# DEVELOPMENTAL DYNAMICS OF THE HUMAN BRAIN TRANSCRIPTOME

# DEVELOPMENTAL DYNAMICS OF THE HUMAN BRAIN TRANSCRIPTOME

By

KEON ARBABI, B.Sc.

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for the Degree

Master of Science

McMaster University

© Copyright by Keon Arbabi, December 2020

MASTER OF SCIENCE (2020) McMaster University

(Neuroscience) Hamilton, Ontario

TITLE: Developmental Dynamics of the Human Brain Transcriptome

AUTHOR: Keon Arbabi

SUPERVISOR: Dr. Kathryn M. Murphy

NUMBER OF PAGES: 146

## Abstract

Large-scale transcriptomic studies are among of the most comprehensive accounts we have of the biological processes underlying human brain development and ageing. However, many analyses and descriptive models applied to gene expression data implicitly assume that developmental change is continuous and uninterrupted. Perhaps this bias is often overlooked because the emphasis is on *what* is changing during development rather than *how* development itself is changing. Indeed, despite the richness of transcriptomic data and its capacity to recapitulate higher-order functions, few have used it to understand the dynamics of brain development. Gene expression is determined by complex, high-dimensional interactions of the gene regulatory network. Dynamic systems theory states that the interactions of components in any complex systems will converge on certain stable patterns, also known as attractor states. To approximate these stable states, the current study leveraged robust and sparse k-means clustering to identify tissue samples with similar patterns of gene expression across the transcriptome. Sample ages were then used to visualize when in developmental time these stable patterns are present. The resulting model describes the developmental dynamics of the brain transcriptome as a series of non-linear, overlapping states that progress across the lifespan.

