression for gyri, sulci and negative regions</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

The essential role of nerve system is to ensure survival of individual and the species it belongs to. The nerve system gathers external and internal information, processes those information and response properly to the circumstances to which individual is being exposed. From evolutilional point of view, more elaborate systems are generally more successful in terms of ensuring survival. This is due to the capacity of generating more complex behaviors when being presented to the challenge of the complex environments. The highly evolved mammalian species exhibit far more complicated behaviours than other species. One pronounced change in mammalian brain is that the volume of cerebral cortex appears to increase as mammalian evolving. In addition, the convolutional structures appeared in mammalian brain allows the area of surface to be considerably expanded, leading to the increase of functionalities in the brain. Both changes are considered to be highly correlated with the increase capacity of intelligence.

One of the most significant shape and anatomical features of the evolved mammalian brains lies in their increasingly convoluted cerebral cortex [1]. The convex structures on the cerebral cortex are named gyri, while the fissures surrounding the ridges are known as sulci. In human brain mapping, the shape characteristics of cortical gyri and sulci have been extensively examined [2], [3], [4], [5], [6], in both healthy and diseased brains. However, quantitative modeling and description of the intrinsic relationships between gyral/sulcal folding patterns, as well as their connectivity patterns and functions, have been much less investigated to date. In recent years, increasing attention and interests have been devoted to this new research direction, and several inspiring and interesting discoveries have been reported [7], [8], [9].

For example, in recent macro-scale study based on diffusion magnetic resonance imaging (dMRI) [8], a keen correlation between cortical convolution and connectivity of streamline fibers was identified. Specifically, the analysis on diffusion tensor imaging (DTI) and high angular resolution diffusion imaging (HARDI) arguably demonstrated that the termination of dMRI-derived streamline fibers concentrate
on gyrus instead of sulci. It has also been shown that this phenomenon is conserved in cerebellum of human beings, chimpanzees and macaques, and mice [7], [8], [9]. Similar observation has been made in another study on human fetus brain [10], which demonstrated that radial organization persisted longer in the crests of gyri than at the depths of sulci in fetus brain. These findings suggest a possible regulation mechanism between brain wiring formation and gyrification progress during development. Moreover, inspired by these structural findings, a recent study using human resting state functional magnetic resonance imaging (fMRI) data demonstrated that the functional interactions between gyral-gyral regions are significantly stronger than those between sulcal-sulcal regions [9]. These studies suggest a well conserved phenomenon across cerebrum and cerebellum, as well as across species. To fully understand the mechanism and causality behind this phenomenon, it would be interesting to explore the structural and functional implications of cortical gyri and sulci folding patterns at the molecular and cellular levels [11].

Histology is a common approach in analyzing anatomy of neurons and their axons. With a wide range of dyes and axonal tracers, qualitative and quantitative morphological analysis of neuronal connectivity can be performed in mammalian brains [12], [13], [14], [15]. Compared with dMRI, the merit of the histological approach is that it offers direct observation of neurons and their axons under a microscope that can then be taken as ground truth. For instance, traditional myelin and Nissl staining has been adopted in the validation frameworks of several dMRI studies [16], [17], [18], [19]. However, one of the limitations of such approaches lies in that only a small portion of the projections can be observed in an experiment. Therefore, traditional studies have focused on one specific anatomical pathway at a time. Recent growing research interest in functional brain interaction and large-scale projections raises the demand for a comprehensive map of neuronal networks, i.e., a connectome. With the advances in anatomical tracing techniques, a few research teams are building a complete circuitry of the mouse brain that provides an ideal experimental platform to understand brain wiring patterns [20].

One example of the above-mentioned effort is the Allen Mouse Brain Connectivity Atlas (ACA) [21] that has been recently made available (http://connectivity.brain-map.org/). In this comprehensive mouse brain neuron connectivity map, axonal projection has been traced anterogradely via recombinant adeno-associated virus (rAAV) from different injection sites covering the whole brain and visualized using
serial two-photon tomography. Moreover, the Allen Institute for Brain Science also produced the Allen Mouse Brain Atlas (ABA) [22], a genome-wide gene expression database built upon a series of images of in situ hybridization. This atlas compiles comprehensive expression energy of 4,084 genes across the whole mouse brain, providing valuable resources for researchers to systematically analyze and characterize spatial gene expression patterns in the mouse brain. In particular, by joint analysis of the ACA and ABA, exploring gene expression pattern spatially together with neuronal connectivity is of great interest, as it could reveal novel insights on how genes are associated with the formation and function of neuronal projections in the brain. To facilitate the integrative analysis of the gene expression and the connectivity data, a unified 3-D reference space, the Allen Reference Atlas (ARA) [23], was created. The ARA is a grid-voxelized map labeled with structures of anatomical ontology for every voxel across the mouse brain. All the images in the ABA and the ACA are aligned to the ARA, thereby enabling the integrative analysis of both atlases.

Despite the fact that current neuroimaging studies have shown that the white matter fibers connected to gyri are significantly denser than those of sulcus, the neuronal connectivity and gene expression differences between gyrus and sulcus are still lacking at the micro-scale. The primary goal of this thesis is to bridge this gap by computationally analysing gene expression and neuronal connectivity in gyri and sulci using ACA an ARA datasets.

1.1 MOTIVATION

In primate brain, the structures of gyri and sulci significantly expand the surface of cerebral cortex, leading to the possibility of forming more sophisticated functions than those species with less pronounced convolutional structures. Recent advance in neuronimaging studies have revealed that neuron fibers are heavily terminated in gyri comparing to those in sulci. Based on this macro-scale funding, the fiber pushing hypothesis regarding the formation of gyri is proposed. However, the cellular and molecular level studies pertaining to the intrinsic relations of gyri and sulci are still lacking. Since the brain functions are established based on the connectivity of neurons at cellular level, and gene expressions of neural cells are the determining forces underlying the neuronal connectivity of the brain at molecular level. Hence, studying the relation of gene expressions and neuronal connectivity between gyri and sulci is of paramount importance to understand both formation and function of gyri
and sulci.

1.2 OBJECTIVE

Prior macro-scale studies on gyri and sulci mainly focused on primate cerebral cortex due to their pronounced convolutional structures and their related role in the study of intelligence evolution. Therefore, it would be ideal to study the same questions at molecular and cellular level using the subjects of similar species. However, the biological experiments data conducted on the primate are currently neither comprehensive nor publicly available. Whereas, ABA and ADA data sets provide the comprehensive data sets of gene expression and neuronal connectivity covering entire adults mouse brain. However, only structures that exhibits pronounced convolutional shape in mouse brain is the cerebellum. Hence, instead of cerebral cortex, this thesis focus on analysis of mouse cerebellum data.