## Table of Contents


**1. Chapter 1: General Introduction……………………………………………………….....1**

1.1. Dynamics of development……………………………………………………………...2

1.1.1. Traditional developmental theories……………………………………………..2

1.1.2. Dynamic systems theory………………………………………………………..4

1.1.3. Dynamic systems in motor development……………………………………….6

1.1.4. Dynamic systems in functional brain imaging………………………………….7

1.1.5. Dynamic systems in cellular differentiation and gene expression……….……....8

1.1.6. Summary and extension to critical periods……………………………………10

1.2. Developmental transcriptomics of the human brain……………………………….12

1.2.1. Contributions of recent developmental transcriptome studies…………………...13

1.2.2. Potential gaps in development transcriptome studies……………………………14

1.3. The current study……………………………………………………………………...17

1.4. Study rationale, objective, and specific aims……………………………………………19

**2. Chapter 2: Introduction to Clustering…………………………….…………….…….....21**

2.1. Introduction……………………………………………………………………………...23

2.2. Robust and spare k-mean clustering ………………………………………………………23

2.3. Evaluation of clustering methods………………………………………………………..24

2.4. Selection of *k*………………………………………………………………………………29

2.5. Minimizing biases…………………………………………………………………………30

2.6. Summary………………………………………………………………………………...31

**3. Chapter 3: Developmental Clusters in Primary Visual Cortex………………………......32**

3.1. Introduction……….……………………………………………………………………..33

3.2. Methods……………………………………………………………………………………33

3.2.1. Data acquisition and processing………………………………………………….33

3.2.2. Analysis of neuroanatomy………………………………………………………..34

3.2.3. Differential gene expression analysis…………………………………………….34

3.2.4. Cell-type analysis……………………………………………………………...35

3.3. Results…………………………………………………………………………………...36

3.3.1. A small selection of genes separates prenatal and postnatal samples…………..…36

# Table of Contents


3.3.2. Developmental clusters in primary visual cortex……38

3.3.3. Further investigation of feature genes: *VAMP1*……45

3.3.4. Differential gene expression analysis……47

3.3.5. Cell-type analysis……50

3.4. Summary……54

4. **Chapter 4: Developmental Clusters in Multiple Brain Regions……56**

4.1. Introduction……57

4.2. Methods……58

4.3. Results……60

4.3.1. Developmental clustering within brain regions……60

4.3.2. Comparing developmental features across brain regions……62

4.3.3. Clustering of brain regions……67

4.3.4. Developmental clustering within brain regions……69

4.4. Summary……77

5. **Chapter 5: Clustering Bipolar Cases with Developmental Controls……79**

5.1. Introduction……80

5.2. Methods……81

5.2.1. Data acquisition……81

5.2.2. *In silico* data merging……82

5.3. Results……84

5.3.1. Clustering control and BD samples……84

5.3.2. Comparing BD transcriptional subgroups……88

5.3.3. Clustering control and BD samples using curated gene lists……95

5.4. Limitation and future directions……95

5.5. Summary……96

6. **Chapter 5: General Discussion……97**

6.1. Summary of Main Findings……98

6.1.1. Developmental clusters in primary visual cortex……98

6.1.2. Developmental clusters in multiple brain regions……98

# Table of Contents

6.1.3. Clustering bipolar cases and developmental controls……………………………….99

6.2. Significance……………………………………………………………………100

6.2.1. Unsupervised clustering identifies stable patterns of gene expression………….100

6.2.2. Overlapping patterns of gene expression………………………………………..101

6.2.3. Biological significance of features……………………………………………...102

6.3. Limitations……………………………………………………………………..103

6.4. Future directions………………………………………………………………104

**7. Chapter 7: References……………………………………………………..……106**

## List of Figures and Tables

**Figure 1.** *Internal validation of cluster algorithms*……………………………………………25

**Figure 2.** *Stability validation of clustering algorithms*…………………………………………..26

**Figure 3.** *Clustering validation comparison with RSKC*……………………………………...28

**Figure 4.** *Selection of k*……………………………………………………………………...30

**Figure 5.** *Clustering and feature selection of prenatal and postnatal clusters.*……………………36

**Figure 6.** *Developmental trajectories for top feature genes*……………………………………..37

**Figure 7.** *Developmental clusters in V1C*……………………………………………………..39

**Figure 8.** *Sample sex and hemisphere by cluster*……………………………………………...40

**Figure 9.** *Top feature genes for postnatal clusters in V1C*……………………………………41

**Figure 10.** *Developmental expression of top postnatal feature genes*…………………………...42

**Figure 11.** *In situ hybridization of VAMP1 in visual cortex*………………………………......46

**Figure 12.** *Sum of differentially expression genes per cluster*…………………………………...48

**Figure 13.** *Differential gene expression analysis for cluster 2A*………………………………49

**Figure 14.** *Cell-type analysis of differentially expressed genes per cluster*……………………..51

**Figure 15.** *Developmental clusters across distinct brain regions*……………………………….61

**Figure 16.** *Median age of developmental clusters compared across brain regions*……………..62

**Figure 17.** *Pre- and postnatal developmental feature genes across brain regions*……………...64

**Figure 18.** *Postnatal developmental feature genes across brain regions*………………………..66

**Figure 19.** *Gene network for developmental and region features*………………………………70

**Figure 20.** *Developmental clustering of samples in region clusters*……………………………..71

**Figure 21.** *Sample brain region and hemisphere per cluster*……………………………………72

**Figure 22.** *Sample sex per cluster*………………………………………………………..…73

**Figure 23.** *Enriched GO biological functions in clusters with sex differences*
*in DFC, MFC ,OFC*…………………………………………………………………………...75

**Figure 24.** *Enriched GO biological functions in clusters with sex differences*
*in M1C, S1C, VFC.*…………………………………………………………………………...76

## List of Figures and Tables

**Figure 25.** *Data merging*……………………………………………………………………83

**Figure 26.** *Clustering of control and BD samples*……………………………………………85

**Figure 27.** *Expression of BD markers*………………………………………………………86

**Figure 28.** *Age of control and BD samples in each cluster*…………………………………...87

**Figure 29.** *Enriched biological processes in cluster 2B*………………………………………90

**Figure 30.** *Enriched biological processes in cluster 2C and 2D*……………………………...93

**Figure 31.** *Clustering control and BD samples using curated gene lists*……………………..94

**Figure 32.** *PHATE analysis of developmental samples*………………………………………105

**Table 1.** *Regions of region clustering*…………………………………………………………69

**Table 2.** *Subject and clinical variables for BD cases*…………………………………………89

## List of Abbreviations

| | |
|---|---|
| A1C | Primary auditory cortex |
| AA | Accelerated ageing |
| AD | Alzheimer's disease |
| ASD | Autism spectrum disorder |
| AMPA | α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid |
| AMY | Amygdala |
| ANOVA | Analysis of variance |
| BD | Bipolar disorder |
| BDNF | Brain-derived trophic factor |
| BOLD | Blood-oxygen-level-dependent |
| CBC | Cerebellar cortex |
| CNS | Central nervous system |
| DE | Differential expression |
| DFC | Dorsolateral prefrontal cortex |
| DIANA | Divisive clustering |
| DNA | Deoxyribonucleic acid |
| DST | Dynamic systems theory |
| ECM | Extracellular matrix |
| FANNY | Fuzzy clustering |
| FDR | False discovery rate |
| fMRI | Functional magnetic resonance imaging |
| GABA | γ-aminobutyric acid |
| GEO | Gene Expression Omnibus |
| GM | Grey matter |
| GO | Gene Ontology |
| GRN | Gene regulatory network |
| HIP | Hippocampus |
| IDE | Insulin-degrading enzyme |
| IF | Intermediate filament |
| IPC | Posterior inferior parietal cortex |
| ITC | Inferior temporal cortex |
| LOESS | Locally estimated scatterplot smoothing |
| LTD | Long-term depression |
| LTP | Long-term potentiation |
| M1C | Primary motor cortex |
| MD | Mediodorsal nucleus of the thalamus |
| MD | Monocular deprivation |
| ME | Monocular enucleation |

## List of Abbreviations

| | |
|---|---|
| MFC | Medial prefrontal cortex |
| MRI | Magnetic resonance imaging |
| NCX | Neocortex |
| NIH | National Institute of Health |
| NMDA | N-Methyl-D-aspartic acid |
| NPC | Neuronal progenitor cell |
| OD | Ocular dominance |
| OFC | Orbital prefrontal cortex |
| PAM | Partitioning around medoids |
| PCA | Principal component analysis |
| PCW | Post-conception weeks |
| pH | Potential of hydrogen |
| PNN | Perineuronal net |
| PV | Parvalbumin |
| RBP | RNA-binding proteins |
| RNA | Ribonucleic acid |
| ROS | Reactive oxygen species |
| RSKC | Robust and sparse k-means clustering |
| RSN | Resting state network |
| S1C | Primary somatosensory cortex |
| SNARE | Soluble N-ethylmaleimide-sensitive factor attachment receptor |
| SOM | Self-organizing maps |
| STC | Posterior superior temporal cortex |
| STR | Striatum |
| STT | Somatostatin |
| SVZ | Subventricular zone |
| tSNE | T-distributed stochastic neighbour embedding |
| TTX | Tetrodotoxin |
| V1C | Primary visual cortex |
| VFC | Ventrolateral prefrontal cortex |
| WBSS | Weighted between sum of squares |
| WGCNA | Weighted gene co-expression network analysis |

# Chapter 1. General Introduction

## 1.1 Dynamics of development

Stages and states have served as essential heuristics in developmental science since its inception. In general, a stage or state describes a discrete period of time characterized by certain psychological or physical abilities. Stage-like processes have also been described in neurobiological studies of brain development. For example, the *critical period* refers to a time-restricted window of enhanced plasticity during which experience is essential for normal development (Hensch, 2005). Animal models have provided clear molecular, cellular, and structural evidence of critical periods in the primary sensory systems of several species (Berardi, Pizzorusso, & Maffei, 2000). Focused sequencing stuides in these systems demonstrate that the critical period is driven by unique and transient patterns of gene expression (Lyckman et al., 2008). When gene expression is measured on a larger scale, however, distinct stages are rarely considered. Modern transcriptomic stuides include hundreds of tissue samples spanning numerous brain areas and developmental periods. Thus, gene expression data is preferentially analyzed with methods that assume *continuous change* across the lifespan. This conceptual bias is often overlooked because the emphasis is on *what* is changing during development rather than *how* development itself is changing.

In the current section, I review traditional theories of developmental dynamics that culminate in modern dynamic systems theory. In essence, dynamic systems theory posits that developmental trajectories are nonlinear and converge on stable "attractor states" over time. I demonstrate the utility of this descriptive model at various levels of analysis, including motor development, functional brain networks, and cellular differentiation. Accordingly, I propose that transcriptomic stuides may capture the developmental dynamics of the human brain and discuss the potential benefit of approaching this data with states and nonlinearity in mind.

### 1.1.1 Traditional developmental theories

*Development* is derived from the Old French word *desvoloper*, which means "to unwrap" or "to unfold." (van Geert, 2003). In this sense, developmental psychologists and neurobiologists are chiefly concerned with how the brain and the behaviours it produces unfold over time. There has been much debate, particularly in developmental psychology, regarding the *dynamics* of this process. That is, what is the *shape* of developmental change? Is it best modeled by a continuous

curve, or are there discontinuities representing spontaneous re-organization of structure and behaviour (Flavell, 1971)? Returning to the etymological roots of the word, let us imagine development as the process of unfolding a folded piece of paper. This metaphor illustrates some fundamental aspects of traditional views of development. First, the process consists of a finite number of steps. Second, the nature of transformation is determined by the initial state of the folded paper. Third, all the states are qualitatively different (i.e. forms of the paper). And lastly, the process of transformation is the same for all states, namely unfolding.

Jean Piaget (1896 – 1980) noticed that children demonstrate large-scale shifts in behaviour elicited by cognitive tasks, which are strongly correlated with age. Piaget used these stable individual differences to describe four stages of cognitive development: (I) sensorimotor, (II) preoperational, (II) concrete operational, and (IV) formal operational. Stages were defined by qualitative differences across a number of fundamental concepts such as quantity, time, causality, and relationships (Piaget & Cook, 1952, Piaget & Inhelder, 1969). Piaget regarded cognitive development as a process of biological maturation and interaction with the environment. In this view, children come to understand their surrounding by assimilating new information into knowledge structures known as *schemas*, which are updated through accommodation. When a child's internal schemas can no longer accommodate new information, the current state dissolves into disequilibrium before a new state that is sufficiently resistant to further change is achieved (Feldman, 2004). Piaget's theory fits well with our folded paper metaphor; it describes cognitive development as a finite, invariant sequence of states, where the distinction between states is qualitative and determined by the properties of the previous step, and transition is driven by consistent forces (i.e. assimilation, accommodation, and equilibration). Piaget's theories of development have been criticized for a lack of explanatory power and have lost considerable support over the years (Fischer & Silvern, 1985; Brainerd, 1987). Despite this, Piaget's interpretation of the mind as a dynamic biological organ that adapts the individual to the environment, and the environment to the individual, involved concepts that would become extremely popular and well-regarded.

The 1950s and 1960s witnessed the birth of many important ideas regarding the dynamics of development. British embryologist Conrad Hal Waddington (1905 – 1975) was interested in how complex and distinctive tissues – bones, muscle, lungs, and so on – could emerge from a

homogenous single cell. He opposed the widespread preformist view that the adult form of an organism is entirely coded for by genes. Instead, he believed that an animal is "a developmental system," wherein genes establish the starting point of development and the process of development itself actively determines the animal's final form (Waddington, 1957; Griffiths & Tabery, 2013). In this view, genes code for the production of certain tissues that, once formed, can influence the expression of genes and formation of other tissues (Slack, 2002). These ideas are illustrated by Waddington's famous *epigenetic landscape.* Taking our piece of paper from earlier and unfolding it completely, we notice that the creases form hills and valleys. Placing a marble on this surface causes it to move towards valleys and away from hills. In this metaphor, the marble represents the organism in a space defined by developmental dimensions or possibilities. The marble can start anywhere on the landscape, but development creates hills and valleys of increasing complexity and depth. Over time, the organism will fall towards states of differentiation and stability. Importantly, movement of the marble simultaneously warps the landscape, so it is possible to arrive at the same valley through various paths. This bi-directional and probabilistic nature of epigenesis was later refined by Gilbert Gottlieb (1929 – 2006), who highlighted the influence of environmental and physiological factors (Griffiths & Tabery, 2013). In sum, Waddington and Gottlieb described a process by which an organism becomes increasingly differentiated and structured via a bi-directional relationship between genes, the environment, and the organism itself.

### 1.1.2 Dynamics systems theory

Dynamic systems theory (DST) provides a contemporary theoretical approach to studying development that continues traditions started by Piaget and Waddington. Systems theory has its roots in mathematics, physics, and chemistry. In the 1970s, Belgian chemist Ilya Prigogine (1917 – 2003) became interested in chemical reactions that appeared the violate the second law of thermodynamics; they would spontaneously create complex spatial and temporal patterns from less ordered chemical reactants. In these "autocatalytic" reactions, reactants produce a chemical compound that facilitates the formation of another product, which in turn has varying facilitatory or inhibitory effects on other compounds in the reaction. Provided with enough heat, this reaction oscillates violently between spontaneous chemical states (Agladze, Krinsky, & Pertsov, 1984). Prigogine noted that such reactions share important properties with biological systems: First,

they are thermodynamically open and able to take energy from their environment to increase their own order. Second, they live in thermodynamic nonequilibrium, where constant fluctuations can become amplified and shift the system into a new state (Prigogine & Nicolis, 1971). Developmentalists immediately recognized the relevance of dynamic system principles to age-old questions of self-organization and changing states in development (Thelen & Smith, 1994).

By their very nature, dynamic systems are high-dimensional; comprised of countless interacting elements spanning multiple, often heterogeneous, levels (e.g. genes, neurons, circuits, and behaviour). Despite the immeasurable degrees of freedom, the interactions of system components converge on certain "preferred" spatial and temporal patterns (Thelen & Smith, 2007). In this sense, dynamic systems are self-organizing because they spontaneously create "structure" through self-sustaining patterns of interactions (van Geert, 2020). Some terminology will aid further discussion. The *state* of a system is a snapshot of its ongoing processes and is defined by the current values of the *order parameters* being used to characterize the system. A set of parameters that the system actively reproduces over time is called an *attractor state.* The concept of an attractor state is not unlike a valley in Waddington's epigenetic landscape, which is a resting point that developmental trajectories in the space converge on. A stable attractor state can be visualized as a marble lying at the bottom of a steep valley – it would require a lot of energy to send over the top, and if perturbed, it will quickly return to the bottom. A series of neighbouring, shallow valleys would facilitate *metastability,* where the system may switch from one attractor to another (Thelen & Smith, 2007). Metastable states are particularly interesting the context of development, and we will revisit them later. The stability of a state can be theoretically indexed by the statistical probability of that state existing relative to other potential configurations and by its resistance to external perturbations. Stability is also related to the system's response to internal fluctuations. Recall that stability is not merely the default state of a system, but a highly specific pattern of interactions that the system expends energy to reproduce (van Geert, 2020). These interactions carry *intrinsic noise* that act as stochastic forces on the stability of a state. Crucially, intrinsic noise can provide potential energy for shifting the system into a new state over time. These ideas come together to describe development as a fundamentally dynamic process, where the interaction of countless elements organically produces stable states that ultimately define function and behaviour.