The primary goal of this thesis’s study is to systematically investigate the differences between gyri and sulci at both molecular and cellular level in mouse brain. To this end, this primary goal can be completed through following sub-objectives:

- **Developing a tool for manual voxel selection**
  Developing a tool which allows the manual determination of whether voxel in 3D mouse brain atlas belonging to gyri or sulci. One challenge in pursuit of this objective is to segment structures of gyri and sulci from cerebellum of mouse brain in 3D atlas. ACA, ABA are the voxelized brain 3-D mouse brain atlas which provide gene expression and neuronal connectivity data, respectively. in addition, ARA provides corresponding voxelized 3D annotation of brain structures based on neural anatomy. However, the ARA annotations do not encode the gyrus and sulcus information explicitly and there are no readily available tools for automatic segmentation of the cerebellum gyral crown, sulcal roots and the walls in-between. Thus, the first objective of this work is to manually determine the voxels of gyri and sulci by developing a visualizing tool which enables the manual selection of voxels and categorizing them as two groups: gyri and sulci.

- **Analysing gene and connectivity correlation with in group and between groups**
  In this objective, the coordinates of identified voxels extracted by manual voxel
selection tool are applied to ACA and ABA datasets to extract gyri and sulci's corresponding gene expression and connectivity data, respectively. Based on the known fact from prior studies that similar gene expression would result in a similar phenotype. This fact elucidates the possible biological mechanism on how molecular level gene expression leads to phenotypes of the cells at the cellular level. Namely in this research, gene expression and neuronal connectivity are expected to be similar within the gyri or sulci group and dissimilar between groups. Therefore, the subsequent study of this objective is to analysis both correlations of gene expression and connectivity in the group of gyri and sulci and between the group. Pursuing this objective would demonstrate how gyri differ from sulci in terms of gene expression and neuronal connectivity at micro-scale.

- **Analysing breath of connectivity in gyri and sulci**
  As prior studies in macro-scale showed that gyri contains long-range neuronal projections connecting to different regions of brain, thereby performing very different neuronal functions as compared to sulci. Hence, gyri are expected to have very diverse connectivity patterns. In comparison, sulci are usually connected to fewer number of regions and is therefore considered as local information processing units [9]. The third objective is to study the connectivity breadth of each gyrus or sulcus to the rest of brain at cellular level in micro-scale which may be consistent to those findings in macro-scale.

- **Computationally predicting the class of voxels**
  With just grid-voxels gene expression data, the objective is to apply pattern classification algorithms to determining whether they belong to gyri and sulci and how well that they can be correctly classified. This experiment could computationally quantizing the differences in gene expression pattern between of gyri and sulci and connecting those differences to characteristics of neuronal connectivity.

1.3 **THESIS STATEMENT**

Gyri and sulci are the convolutional structures which are primarily observable in
primate brain. Revealed by prior studies using neuroimaging techniques, such convolutional structures exhibits significant differences in terms of connectivity. Neuroimaging techniques are considered as macro-scale assessment in studying of brain. Whereas, the underlying biological mechanism lies in the micro scale. Namely, the molecular level of gene expression profile and their resultant cellular level phenotypes. Thus, by studying relations of gene expression and neuronal connectivity in gyri and sulci of mouse cerebellum, this thesis revealed the differences between gyri and sulci at micro-scale, which is consistent with the finding in the macro-scale.

1.4 CONTRIBUTIONS

Despite the fact that current neuroimaging studies have shown that the white matter fibers connected to gyri are significantly denser than those of sulci, the neuronal connectivity and gene expression differences between gyrus and sulcus are still lacking at the micro-scale. To bridge this gap, I studied the neural connectivity and gene expression patterns of gyri and sulci in the mouse cerebellum. As illustrated in Figure 4, I first built in-house software based on the ARA to enable visualization of gyri and sulci at the voxel level. By using this software tool, I manually selected 12 gyri and 14 sulci covering one hemisphere of mouse brain cerebellum. The resulting 3-D coordinates of those voxels were applied to the ABA and ACA to extract the corresponding data of connectivity and gene expression belonging to those identified gyri and sulci. In addition, data from voxels belonging to neither gyri nor sulci were also extracted to form the negative group. I compared the neuronal connectivity between gyri and sulci and the results showed that, in general, gyri project to more brain regions and have substantially stronger overall connectivity than those of sulci. Investigation of their connectivity patterns has shown that correlations are significantly higher within groups than between groups of gyri and sulci. This result is consistent with the findings in neuroimaging stating that gyri and sulci have different patterns of white matter fibers connections. I subsequently compared the gene expression patterns among gyri, sulci and negative voxels. The results demonstrated that gene expression within sulci has the highest correlations when compared with gyrus-gyrus, negative-negative, gyrus-sulcus and gyrus-negative. To further investigate whether gyri, sulci and negatives have distinctive gene expression signatures, I conducted a series of classification tasks to discriminate gyrus, sulcus, and negative samples. The prediction of gyrus against negative yielded an accuracy value
of 82% and an area under the receiver operating characteristic (ROC) curve (AUC) of 87%. Sulcus against negative classification achieved an accuracy of 84% and an AUC of 86%. The highest among all the classification tasks was the gyrus against sulcus classification, which produced an accuracy of 92% and an AUC of 97%. Such high predictive performance demonstrates that gyrus and sulcus can be accurately identified based on their gene expression patterns. In addition, gyrus and sulcus can be recognized from negative samples with an accuracy that is consistently over 80%. This also demonstrate that high predictive performance can be obtained by using a very small subset of the genes.
CHAPTER 2

TERMINOLOGY

This thesis is focused on computational analysis of convolved structures in the brain, which falls into the scope of computational neuroscience. Hence, the aim of this chapter is to briefly discuss some of the terms in neuroscience related to experiments of computational analysis in the later chapters.

2.1 GYRUS AND SULCUS

The cerebral surface appears to increase as species evolve. In many mammal species, convoluted cerebral cortex can be observed. This is particularly pronounced in primate. Such convolution of brain is considered to expand surface area of cortex and allow greater functionality to fit into a smaller cranium, thus play an crucial important role in the species evolution. In rodent, instead of cerebral cortex, however the similar convolutional structure can only be observed in cerebellum. In neuroanatomy, gyrus refers to the concave structure and sulcus is a groove one in the brain. Sulcus is surrounded by gyrus, leading to folded appearance in the brain of many mammal species.

Fig. 1. Illustration of gyrus and sulcus in the cerebral cortex.
2.2 IN SITU HYBRIDIZATION

In situ hybridization is a RNA or DNA detecting technique which uses a labeled hybridization probe to localize a specific sequence of DNA or RNA strand that is complementary to the sequence of probe. During ISH experiments, the target transcript RNA on the section of tissue are usually fixed in place in order to be accessed by hybridization probe. With an appropriate temperature, the probe is then hybridized to the target DNA or RNA on the tissue section due to the complementarity between the probe and target (The process is illustrated in Figure 2 and the actual resultant ISH images from ABA are shown in Figure 5 ). The remaining excess probe on the tissue section that are not hybridized is washed away. Since probe is labeled with molecular marker which is usually radioactive or florescent molecules, the target DNA or RNA which hybridized with the probe due to the sequence similarity can be detected and localized by imaging techniques such as autoradiography and fluorescence microscopy.

Fig. 2. Illustration of In Situ Hybridization.
2.3 VIRAL NEURONAL TRACING

The tremendous amount of neurons are connected to form neural networks which allows complex information processing. Such connectivity is the basis of brain functionality. Hence, understanding how they are connected from a small circuit to form a large networks of entire brain is of paramount importance. This requires a tool to allow researchers to trace and visualize such connections. Viral neuronal tracing is one of such techniques which utilize human modified neurotropic viruses as tracers. Such tracer virus can replicate and spread from one neuron to another through synapse of two neurons. Depend on the viral tracer, the tracing can occurs in either anterograde direction which is from soma to synapse, or the opposite called retrograde. Since the viral tracer are modified to express some fluorescent proteins, the neuronal pathway that the viral tracer has spread through can then be measured by fluorescent imaging technique.

Fig. 3. Sample brain section image of viral neuronal tracing from allen connectivity atlas.

2.4 DIFFUSION MAGNETIC RESONANCE IMAGING

Diffusion magnetic resonance imaging is a technique that measures the diffusion process of molecules, particularly water, in biological tissues. In nature environment, water molecules move randomly according to Brownian motion. In central nervous
system, however it diffuses along white matter axons due to the constrain of myelin membrane and small diameter of axonal fibers. If water molecular are found to diffuse in one direction in a given voxel, then the majority of fibers in this voxel are assumed to be parallel to that direction. Such property is utilized to locate the neural fibers and thus help understand the neuronal connectivity of the brain.

2.5 FUNCTIONAL MAGNETIC RESONANCE IMAGING

The blood flow and oxygenation in the brain are known to be closely related to neural activity. This is, when an region of the brain become active, blood flow to this region increases accordingly. Based on this fact, functional MRI (fMRI) is a MRI technology that measures brain activity by detecting related changes in blood flow.
CHAPTER 3

MATERIAL AND METHODS

For the experiments in this work, ABA, ACA and ARA datasets were used, which represent gene expression, neuronal connectivity and a 3-D anatomical reference atlas of the mouse brain, respectively. The data processing steps are outlined in Figure 4. Since the shapes of gyrus and sulcus in primate cerebral cortex are similar to those of cerebellum (CB) in mouse, This work limited the gyrus and sulcus localization only to regions of the CB in the mouse brain. I manually identified voxels of gyri and sulci from CB using in-house software built for this study. The resulting coordinates of voxels of each gyrus or sulcus were applied to the matrices of the ABA and ACA to extract corresponding gene expression and neuronal connectivity. Subsequently, this study then conducted a series of correlation and prediction experiments to elucidate the neural connectivity and gene expression differences between gyri and sulci.

3.1 ALLEN MOUSE BRAIN ATLAS

The ABA is considered a the most complete of the molecular brain atlas to allows researchers to understand brain gene expression at a genomic level. It contains high-throughput in situ hybridization (ISH) gene expression pattern images for the 56-day old male mice [24], [25], [26]. The ABA is a genome-wide, 3-D database of gene expression in the adult mouse brain consisting of approximately 20,000 genes in the sagittal section and over 4,000 genes with restricted expression patterns in the coronal section. This allows researchers to understand brain gene expression at a genomic level. For each gene expression experiment, ISH method for probing a specific gene expression was applied to multiple section of one mouse brain specimen to generate a serial of corresponding 2-D images. These images were then processed by an informatics data processing pipeline [27]. In this process, they were co-registered to a standardized three-dimensional reference space of the allen reference atlas to generate grid voxel-level data. The resulting data is quantified gene expression values generated from many mouse brains at fine grid voxel-level (resolution 100 um-300um), and each voxel is annotated with the brain structure it belongs. Coronal
Fig. 4. Flowchart illustrating the data processing steps for the three data sets. A: Steps for gene expression data processing. B: Steps for connectivity data processing. C: Steps for manual selecting voxels in gyri and sulci.

data was used in this work, since this set of genes show restricted expression patterns and thus are of particular interest. the allen informatics data processing pipeline consists of 6 components: a preprocessing module, a 3-D reference model, an alignment module, an expression detection module, an expression gridding module and a structure unionizer module.

- preprocessing module
  The preprocessing module white balance and normalize intensity of image, followed by morphological filtering and connected component analysis to remove noise and connect broken segments.
• **3-D reference model**
  The brain was sliced to 528 sections covering an entire brain, each with 25um thick. Such high-frequency sections images were then section-to-section registered to low-frequency histology images combined with MRI registration. Subsequently, a brain volume was the reconstructed from those section images with over 800 structures annotated 3-D reference space.

• **alignment module**
  The alignment module co-register all image sections of a specimen with the 3-D reference model to reconstruct a consistent 3-D volume though all specimen.

• **expression detection module**
  Tissue area segmentation, small and isolated object segmentation and dense and clumped object segmentation algorithms are applied to each ISH image to detect gene expression. The output result is a grayscale mask of identified pixels image which correspond with gene expression.

• **expression gridding module**
  The Gridding module operates on a per image-series basis. Each image is divided into a 200 m x 200 m grid. Based on this gridding, it create a low-resolution, grid-voxel level averaged gene expression using the common coordinate space of the 3-D reference model. This 3-D grid-voxel level data enables spatial comparison between different specimens. This module operates on a per image-series basis.

• **structure unionizer module**
  Structure unionizer module combines grid-level data to compute the gene expression for each structure delineated in the 3-d reference atlas, enabling researchers to compare gene expressions in a hierarchy of multiple structure levels.

### 3.2 ALLEN MOUSE BRAIN CONNECTIVITY ATLAS

The Allen mouse brain connectivity atlas contains 3-D high-resolution maps of neural connectivity in the adult mouse brain. The neural projection were detected by injection of viral tracers and captured by using serial two-photon tomography. The viral tracers were injected at total 1019 injection sites corresponding to 210 unique anatomical structures (primary injection structures) of the brain. The 3D locations
Fig. 5. Sample ISH images of allen developing mouse brain atlas for genes Neurod6, Gfap, and Ugt8a, respectively. Selected images from approximately the same location were shown for coronal (top) and sagittal (bottom) sections.

of injections sites are given in Figure 6. For each experiment, 140 high-resolution, 2-D images were used to generate a brain-wide projection dataset. These images were subsequently co-registered into the ARA and converted into grid voxel-level data by the informatics data processing pipeline [27]. Such co-registership facilitates the integrative analysis of the connectivity atlas and other mouse brain atlases such as the gene expression atlas.