Stability and change are two sides of the same coin. We have already discussed how dynamic systems produce stability, but how do they create progressive patterns of change – as seen in development? A system is stable when components cooperate and form a stable pattern of interactions that resists internal and external fluctuations. However, if the intrinsic noise from component sub-systems increases, or external pressures on the system change, then the old pattern of interactions becomes incoherent (Thelen & Smith, 2007). This describes *criticality*, where any minor perturbation can cause a cascade of changes that shifts the system into a qualitatively new pattern over a short period of time (van Geert, 2020). From a developmental perspective, the discontinuities in function and behaviour caused by these shifts may give the impression that a new stage has emerged, as Piaget had hypothesized. Nevertheless, several studies have demonstrated that these changes are as not global as Piaget had thought and are likely confined to specific developmental domains (Fischer & Bullock, 1981; Collins, 1984). Moreover, not all changes in a system are discontinuous shifts. Some parameters may change in a linear and continuous manner. Studying early language development, van Dijk and van Geert (2007) argue that continuity and discontinuity are not necessarily dichotomous but can be conceived as two extremes of one continuum. For example, the sudden emergence of a new grammatical structure that can be characterized as a discontinuity does not imply that the former grammatical behaviour completely disappears. The two behaviour modes co-exist for some time, and the replacement of the old mode by the new may occur in a linear, continuous fashion. This transition will likely differ between individuals, resulting in patterns that are more discontinuous in some individuals and more continuous in others (van Dijk & van Geert, 2007).

### 1.1.3 Dynamic systems in motor development

Dynamic systems theory has made major contributions to developmental psychology over the past 20 years (Spencer, Austin, & Schutte, 2012). In particular, the field of infant motor development has witnessed a revitalization due to the introduction of dynamic systems concepts and tools (Newell, Liu, & Mayer-Kress, 2003). In the traditional view, the development of motor actions occurs in a series of relatively fixed milestones that are mostly under genetic or biological control (Adolph & Berger, 2006). This model struggles to explain qualitative changes in behaviour over multiple time scales. For example, an infant can go from crawling to walking in the matter of seconds, but how does this transition relate to the more gradual shift in time

spent crawling verses walking that occurs over months (Spencer, Perone, & Buss, 2011)? In the dynamic systems perspective, behaviour arises from multiple components interacting and combining freely from moment to moment; like musicians playing jazz (Thelen & Smith, 1994). The resulting behavioural attractors are a product of the current context, task, and developmental history of the individual, and certain attractors become more likely (stable) over developmental time. In detail, crawling is a behaviour that infants use to move around when they have the strength and coordination to use their arms and knees, but cannot yet balance themselves or walk up-right. When infants learn to walk, the crawling pattern becomes destabilized by the pattern of bipedal locomotion. Contrary to the traditional theory of motor development, crawling is not hard coded for by genes or the nervous system. It self-organizes as a task-specific solution, and is later replaced by a more efficient solution, namely walking (Thelen & Smith, 2007). The concepts of dynamic systems theory have also been successfully applied to a number of other research topics including language development (van Geert, 1991), cognitive development (Smith & Thelen, 2003; Spencer et al., 2007), and socioemotional development (Lewis, Lamey, & Douglas, 1999).

### 1.1.4 Dynamic systems in functional brain imaging

Systems theory has been indispensable in furthering our understanding of the dynamics of the human brain. For decades, neuroimaging studies have measured task-specific brain activation relative to some putative baseline state, in which brain activity is thought to be low. Many studies also presume that spontaneous brain activity at baseline is sufficiently random enough to be averaged out in statistical analysis (Deco, Jirsa, & McIntosh, 2011). However, Biswal et al. (1995) demonstrated that brain regions that are active together during motor activity maintain a high correlation of BOLD (blood-oxygen-level-dependent) fMRI (functional magnetic resonance imaging) signal fluctuations at rest. These consistent, distributed patterns of activity in the absence of an overt task are referred to as resting-state networks (RSNs). RSNs map onto anatomical and functionally connected brain regions, and are disrupted in brain disorders (Hagmann et al., 2008; Garrity et al., 2007). Accordingly, their activity dynamics may reflect some important aspect of the function of neural circuits. In the view of Deco, Jirsa, and McIntosh (2011), brain dynamics at rest represent a constant state of inner state exploration, where the brain can freely generate network configurations that would be optimal for an

incoming input: "The resting state is like a tennis player waiting for the service of his or her opponent. The player is not static, but continues to move with small lateral jumps left and right to be able to react more effectively to the impending serve" (Deco, Jirsa, McIntosh, 2011).

As this metaphor illustrates, one key feature of RSNs are their intrinsic fluctuations or "noise." This noise may reflect the imprecise nature of biological operations on the level of individual neurons (e.g. variation in the activation of ion channels) and ensembles of neurons (e.g. variation in transmission timing). On the level of nonlinear dynamic systems, we have discussed the contribution of intrinsic noise to the formation of new states. In the absence of noise, there is little capacity for the system to explore new states. With noise, the systems can visit several metastable network configurations spontaneously and converge on a configuration that meets the demands of the current situation. McIntosh and colleagues (2010) use empirical data from infants and children to show that brain noise appears to increase with maturation. They hypothesize that this reflects the formation of new functional networks through the refinement of anatomical connectivity and learning during development. Thus, noise enables the exploration of, and is a reflection of, the brain's dynamic repertoire (McIntosh et al., 2010).