3.3 ALLEN REFERENCE ATLAS

The ARA is a standardized 3-D space into which both the ISH and the projection images were registered. It is a color-coded, high-resolution, and web-based digital brain atlas accompanied by a systematic, hierarchically organized taxonomy of mouse brain structures [23]. Over 800 structures were annotated on 2-D images, and interpolated to create annotations in the reconstructed 3-D brain atlas, yielding a fully annotated reference space. The ARA space is divided into regular grids of 100m and 200m resolutions. The ABA and ACA data were processed at the 200m and 100m levels, respectively.
Fig. 6. Visualization of the injection sites for the 1019 viral tracers in the mouse brain. Locations of the 1019 viral tracer injection sites covering one hemisphere of the mouse brain are displayed in the coronal (Left figure) and sagittal (Right figure) sections. Each filled circle represents the location of one particular injection site, and the color indicates the brain structure to which the injection belongs based on the neuroanatomical ontology of the ARA. The numbers of injections in the cerebral cortex and the cerebellum are 474 and 34 out of the total 1019 injections respectively. These numbers correspond to unique injections in cerebral cortex and cerebellum of 64 and 12 out of 209 total unique injections respectively. Figures reproduced from the allen brain atlas data portal with permission.

3.4 MANUAL IDENTIFICATION OF VOXELS IN GYRI AND SULCI

Among all regions in the mouse brain, only structures in the CB exhibit distinctive shapes of convoluted folding patterns similar to gyri and sulci in the cerebral cortex of primate brains. Accordingly, this thesis only focused on structures from the CB to identify gyrus and sulcus voxels. Aiming at investigating the differences between gyri and sulci in terms of connectivity and gene expression. Accurate identification of voxels in gyri and sulci in the mouse brain is the crucial first step in pursuing objective of this study. However, the ARA annotations do not encode the gyrus and sulcus information explicitly. In addition, since there are no readily available tools for automatic segmentation of the cerebellum gyral crown, sulcal roots and the walls in-between (as far as I know), this work chose to manually determine the voxels of gyri and sulci. Specifically, using the 3-D coordinates of the annotated voxels in the ARA at 100m resolution, an in-house software developed in MATLAB 7.0
(The Mathworks, Natick, MA) to visualize these voxels was build. I was then able to navigate through the sagittal planes of the brain section-by-section using our in-house tool and manually select those voxels belonging to gyri and sulci. To facilitate manual identification of voxels of gyri and sulci, all structures in the CB were color-coded, and each structure was visualized to distinguish the molecular and granular layers. For each sagittal section, the voxels of gyri were selected from the convex regions of each structure, and voxels of sulci were determined from concave regions of two adjacent structures. To maintain high accuracy during voxel localization, section images from the ARA in the sagittal plane were displayed side-by-side and utilized simultaneously, as illustrated in Figure 4b. Additionally, the procedure illustrated in Figure 8 was undertaken to minimize the potential bias due to manual selection, and to deal with the discrepancies of selecting voxels between different experimenters. Voxels that were located at the region of 1/4 of convex crest, as opposed to the whole annotated structure, were identified as belonging to gyrus. The same criterion was used to identify sulcus voxels from the concave region of two adjacent annotated structures. The resulting coordinates of selected voxels of each individual gyrus and
sulcus were saved for subsequent analyses.

Fig. 8. A snapshot of the in-house software tool for identifying the gyrus and sulcus voxels. An ARA sagittal image (A) is compared with its corresponding section by the in-house software for voxel localization. The orange box with dash lines contains one annotated structure, and the region of gyrus enclosed by the red box illustrates the criteria of identifying gyrus region used in our study. Similarly, the blue box with dash line contains a concave region of two adjacent structures and the purple box shows the criteria of identifying sulcus (B). The process of voxel selection is illustrated through the small blue sub-image in (B) which represents the axial plane, and the red line indicates the sagittal section coordinates of the displayed data. The red dots in (B) are the voxels identified manually to be gyrus voxels while the purple ones represent sulcus voxels.

3.5 PROCESSING OF ACA DATA

The ABA data portal provides a set of application programmable interfaces (APIs) [28] to allow access to various data. These APIs was used to download the voxel-level grid data in the ABA and ACA. Prior to performing the connectivity analysis of gyri and sulci, the connectivity values of voxels corresponding to injection sites were set to zero to remove the effect of over intensified signals due to their close vicinity to injections. Figure 4 illustrates the steps involved in connectivity data processing in this study. The connectivity matrix for each individual gyrus and sulcus was extracted by applying the 3-D coordinates of each gyrus and sulcus voxels to the aggregated voxel-level connectivity matrix, yielding 26 matrices representing
12 gyri and 14 sulci. In each matrix, the number of rows equals to the number of selected voxels for a specific gyrus or sulcus, and the number of columns equals to the number of injection experiments. Consequently, values in each row of those matrices represent connectivity strength of a specific voxel from all injection sites available. The rows of each of these matrices were further averaged to represent the overall neuronal connectivity strength for each individual gyrus or sulcus. Statistical analysis of the connectivity data found that the distribution is highly skewed with a mean of 0.1994 and a standard deviation of 6.335 on all identified voxels. This implies that most of the data values are distributed below 1, and a small number of them are much larger. This causes difficulty in data visualization when comparing gyri with sulci. Therefore, this work converted the raw connectivity data by a logarithmic transform and used the transformed values for subsequent analyses and visualizations. Heat maps (Figure 11) were constructed to elucidate the connectivity strength over all injection sites for each gyrus and sulcus. To determine whether gyri, sulci and negative samples have different connectivity patterns, their Pearson correlation at both voxel-level and gyrus-sulcus level were compared, as illustrated in Figure 12. Moreover, the overall connectivity was compared in the form of 1 gyrus against 14 sulci and vice versa, as shown in Table 2.

3.6 PROCESSING OF ABA DATA

The voxel-level ABA data were provided at the 200 m resolution. Since the manually selected voxels in this study were generated from the 100 m annotation data, the gene expression data was up-sampled to the 100 m resolution using linear interpolation to enable an integrative analysis of the two datasets. Then the gene expression energy values at voxel level corresponding to each gyrus and sulcus were extracted from the up-sampled data. Similar to the connectivity data, the Pearson correlation of gene expression in gyri and sulci at both voxel-level and gyrus-sulcus level were compared. For the gyrus-sulcus level comparison, the expression energy for each gene was averaged across all voxels that belonged to a particular gyrus or sulcus, yielding a vector corresponding to 4084 genes for every identified gyrus or sulcus. These data were further used as inputs to compute structure level correlations. The corresponding results are given in Figure 12.

3.7 SPARSE MODELS
Sparse linear models for the prediction analysis were used in this work, as such models yield competitive performance while retaining interpretability. In order to identify the most predictive genes while making accurate predictions, L1-norm regularized models were employed, known as LASSO [29]. These sparse models set some of the model parameters to zero as a result of minimizing the L1-norm of the parameter vector. Since each element in the parameter vector corresponds to a specific gene, setting an element of the parameter vector to zero is equivalent to eliminating the corresponding gene from the model. The predictive performance was measured by the area under the receiver operating characteristic (ROC) curve (AUC) and the accuracy. The receiver operating characteristic curve (ROC), is a graphical plot which illustrates the performance of a binary classifier system as its discrimination varies. This curve is created by plotting the fraction of true positives out of the total actual positives vs. the fraction of false positives out of the total actual negatives. When using normalized units, the area under the curve (AUC) is equal to the probability that a classifier will classify a randomly chosen positive sample higher than a randomly chosen negative one.

The validation procedure was employed in this prediction analysis, where 2/3 of the data were used to train the classifier and the remaining 1/3 are used for testing. Formally, the L1-norm regularized models optimize the following objective function:

$$\min_{w,b} \sum_{i=0}^{N} L(x_i^Tw + b, y_i) + \lambda ||w||_1$$

where $x_i$ is the gene expression vector corresponding to a specific voxel, $w$ is the weight vector, $b$ is the bias term, $y_i \in \{-1, 1\}$ is the label where 1 corresponds to a positive sample and -1 corresponds to a negative sample, $N$ is the number of voxels, and $L$ is a loss function measuring the prediction error. The logistic regression loss function was employed in this work, as it leads to competitive performance for classification tasks. This gives rise to the following minimum negative log-likelihood optimization problem (Fan, Chang, Hsieh et al. 2008):

$$\min_{w,b} \sum_{i=0}^{N} \log(1 + \exp^{-y_i(x_i^Tw+b)}) + \lambda ||w||_1$$

In Equations (3.7) and (3.7), $\lambda$ acts as a regularization parameter. Tuning this parameter leads to control over the number of selected genes. When $\lambda$ is set to a
high value, most of the elements in w are set to 0, leading to a very small number of genes selected. Conversely, setting \( \lambda \) to 0 will produce a full model, resulting in low interpretability. In the experiments, \( \lambda \) was tuned so that the number of selected genes in this model was between 50 and 250. One observation is that when re-applying linear models on the data using only the selected genes from the sparse model, the almost same accuracy can be obtained as it was achieved from using all the genes. This result is consistent with previous findings that a small number of genes usually contribute to the generation of neuronal connectivity [30].

3.8 STABILITY SELECTION

To provide a robust and principled approach for estimating the regularization parameter, stability selection [31] was applied to tune \( \lambda \) in order to obtain a set of robust genes. In stability selection, a set of \( \lambda \) values instead of a single one was chosen. For each \( \lambda \) value, the selection probability for each variable (gene) was computed, defined as the probability of each gene been selected when randomly resampling from a sub-sample of size \( N/2 \) drawn from the data without replacement. The selection probability for the \( i^{th} \) variable is defined as

\[
 k_i^\lambda = P\{x_i \subseteq S(\eta, \lambda_j)\} 
\]

where \( S(\eta, \lambda_j) \) is the set of selected genes when using the \( \eta \) subsample and the regularization parameter \( \lambda_j \), and \( P\{x_i \subseteq S(\eta, \lambda_j)\} \) denotes the frequency that the \( i^{th} \) variable is selected among the resampling. It has been shown that 100 resampling is usually sufficient for robust estimation [31]. The overall selection probability for the \( i^{th} \) variable (gene) across all different regularization values was defined as

\[
 G_i = max_{\lambda_j} \{ S(\eta, \lambda_j) \} 
\]

where the frequency of a specific gene is equivalent to its maximum frequency across all \( \lambda \) values. A gene ranking was then obtained by sorting the total list of genes based on their selection frequency.
CHAPTER 4

RESULTS AND DISCUSSION

The primary goal of the experiments in this thesis was to investigate whether gyri differ from sulci in terms of their neuronal connectivity and underlying gene expression patterns. The manual data annotation in this work generated a total of 4,739 voxels distributed across 12 gyri and 3,291 voxels distributed across 14 sulci. Detailed information on the selected gyri and sulci are given in Tables 1 and 2. The resulting coordinates of the selected voxels in gyri and sulci were used to produce the gene expression and neuronal connectivity data for each gyrus and sulcus. The entire pipeline of the analysis in this work is illustrated in Figure 4.

<table>
<thead>
<tr>
<th>Gyri and sulci abbrev</th>
<th># voxels</th>
<th># connected viral tracer injection sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN1</td>
<td>803</td>
<td>911</td>
</tr>
<tr>
<td>AN2</td>
<td>587</td>
<td>879</td>
</tr>
<tr>
<td>CENT2</td>
<td>190</td>
<td>863</td>
</tr>
<tr>
<td>CENT3</td>
<td>495</td>
<td>792</td>
</tr>
<tr>
<td>COPY</td>
<td>108</td>
<td>912</td>
</tr>
<tr>
<td>CUL</td>
<td>796</td>
<td>948</td>
</tr>
<tr>
<td>DEC</td>
<td>266</td>
<td>808</td>
</tr>
<tr>
<td>PL</td>
<td>114</td>
<td>917</td>
</tr>
<tr>
<td>NOD</td>
<td>202</td>
<td>869</td>
</tr>
<tr>
<td>PRM</td>
<td>403</td>
<td>922</td>
</tr>
<tr>
<td>SIM</td>
<td>572</td>
<td>896</td>
</tr>
<tr>
<td>UVU</td>
<td>203</td>
<td>816</td>
</tr>
<tr>
<td>AN1-AN2</td>
<td>441</td>
<td>787</td>
</tr>
<tr>
<td>AN2-PRM</td>
<td>295</td>
<td>845</td>
</tr>
<tr>
<td>CENT2-CENT3</td>
<td>223</td>
<td>526</td>
</tr>
<tr>
<td>CENT2-SU</td>
<td>86</td>
<td>754</td>
</tr>
<tr>
<td>CENT3-CUL</td>
<td>423</td>
<td>839</td>
</tr>
<tr>
<td>CUL-DEC</td>
<td>133</td>
<td>535</td>
</tr>
<tr>
<td>CUL-SIM</td>
<td>556</td>
<td>852</td>
</tr>
<tr>
<td>NOD-LING</td>
<td>51</td>
<td>575</td>
</tr>
<tr>
<td>PRM-COPY</td>
<td>286</td>
<td>433</td>
</tr>
<tr>
<td>PYR-UVU</td>
<td>226</td>
<td>538</td>
</tr>
<tr>
<td>SIM-AN1</td>
<td>224</td>
<td>585</td>
</tr>
<tr>
<td>UVU-NOD</td>
<td>233</td>
<td>630</td>
</tr>
<tr>
<td>CENT3-SU</td>
<td>10</td>
<td>400</td>
</tr>
<tr>
<td>FOTU-PYR</td>
<td>104</td>
<td>254</td>
</tr>
</tbody>
</table>

4.1 GYRI CONTAIN STRONGER CONNECTIONS THAN THOSE OF SULCI AND CONNECT TO BROADER REGIONS OF THE BRAIN
To compare the connectivity patterns between gyri and sulci, the connectivity data across the 1,019 viral tracer injection sites covering all major brain structures (Figure 6) at both voxel level and gyrus-sulcus level were analysed.