### 1.1.5 Dynamic systems in cellular differentiation and gene expression

The dynamic systems principles at play at the level of brain networks also scale to the cellular and molecular levels. A central question in developmental biology is that of multipotency and cell fate; specifically, how do genetically identical cells differentiate into vastly distinct, stable, and specialized cell types? One major influence are *extrinsic* factors that act nonuniformly on cell populations and thus trigger a cellular response in only a subset. For example, during brain development, the generation of neural diversity depends on temporal and layer-specific spatial patterning of neural progenitors and the transcription factors they produce (Holguera & Desplan, 2018). However, even a clonal culture of cells exposed to uniform environmental conditions comes to show enduring phenotypic heterogeneity (Orkin & Zon, 2002). This *intrinsic* variability is linked to stochasticity in gene expression. *Gene expression noise* is a consequence of the random molecular events underlying transcription and translation, including the formation and decay of single molecules and multi-component complexes (Kaern et al., 2005). Recent genome mapping studies have suggested that variability in nuclear and chromatin organization also contribute (Finn & Misteli, 2019). Gene noise is accentuated when

transcription follows slow kinetics, resulting in uneven transcriptional products that are amplified by a high translation rate. This phenomenon is called *translational bursting* (Kaufmann & van Oudenaarden, 2007). Because fluctuations in protein production are generally not synchronized between cells, they eventually create a heterogeneous population over time. In addition to relatively fast fluctuations caused by gene expression noise, studies in mammalian cells demonstrate slow fluctuations in protein levels on the order of days (Sigal et al., 2006; Chang et al., 2008). In clonal cultures of mouse haematopoietic progenitor cells, this process resulted in spontaneous "outlier" cells, reconstituting the distribution of the parental population over the course of a week. Although the unique gene expression profiles reverted, they lasted long enough to create a differential proclivity for the erythroid or myeloid lineage (Chang et al., 2008). Thus, intrinsic variability in gene expression appears to have a developmental role.

While several formal frameworks have attempted to explain how slow fluctuations cause intrinsic non-genetic heterogeneity, the most comprehensive account comes from dynamic systems theory. In this framework, gene expression is determined by the complex, high-dimensional interactions of the gene regulatory network (GRN), which encompasses virtually the entire genome (Huang, 2009). Again, the dynamics of this network are best conceptualized as a *potential landscape*, where the gene expression profile associated with an attractor state can be interpreted as defining a particular cellular phenotype. Huang (2009) uses GRN consisting of two mutually inhibiting genes (*A* and *B*) to demonstrate basic concepts in network dynamics. In this example, each point in state space represents a combinatorically possible network state resulting from the expression levels of *A* and *B*. However, not all states are equally probable. A network state with high expression levels of both *A* and *B* would be highly unstable and hence unlikely to exist. This state forms a "hill" between two stable attractors – one with high *A* expression and low *B* expression, and one with the reverse. External regulatory signals (e.g. transcription factors) and larger perturbations (e.g. strong gene expression noise) can alter the expression of multiple genes and induce a transition across hills separating attractor states. This discontinuous transition embodies a differentiation event for an individual cell (Huang and Kauffman, 2009). Accordingly, cell-to-cell heterogeneity in a population of cells is best characterized as a "cloud" of individual points moving over the potential landscape. Kauffman (1993) describes the landscape of a complex network as "rugged," containing multiple nested attractors and sub-attractors. Thus, the simultaneous occupation of multiple major attractors by the cloud represents

a population with a number of distinct cell-type. On the other hand, the cloud can disperse over several sub-attractors within a cell-type attractor (Huang, 2009). Facilitated by gene expression noise, transitions between these sub-attractors manifest as the slow fluctuations in protein expression seen in empirical studies.

### 1.1.6 Summary and extension to critical periods

The brain is a dynamic system – at all times and on all levels. Dynamics is the language of stability and change, thus developmental science is the study of when systems are stable or changing, and what makes them change (Thelen & Smith, 2007). We have discussed the tendency of complex systems, consisting of many embedded levels, to self-organize into cohesive patterns. Many of the systems we have covered are highly non-linear, meaning that small changes in one or more components can lead to large-scale reorganizations of the system. The idea that *the brain naturally passes through periods of stability punctuated by spontaneous change* is our main takeaway from dynamic systems theory.

In one final example, I will demonstrate the utility of dynamic systems theory for unpacking a classical property of brain development. For half a century, neuroscientists have known that plasticity is enhanced during specific windows of opportunity. During these *critical periods*, experience provides information that is essential for normal development and that permanently alters function (Hensch, 2005). Animal models have provided clear molecular, cellular, and structural evidence of critical periods in the primary sensory systems of several species (Berardi, Pizzorusso, & Maffei, 2000). I will focus on the primary visual cortex (V1C), which is the premier model of experience-dependent plasticity. Neuronal activity is important for the formation and maturation of neural circuits in this area (Katz & Shatz, 1996). For example, thalamocortical axons that serve either the right or left eye terminate in layer IV of the visual cortex to produce alternating ocular dominance (OD) columns (Hubel, Wiesel, & LeVay, 1976; Shatz & Stryker, 1978). Occluding one eye during development causes expansion of the OD columns serving the open eye and retraction of thalamic inputs serving the occluded eye, which become reduced in size and complexity (Antonini & Stryker, 1996; Antonini, Fagiolini, & Stryker, 1999). This structural reorganization and concomitant shift in spiking response of neurons in visual cortex following monocular deprivation (MD) occurs only during a critical period defined by behavioral amblyopia (Gordon & Stryker, 1996). However, the critical period

itself is plastic. The onset of plasticity in mice can delayed by preventing the maturation of GABA (γ-aminobutyric acid)-mediated transmission via *Gad65* knockout (Hensch et al., 1998), or by dark-rearing from birth (Morales, Choi, & Kirkwood, 2002). Conversely, the critical period can be expedited by enhancing GABA transmission with benzodiazepines shortly after eye-opening (Fagiolini et al., 2004), or by promoting the rapid maturation of inhibitory interneurons via transgenic overexpression of BDNF (brain-derived trophic factor) (Huang et al., 1999). There is also evidence that critical period closure is not absolute and may even be "reversed." For example, although brief MD in the rodent model only produces OD plasticity within the critical period (Gordon & Stryker, 1996), a slightly longer duration of MD causes a robust OD shift in adults (Sawtell et al. 2003). Moreover, preceding a brief MD with a prolonged binocular deprivation in adult rats induces a "juvenile-like" OD plasticity (He et al., 2006). Studies from our lab have also demonstrated that the proteins associated with the molecular mechanisms underlying critical period plasticity show protracted developmental trajectories and dynamic changes well beyond the closure of the critical period (Pinto et al., 2010; Siu et al., 2017).