An example of connectivity from one selected injection site is jointly visualized with cerebella cortex in Figure 10. At the voxel level, for each selected voxel (gyri or sulci), a vector was extracted from the ACA representing connectivity strength from all the 1,019 injection sites. For the gyrus-sulcus level analysis, each gyrus/sulcus was represented by a vector containing 1019 entries, where the value in each entry was derived by averaging the values of connectivity from one of the 1019 viral injection sites to all the voxels that the gyrus/sulcus contains. The voxel level and gyrus-sulcus level connectivity are visualized as heat maps in Figure 11. As illustrated in Figure 11 as well as Figure 10, gyri contain substantially stronger overall connectivity from the injection sites than that of sulci.

![Graph showing connectivity degrees for gyri and sulci](image)

**Fig. 9.** Comparison between connectivity degrees of both gyrus and sulcus are shown. The connectivity degree is the number of injection sites a gyrus/sulcus is connected from out of the total number of injection sites available. A gyrus/sulcus is considered connected from a specific injection site if it has a nonzero connectivity strength value with that injection site in its connectivity vector. On average, gyri are connected to 877 (86.1%) injection sites, while sulci are connected on average to only 609 (59.7%).
To quantitatively analyze the breadth of neuronal projections in the gyri and sulci of the cerebellum, comparing their connectivity degree at the gyrus-sulcus level was performed. The connectivity degree of a specific gyrus or sulcus is the number of injection sites it is connected to out of the total number of injection sites available. A gyrus/sulcus is considered to be connected to a specific injection site if it has a non-zero connectivity energy value from that injection site in its connectivity vector. Such a connectivity measure allows us to characterize the breadth of neuronal projections from the rest of the regions in mouse brain to a particular gyrus or sulcus. From Figure 9, it can be observed that, on average, gyri are connected to 877 (86.1%) injection sites, while sulci are connected to only 609 (59.7%) injection sites. This demonstrates that gyri exhibit broader connectivity as compared to that of sulci. In addition, a thorough element-wise comparison of each gyrus and sulcus pair using the gyrus-sulcus level data was performed. This is, comparing connectivity vectors of each gyrus-sulcus pair by counting the percentage of injection sites in which the gyrus connectivity strength exceeds that of sulcus. As shown in Table 2, all extracted gyri have connectivity strength stronger than all other sulci.
Fig. 11. Heat maps for the connectivity of gyri and sulci from the 1019 injection sites. (Top) Voxel-level connectivity where the upper and lower rows correspond to gyri and sulci respectively. Columns represent gyrus/sulcus connectivity from the 1019 injection sites. (Bottom) The average connectivity across the voxels of individual gyrus and sulcus from the 1019 injection sites. The upper 12 rows represent 12 gyri and lower 14 rows represent 14 sulci. The color in both A and B indicates the connection strength. Note that the data displayed here are the logarithmic values of the raw connectivity data for better visualization.
Pairwise comparison of connectivity strength between gyri and sulci. Each value in the table denotes the percentage that a gyrus (column) exceeds a sulcus (row) in pair-wise comparison of connectivity strength to each injection site for a total of 1019 sites. Each column represents a gyrus (abbrev name is given on the corresponding column of top row) against 14 sulci, and each row represents a sulcus (abbrev name is given on the corresponding row of left most column) against 12 gyri. The bottom row denotes the average for each column and the rightmost column denotes the average for each row. The names of gyri are the abbreviation of structure names based on the brain ontology, and the names of sulci are given by joining the two abbreviations of structure names, since its voxels belong to two structures. The names of sulci with SU after - denote that they belong to only one brain structure.

<table>
<thead>
<tr>
<th>E(G)&gt; E(S)%</th>
<th>AN1</th>
<th>AN2</th>
<th>CENT2</th>
<th>CENT3</th>
<th>COPY</th>
<th>CUL</th>
<th>DEC</th>
<th>FL</th>
<th>NOD</th>
<th>PRM</th>
<th>SLM</th>
<th>UVU</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN1-AN2</td>
<td>73.9</td>
<td>65.2</td>
<td>61.1</td>
<td>40.4</td>
<td>80.5</td>
<td>77.6</td>
<td>65.0</td>
<td>71.9</td>
<td>75.5</td>
<td>79.5</td>
<td>68.0</td>
<td>69.3</td>
<td>69.0</td>
</tr>
<tr>
<td>AN2-PRM</td>
<td>73.0</td>
<td>66.3</td>
<td>56.77</td>
<td>38.6</td>
<td>82.6</td>
<td>74.0</td>
<td>63.4</td>
<td>69.7</td>
<td>72.2</td>
<td>82.4</td>
<td>64.0</td>
<td>61.8</td>
<td>67.1</td>
</tr>
<tr>
<td>CENT2-CENT3</td>
<td>87.0</td>
<td>81.2</td>
<td>81.5</td>
<td>63.9</td>
<td>88.4</td>
<td>88.1</td>
<td>80.5</td>
<td>84.8</td>
<td>87.1</td>
<td>87.4</td>
<td>84.0</td>
<td>79.9</td>
<td>82.8</td>
</tr>
<tr>
<td>CENT2-SU</td>
<td>55.2</td>
<td>50.5</td>
<td>45.3</td>
<td>30.4</td>
<td>73.3</td>
<td>59.9</td>
<td>48.8</td>
<td>57.2</td>
<td>60.9</td>
<td>68.6</td>
<td>48.2</td>
<td>53.0</td>
<td>54.3</td>
</tr>
<tr>
<td>CENT3-CUL</td>
<td>60.8</td>
<td>55.0</td>
<td>49.9</td>
<td>26.5</td>
<td>77.4</td>
<td>68.0</td>
<td>52.4</td>
<td>62.1</td>
<td>66.8</td>
<td>75.3</td>
<td>52.4</td>
<td>54.8</td>
<td>58.5</td>
</tr>
<tr>
<td>CUL-DEC</td>
<td>81.7</td>
<td>76.2</td>
<td>74.6</td>
<td>63.7</td>
<td>83.3</td>
<td>84.5</td>
<td>74.0</td>
<td>82.0</td>
<td>80.2</td>
<td>82.7</td>
<td>80.2</td>
<td>75.7</td>
<td>78.2</td>
</tr>
<tr>
<td>CUL-SIM</td>
<td>65.4</td>
<td>56.4</td>
<td>52.5</td>
<td>32.5</td>
<td>78.4</td>
<td>73.1</td>
<td>58.0</td>
<td>65.9</td>
<td>70.2</td>
<td>75.6</td>
<td>59.7</td>
<td>58.4</td>
<td>62.2</td>
</tr>
<tr>
<td>NOD-LING</td>
<td>74.9</td>
<td>70.1</td>
<td>68.4</td>
<td>50.7</td>
<td>84.1</td>
<td>78.0</td>
<td>69.0</td>
<td>75.5</td>
<td>79.8</td>
<td>82.8</td>
<td>71.0</td>
<td>69.6</td>
<td>72.8</td>
</tr>
<tr>
<td>PRM-COPY</td>
<td>86.8</td>
<td>81.8</td>
<td>81.2</td>
<td>69.5</td>
<td>89.0</td>
<td>88.3</td>
<td>81.2</td>
<td>85.2</td>
<td>85.4</td>
<td>86.7</td>
<td>83.5</td>
<td>81.1</td>
<td>83.3</td>
</tr>
<tr>
<td>PYR-UVU</td>
<td>87.3</td>
<td>83.8</td>
<td>82.5</td>
<td>74.5</td>
<td>89.0</td>
<td>91.1</td>
<td>81.2</td>
<td>85.3</td>
<td>86.8</td>
<td>88.0</td>
<td>86.5</td>
<td>84.6</td>
<td>85.0</td>
</tr>
<tr>
<td>SIM-AN1</td>
<td>77.9</td>
<td>71.7</td>
<td>69.6</td>
<td>56.9</td>
<td>81.7</td>
<td>78.8</td>
<td>68.9</td>
<td>77.9</td>
<td>78.2</td>
<td>80.5</td>
<td>74.8</td>
<td>71.1</td>
<td>74.0</td>
</tr>
<tr>
<td>UVU-NOD</td>
<td>86.4</td>
<td>78.9</td>
<td>76.0</td>
<td>62.8</td>
<td>89.2</td>
<td>89.0</td>
<td>80.3</td>
<td>84.1</td>
<td>85.5</td>
<td>89.3</td>
<td>84.3</td>
<td>79.7</td>
<td>82.1</td>
</tr>
<tr>
<td>CENT3-SU</td>
<td>81.5</td>
<td>75.8</td>
<td>75.8</td>
<td>64.2</td>
<td>85.8</td>
<td>83.9</td>
<td>76.7</td>
<td>80.2</td>
<td>83.9</td>
<td>85.7</td>
<td>78.2</td>
<td>76.9</td>
<td>79.0</td>
</tr>
<tr>
<td>FOTU-PYR</td>
<td>91.6</td>
<td>88.4</td>
<td>89.9</td>
<td>85.5</td>
<td>90.6</td>
<td>92.9</td>
<td>87.2</td>
<td>89.7</td>
<td>89.5</td>
<td>89.7</td>
<td>90.1</td>
<td>88.4</td>
<td>89.4</td>
</tr>
<tr>
<td>Average</td>
<td>E(G)&gt;E(S)%</td>
<td>77.4</td>
<td>71.5</td>
<td>68.9</td>
<td>54.3</td>
<td>83.8</td>
<td>80.5</td>
<td>70.5</td>
<td>76.5</td>
<td>78.7</td>
<td>82.4</td>
<td>73.2</td>
<td>71.7</td>
</tr>
</tbody>
</table>
4.2 GYRI AND SULCI HAVE DIFFERENT CONNECTIVITY PATTERNS

Motivated by the above results that gyri have broader and stronger connectivity than sulci, I carried out a connectivity pattern correlation analysis to obtain further insights on how gyri and sulci are connected in the mouse brain. Based on the axon pushing hypothesis [7], [8], it is believed that gyri would have long-range neuronal projections connecting to different regions of brain, thereby performing very different neuronal functions as compared to sulci. Thus, voxels in gyri are expected to have very diverse connectivity patterns. In comparison, sulci are considered as local information processing units [9] and are usually connected to fewer number of regions, thereby exhibiting similar connectivity patterns. Here I tested this hypothesis using the ACA data. I computed the Pearson correlation coefficient between the connectivity vector data at both voxel and gyrus-sulcus levels. The results are shown in Figure 12. At the voxel level, sulcus-to-sulcus comparisons demonstrate striking similarity in terms of connectivity pattern correlation. Moreover, they show higher correlation with negative voxels than with gyrus voxels. Conversely, gyrus-gyrus correlations indeed showed that gyri have distinctive connectivity patterns, since the correlations within different gyri are much lower than those within different sulci. At the gyrus-sulcus level, the gyrus-gyrus and sulcus-sulcus connectivity correlations are higher than gyrus-sulcus correlations. This indicates that gyri and sulci, as separate groups, have different neuronal connectivity patterns to the rest of the brain regions. These results are consistent with the hypothesis that different gyri are connected to different regions of brain while sulci are connected locally.
Fig. 12. The correlations of connectivity for gyri, sulci and negative regions. Left: Voxel-level connectivity correlation between gyri, sulci and negative regions. Right: Gyrus-sulcus level connectivity correlation, where data for each individual gyrus or sulcus were averaged across all corresponding voxels. The connectivity data used for the correlation in the figures were the logarithmic values of the raw data.
4.3 GYRI AND SULCI EXHIBIT DIFFERENT GENE EXPRESSION PATTERNS

Given the results that gyri differ with sulci in terms of neuronal connectivity, it can be expected that gene expressions in gyri and sulci would have correlation patterns that are consistent with those of neuronal connectivity, since connectivity and gene expressions were found to be correlated [32]. Using the same approach in extracting connectivity data from the ACA, gene expression data of voxels in gyri and sulci from the ABA was obtained. Figures 13 shows the Pearson’s correlation of gene expression between gyri, sulci and negative samples at voxel level and gyrus-sulcus level, respectively. The resulting sulci-sulci correlations are the highest among all group pairs. Additionally, sulci show a higher correlation with negative samples in comparison with gyri-negative correlation. In contrast, gyri tend to have different gene expression patterns that are more similar to neighboring sulci than among each other. This interesting property of gyri indicates that they may perform different functional tasks corresponding to the different afferent and efferent neuronal projections they have, therefore exhibiting different neural connectivity patterns. This is consistent with gyri’s demonstrated property of having denser axonal fiber connections than sulci, as reported in neuroimaging experiments [7], [8], and are thus believed to play an important role in large-scale signal processing [9]. Conversely, sulci usually connect to fewer axonal fibers than gyri and are expected to play a role in local signal processing [9]. Thus they may have similar connectivity and gene expression patterns. In addition to their role in long-range signal processing, gyri may also pass signals to local regions, which in this case are sulci. Such a neuronal signal processing architecture suggests that gyri tend to be connected to nearby sulci. These distinctive gene expression patterns of gyri and sulci lead to conclusions that are consistent with previous findings in neuroimaging experiments.
Fig. 13. The correlations of gene expression for gyri, sulci and negative regions. Left: Voxel-level gene expression correlations between gyri, sulci and negative regions. Right: Gyrus-sulcus level gene expression correlation, where data for each individual gyrus or sulcus were averaged across all corresponding voxels.
4.4 GYROS, SULCUS AND NEGATIVE VOXELS CAN BE ACCURATELY CLASSIFIED BASED ON THEIR GENE EXPRESSION PATTERNS

Since aforementioned analysis has shown that gyrus, sulcus and negative voxels possess different gene expression patterns, one interesting question is whether such differences would allow predicting the voxels into the correct category based on their gene expression patterns. This hypothesis was tested by predicting the label (gyrus, sulcus, or negative) of each voxel based only on its gene expression pattern using pattern classification algorithms. In addition, feature selection was performed using sparse methods to obtain the top genes that contribute most to the prediction assay. To ensure the validity of the top genes, the prediction based only on the selected genes was recalculated. In this experimental approach, 2/3 of the randomly sampled data were used to construct the model. A number of sparse models were constructed for a wide range of λ values ranging between 0.0005 and 1 to induce different levels of sparsity. The remaining 1/3 of the samples were used to compute the accuracy of the model.