The critical period can be thought of as a unique state that the visual cortex inhabits early in life. The sheer amount of ongoing change in the constituent parts (e.g. synaptogenesis, synaptic pruning, synaptic spine motility, glutamatergic and GABAergic receptor maturation, intra-cortical myelin formation) maximizes the intrinsic noise of the system and pushes it to the edge of criticality. Under these conditions, the system is exquisitely sensitive to extrinsic influences, namely visual experience. Even the shadows of blood vessels in the eye are imprinted on the visual cortex (Adams & Horton, 2002). During normal development, visual experience enables the systems to transition from an energetically expensive and unstable state – defined by exuberant axonal processes and immature synapses – to a preferable state of stability where the adult cortex is organized into functional patterns serving each eye. If input from one eye is absent, as in MD, the adult attractor state forms without its influence. Studies demonstrating that the onset and closure of the critical period can be manipulated by pharmacological and environmental agents highlight that it is the product of a dynamic system, in contrast to a deterministic window that is hard coded within the organism. The dynamic view of the visual cortex is also supported by studies exhibiting "reversal" of the critical period and reinstatement of plasticity. These results can be conceptualized as a return of the system to a set of attractors that were present during the critical period. Nevertheless, these studies cannot fully induce the

same level of plasticity found during the critical period in adults. The limited and irreversible nature of the critical period is a property of dynamic systems in general. Self-sustaining patterns of interactions that define the structure and function of the brain persist because they are able to consume and dissipate energy. The more energy dissipation takes place, the more the process of organization – such as the emergence of states or attractors – becomes irreversible. This is because the number of possible paths back to the initial state grows exponentially, and the probability of the system getting onto a path back to a previous state is extremely low (van Geert, 2020). It is likely, however, that the system exhibits transient states that are reflective of a future configuration. Interestingly, this is exactly what is found in studies investigating synaptic phenotypes in animal models of monogenic neurodevelopmental disorders (Meredith, Dawitz, & Kramvis, 2012; Kroon, Sierksma, & Meredith, 2013).

The following section will review the paucity of developmental dynamics concepts in cross-sectional transcriptomic studies of the human brain, and the benefit of approaching this data with states and nonlinearity in mind.

## 1.2 Developmental transcriptomics of the human brain

The brain is an exceptionally complex structure comprised of billions of glial and neuronal cells that organize into dense functional networks, capable of communicating within and across brain regions to produce cognition and other sophisticated behaviours. The development of the brain is equally miraculous. During the first few years of life it nearly quadruples in size, growing through the birth of new neurons, multiplication of glial cells, expansion of myelination, formation of synaptic connections, and the pruning of unused ones (Breen et al., 2018). In humans, the brain continues to undergo refinement and maturation well into adolescence, and is further shaped by ageing and degeneration in late adulthood (Keil, Qalieh, & Kwan, 2018). These processes are coordinated by cascades of molecules that are synchronized in space and time via tightly regulated gene expression. Accordingly, much of our current knowledge of biological changes underlying human brain development and ageing has been deduced from transcriptomic data. For decades, the expression of individual genes has been investigated with lower-throughput techniques such as *in situ* hybridization and Western blotting. The advent of microarray and RNA-seq technologies have made unbiased, genome-wide

interrogation of the transcriptome possible, and paved the way for a number of large-scale studies.

### 1.2.1 Contributions of recent developmental brain transcriptome studies

In one of the most comprehensive human brain transcriptome stuides to date, Kang et al. (2011) used a high-throughput exon array to analyze 16 human brain regions (hippocampus, striatum, cerebellar cortex, amygdala, mediodorsal nucleus of the thalamus, and 11 neocortical areas) from 57 post-mortem human brains. Tissue samples were collected from early fetal development (5 weeks after conception) through ageing (82 years), covering virtually the entire human lifespan. The results of this work indicate that the majority of brain-expressed genes coding for proteins are temporally, and to a lesser extent, spatially regulated, and that this regulation occurs primarily during prenatal development (Kang et al., 2011). The authors used weighted gene co-expression network analysis (Zhang & Horvath, 2005) to identify 29 modules corresponding to distinct spatiotemporal patterns and biological processes. For example, two large-scale modules were identified with opposing developmental trajectories across regions. The waning module was enriched for gene ontology (GO) categories related to transcription factors and zinc-finger proteins, whereas the waxing module was enriched for processes involved in postnatal brain maturation, such as calcium signalling, synaptic transmission and neuroactive ligand-receptor interaction (Kang et al., 2011). Additionally, the authors analyzed the trajectories of individual genes and lists of manually curated genes associated with specific neurobiological categories. This approach enabled the validation of expression data for several neuronal genes with independent immunohistochemical findings, and the visualization of lifespan changes in several processes including synapse development, dendrite development and myelination.

In a more focused study, Colantuoni et al. (2011) explored the temporal dynamics of the transcriptome in the dorsolateral prefrontal cortex. With the use of two-color arrays, gene expression was quantified in 269 human brain samples spanning gestational week 14 through 80 years of age. Interestingly, a significant number of genes showed a reversal of expression between fetal development and early postnatal life. This pattern of reversals was mirrored decades later in ageing and neurodegeneration (Colantuoni et al., 2011).

The high-quality transcriptomic data from both of these studies are publicly available on the Gene Expression Omnibus (GEO). Advances in sequencing technology coupled with funding

from the National Institute of Health (NIH) and private foundations (e.g. Allen Brain Institute) have contributed to a growing body of brain transcriptome databases (Keil, Qalieh, & Kwan, 2018). These repositories serve as general cartographers of human neurodevelopment and hold the potential to identify important biological mechanisms as well as appropriate tissues and developmental points to query in experimental studies. For example, Shim et al. (2012) performed a focused gene co-expression analysis on the Kang et al. (2011) dataset and identified SOX4 and SOX11 as novel regulators of the *Fezf2* E4 enhancer. The spatiotemporal nature of the same dataset was leveraged by Bae et al. (2014) to determine how a mutation upstream of *GPR56* leads to perisylvian polymicrogyria, a disruption of gyri formation in the area around the sylvian fissure. The authors assessed candidate gene expression in the ventrolateral prefrontal cortex during the fetal development, and identified members of the RFX transcription factor family as potential mutational targets. Lastly, Kwan et al. (2012) used the Kang et al. (2011) dataset to build a list of proteins that interact with *NOS1* in the fetal cortex; a protein-mRNA interaction that is unique to humans. This helped identify FMRP as a post-transcriptional regulator of human NOS1 expression. In sum, cross-sectional transcriptomic stuides of the human brain facilitate our understanding of the biological events that occur during normal development and that are potentially disrupted in disease.

### 1.2.2 Potential gaps in developmental transcriptome studies

Despite their utility, there are a number of caveats to consider when gathering, processing, and interpreting transcriptomic data. Important technical and biological factors include post-mortem tissue quality, RNA extraction and stabilization methods, differences in platform technology, batch-effects, library generation, genome mapping, normalization methods, differential splicing, and mRNA-to-protein correspondence. While crucial, these topics are outside of the scope of the current review. Instead, I will focus on conceptual biases baked into common methods of analyzing developmental gene expression.

Unlike other areas of developmental science, most studies in developmental neurobiology lack an overarching theoretical framework. Often, they ask *what* is changing during development rather than *how* development itself is changing. This is a common approach to exploratory investigations that seek to study a particular phenomenon in an unbiased fashion. Accordingly, transcriptomic studies will survey the entire genome and identify changing patterns of gene

expression in space and time. The rationale is that an increase or decrease in the transcription of a particular gene (depending on the function of its protein product) says something about the biological processes occurring at that point in developmental time. As sequencing technology has become cheaper and more efficient, transcriptomic stuides have been able to add more tissue samples to their analyses. With enough samples spanning from fetal development to senescence, it becomes possible to model continuous change in gene expression throughout the human lifespan. Thus, using linear or non-linear regression to approximate the developmental trajectory of genes is one of the most popular analyses in this area of research. It also produces a striking and intuitive visualization of when things are changing. A less common approach, belonging to a lower-throughput era, is age-group comparison. Samples are binned into age groups, often characterized by certain developmental milestones, and are compared using t-tests or an ANOVA.

Both of these approaches have implicit biases about the *dynamics* of development that are often overlooked. The first assumption of linear regression is that the relationship between X and the mean of Y is linear, followed by the assumptions of homoscedasticity, independence, and normality. While the assumption of normal residuals is not needed for non-linear regression, the main assumption is still that the data is well represented by the model (Berry, 1993). When analyzing gene expression using regression, we assume that there is some underlying *continuous* pattern in the data that is captured by the model. Unsurprisingly, this approach is not suited for describing discontinuities in empirical data, such as sudden increases or decreases in expression. Even sigmoidal functions, which can change in value rapidly, are still continuous if there are no gaps in the curve. In the previous section, we covered several examples that suggest brain development is at times non-linear and discontinuous. Therefore, we must ask whether continuous curves are indeed the descriptive format that neuroscientists wish to use when extracting developmental information from their data. While assigning samples to age-defined groups may appear to agree with discontinuous, stage-like development, the groups are determined *a priori* and cannot reveal anything new about the dynamics of development.

Variability is an important driver of development that is handled differently in various statistical analyses. Due to the cross-sectional design and use of bulk tissue in nearly all human developmental transcriptomic studies, one cannot directly determine the source of variability in

gene expression and parse the biological, technical, and random effects. In regression models, the goal is to estimate the mathematical formula applied to a set of explanatory variables (e.g. age, sex, post-mortem internal, pH, probe quality) that best predicts the dependent variable (e.g. gene expression). Most regression programs use a method such as least squares to estimates the coefficient of each explanatory variable while minimizing unexplained variance in the error term. In other words, regression asks how much of the variance in gene expression can be explained by age in combination with other explanatory variables. Let us consider a hypothetical scenario where two samples have a similar phenotype (age and sex), but are starkly different in terms of expression for a particular gene. In a regression analysis, the differences between these samples is not captured by the explanatory variables and is instead attributed to error, resulting in a poor fit. Thus, when fitting a continuous model onto developmental gene expression, each individual data point is treated as an error-laden instance of a common, underlying biological pattern. It is possible, however, that this "error" represents multiple heterogeneous patterns (or states) that overlap in developmental time. While the ideal approach to understanding the source and function of biological variability would involve measuring gene expression at a cellular resolution across multiple developmental periods within the same individual, this is not yet possible with current technology. In the interim, there should be an emphasis on using analyses that treat variance as an essential element of the higher-order patterns that constitute the system, rather than noise that should be averaged out.

Kang et al. (2011) quantified probes representing 17,565 protein-coding genes in 1,340 tissue samples, producing a total of 23,537,100 data points. The size and high dimensionality of spatiotemporal transcriptomes pose several challenges to computational analyses. It is clear that the complex phenotypes of the brain arise from the collective activity of many genes, so neuroscientists must first summarize this data to identify the biological significance. A common workflow is to group genes that share common functions defined by prior knowledge or experimental data. Gene-set databases like the Gene Ontology project (Ashburner et al., 2000) use hierarchically structured terms that describes gene products in terms of their associated biological processes, cellular components, and molecular function. Genes can be categorized by their involvement in human diseases (Amberger et al., 2015), and single-cell RNA-seq stuides have identified groups of markers genes that are up-regulated in certain cell-types (Tasic et al., 2016). These resources are valuable but incomplete. Alternatively, we can identify groups of

genes with similar expression patterns over a set of samples, as they are likely to be involved in the same biological processes. *Gene co-expression* is useful for functional annotation, pathway analysis and reconstruction of gene regulatory networks. Nevertheless, the most common measure of co-expression is Pearson correlation, which can only capture linear relationships between variables (Mahfouz et al., 2017). Co-expression analyses can be extended using a network-based approach. The most commonly used is weighted gene co-expression network analysis (WGCNA), where genes are organized into modules based on topological overlaps in co-expression patterns (Zhang & Horvath, 2005). The overall expression profile of each module is represented by an eigengene, which is often visualized in developmental time with a continuous curve.