The AUC and accuracy values obtained using the top 25, 50, 100, 200, 500 and all 4084 genes are reported in Table 3. The results indicate that the prediction performance of gyri against negative has a minimum accuracy value of 82% and a minimum AUC of 87%. Sulcus against negative has yielded a minimum accuracy of 84% and a minimum AUC of 86%. The highest amongst all the classification tasks was the gyri against sulci, which had a minimum accuracy of 92% and an AUC of 97%. This high prediction performance demonstrated that gyri and sulci can be clearly classified based on their gene expression patterns. It also can be observed that the task of recognizing gyri and sulci from negative samples yielded accuracies that were constantly above 80%. Moreover, using a small subset of the genes, very high prediction accuracies can be obtained as well. This suggests that these small number of genes act as marker genes that distinguish gyri, sulci, and other voxels.

To compare aforementioned results with random predictions, an experiment was performed that ran prediction 100 times by selecting the training and test voxels at random each time. In addition, a similar experiment was ran for another 100 times but randomly shuffled the labels this time to ensure that the high accuracy previously obtained was not by chance. The average AUC and standard deviation are listed in Table 4. It can be observed that the average performance obtained by shuffling
The prediction AUC and accuracy for the three classification problems: gyrus against sulcus, gyrus against negative, and sulcus against negative. The values reported are generated using different numbers of genes ranging between 25 to all of the 4084 genes

<table>
<thead>
<tr>
<th></th>
<th># of genes</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gyrus/sulcus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>97.55</td>
<td>99.02</td>
<td>99.29</td>
<td>99.55</td>
<td>99.72</td>
<td>99.90</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>92.61</td>
<td>95.80</td>
<td>96.41</td>
<td>96.56</td>
<td>97.25</td>
<td>98.33</td>
<td></td>
</tr>
<tr>
<td><strong>Gyrus/negative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>87.05</td>
<td>88.34</td>
<td>89.65</td>
<td>92.16</td>
<td>93.73</td>
<td>96.35</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>82.00</td>
<td>83.76</td>
<td>85.01</td>
<td>86.87</td>
<td>87.91</td>
<td>90.59</td>
<td></td>
</tr>
<tr>
<td><strong>Sulcus/negative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>86.34</td>
<td>89.04</td>
<td>90.78</td>
<td>93.85</td>
<td>95.90</td>
<td>97.55</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>84.03</td>
<td>85.35</td>
<td>86.69</td>
<td>88.99</td>
<td>91.22</td>
<td>93.00</td>
<td></td>
</tr>
</tbody>
</table>

The labels is very low. This assures that the high prediction accuracy obtained in aforementioned expriments, indicating that the gyri, sulci and negative voxels indeed have significantly different expression patterns.

The prediction AUC and accuracy for the three classification problems: gyrus against sulcus, gyrus against negative, and sulcus against negative. The values reported are generated using different numbers of genes ranging between 25 to all of the 4084 genes

<table>
<thead>
<tr>
<th></th>
<th>Without label shuffling</th>
<th>With label shuffling</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gyrus/sulcus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>99.88 ± 0.04</td>
<td>50.43 ± 1.92</td>
</tr>
<tr>
<td>Accuracy</td>
<td>98.47 ± 0.23</td>
<td>51.68 ± 1.46</td>
</tr>
<tr>
<td><strong>Gyrus/negative</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>96.56 ± 0.21</td>
<td>49.88 ± 1.81</td>
</tr>
<tr>
<td>Accuracy</td>
<td>91.04 ± 0.35</td>
<td>69.76 ± 1.47</td>
</tr>
<tr>
<td><strong>Sulcus/negative</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>97.20 ± 0.20</td>
<td>51.47 ± 2.03</td>
</tr>
<tr>
<td>Accuracy</td>
<td>92.48 ± 0.36</td>
<td>76.36 ± 1.35</td>
</tr>
</tbody>
</table>
CHAPTER 5

CONCLUSION AND FUTURE WORK

One of the most prominent attributes in the primate cortex is its convoluted neuronal structures of gyri and sulci. Recent studies using DTI, dMRI, and HARDI data have showed that gyri differ significantly from sulci in terms of axonal fiber connections, suggesting that they may play different roles in neuronal information processing. However, these current results are limited to macro-scale data analysis. Complete understanding of the biological implications underlying gyri and sulci formation and function requires cellular and molecular level experiments and analyses. To the best of my knowledge, this work represents the first attempt to study the gyrus and sulcus differences at cellular and molecular levels using both neuronal connectivity and gene expression data. The work in this thesis showed that (1) gyri possess overall stronger and broader connections than those of sulci, (2) gyri significantly differ from sulci in both connectivity and gene expression patterns, (3) gene expression in gyri reveals that they tend to connect to adjacent sulci instead of other gyri, (4) high prediction accuracy can be achieved in classifying gyri, sulci, and other voxels, and the main predictive power is contributed by a small number of genes. This thesis primarily focused on predictive and correlative analysis of gene expression and connectivity patterns. In the future, I will perform in-depth analysis of the results generated by this study. In particular, I will make use of existing knowledge compiled into databases, such as the gene ontology, to investigate the functions of genes that distinguish gyri from sulci. The data used in this study were generated by single injection anterograde tracers. Double co-injection tract tracing techniques allow the identification of both input and output of projections [33]. Such types of data have been made available recently at high-throughput scale [34]. I will extend current study to these types of data in the future. Finally, it should be pointed out that the Allen Mouse Brain Connectivity Atlas is an ongoing project that constantly integrates new imaging and mapping technologies, and additional data are released regularly. These data will be considered and examined in future studies.
REFERENCES


VITA

Tao Zeng
Department of Computer Science
Old Dominion University
Norfolk, VA 23529

Tao Zeng received his Master degree in neuroscience from Chinese Academy of Sciences, China in 2008. Afterwards, he worked at Eastern Virginia Medical School as a research fellow for 4 years. In Summer 2012, he joined in Computer Science Department of Old Dominion University and started his research in machine learning and computational neuroscience under the supervision of Professor Dr. Shuiwang Ji. His research regarding gene expression and connectivity patterns of convolutional structures in the brain has been published in a peer-reviewed journal.