In this section, we have reviewed various approaches to analyzing transcriptomic data with the goal of linking gene expression to the etiology of neurodevelopment and ageing. Many studies focus on identifying spatiotemporal expression patterns across genes, as these are thought to reflect developmental events, such as neurogenesis, myelination, and cellular senescence. In general, the emphasis is on processes that change over the lifespan rather than the overall dynamics of the change. Often, development is implicitly assumed to be continuous, following from the use regression-based analyses and visualizations. Several studies have broadly discussed the dynamics of gene expression. Kang et al. (2011) found that over 80% of genes are differentially regulated across brain regions and developmental periods, and that the greatest regional differences in expression occur during prenatal development. Colantuoni et al. (2011) found that the rate of gene expression changes remains relatively low until 50 years of age, when it increases to mirror the changes seen in early postnatal life. However, a number of important questions remain unaddressed: namely, *can gene expression be used to model non-linear, stage-like development and ageing in the human brain?* Further, *can non-linear analyses reveal new insights into the biological changes occurring in healthy development and ageing, as well as in disease?*

## 1.3 The current study

Human brain development and ageing are complex, multidimensional processes. A formal framework for describing systems with these properties is dynamic systems theory. Briefly, the interactions between components of a complex system tend to self-organize into

relatively stable patterns or states (Prigogine & Nicolis, 1971). Development can be conceptualized as the progressive generation and exploration of stable states, which occurs at multiple levels of analysis and over multiple timescales, and in response to internal and external fluctuations (Thelen & Smith, 1994). Crucially, change between states is often sudden and takes the form of a discontinuity. Our understanding of the dynamics of the system occurs through the measurement of some selected property. By doing so, we can take a few properties out of the system, but we can never take the system out of the properties (van Geert, 2020). Currently, one of the most comprehensive measures of brain development is gene expression. Transcriptomic data is exceptionally rich, and several studies demonstrate that it captures high-level processes including gene-regulation and biological function (Dillman et al., 2017; Colantuoni et al., 2011; Kang et al., 2011; Breen et al., 2018). More focused studies have found that special developmental states, such as critical and sensitive periods, are driven by specific patterns of protein expression (e.g. excitatory-inhibitory balance in primary visual cortex) (Hensch, 2005). Nevertheless, most models describing the brain transcriptome over the lifespan are continuous. An important question remains: *can human brain development and ageing be described as a series of stable, biological states?*

To address this question, the current study leverages large-scale transcriptomic datasets available on the Gene Expression Omnibus (GEO) and modern, data-driven analyses. We implement robust and sparse k-means clustering (RSKC) (Witten & Tibshirani 2010; Kondo et al., 2016) to identify groups of samples with similar high-dimensional patterns of gene expression. This approach yielded developmental clusters with clear progression across the lifespan and identified feature genes that predict the separation of clusters. Several feature genes emphasized the developmental influence of well-known mechanisms, such as inhibitory interneurons and intracortical myelin in primary visual cortex (V1C), in addition to other genes that may merit further investigation, such as *VAMP1*. Clusters were unpacked using differential gene expression analysis, allowing enriched biological processes and cell types to be visualized. Data from 16 brain regions were independently surveyed using our analysis pipeline, revealing spatially distinct patterns of developmental clusters. Notably, gene expression in subcortical structures does not separate samples into postnatal developmental clusters. RSKC was also used to cluster brain regions with similar patterns of gene expression in order to identify common developmental clusters. Interestingly, region clusters roughly charted anatomical separation, and

clusters enriched for frontal areas exhibited sample separation by sex. Lastly, our clustering approach was used to test the hypothesis that individuals with bipolar disorder experience accelerated brain ageing. This hypothesis was not supported, but a new way to conceptualize clinical subgroups is presented.

In sum, the current study demonstrates that the lifespan dynamics of the human brain transcriptome can be described by a series of overlapping developmental clusters. These clusters exhibit regional and sex-based differences, and are driven by a small subset of feature genes. Moreover, clustering detected nuanced differences in gene expression during postnatal development that were underrepresented in the original publication. Further testing is required to ascertain whether developmental clusters indeed represent stable biological states, and whether they represent inter-individual differences or intrinsic variability. Even so, developmental clusters present a different way of thinking about human brain development that shifts away from discrete age-defined stages or continuous trajectories to a hybrid classification scheme of overlapping states. In this model, the brain at any point during the lifespan may be represented by two or more states. An intriguing possibility is that cortical tissue transitions between states in either a single step or by fluctuating back and forth until the optimal state is achieved.

## 1.4 Study rationale, objective, and specific aims

**Rationale**: Transcriptomic studies are one of the most comprehensive accounts we have of the biological events underlying human brain development and ageing. Many analysis methods and descriptive models applied to gene expression data implicitly assume that developmental change is continuous and uninterrupted. Perhaps this bias is often overlooked because the emphasis is on *what* is changing during development rather than *how* development itself is changing. Indeed, despite the richness of transcriptomic data and its capacity to recapitulate higher-order functions, few have used it to understand the dynamics of brain development.

**Objective:** To explore the developmental dynamics of the human brain transcriptome by testing whether brain development and ageing can be described as a series of stable, biological states.

**Specific Aims:**

1. Cluster developmental samples from primary visual cortex (V1C) by gene expression and describe the pattern of cluster progression across the lifespan. Identify feature genes

driving the separation of developmental clusters and investigate their biological functions. Assess the contribution of cell type to each cluster.

2. Compare developmental clusters and features genes across 16 brain regions. Cluster regions by gene expression and compare developmental progression.
3. Cluster developmental controls and bipolar cases together. Investigate whether bipolar samples cluster with older controls, if at all, as suggested by the accelerated ageing theoretical framework.

**Approach:** I leverage large-scale transcriptomic datasets available on the Gene Expression Omnibus (GEO) and modern, data-driven analysis. We implement robust and sparse k-means clustering (RSKC) (Witten & Tibshirani 2010; Kondo et al., 2016) to identify groups of samples with similar high-dimensional patterns of gene expression. This algorithm also selected weighted features that drive the separation of clusters. Lists of genes up-regulated and down-regulated in each cluster were unpacked using differential gene expression analysis. These lists were queried for enrichment of cell type markers and Gene Ontology (GO) biological processes.

# Chapter 2. Introduction to Clustering

## 2.1 Introduction

Descriptive models of human brain development have evolved over time to suit the precision of investigative tools. Early theories were informed by behavioural observations made by cognitive psychologists, many of whom described development as a series of distinct stages (Piaget, 1952; Erikson, 1959; Vygotsky, 1987). Morphometric studies of cortical synapses and dendrites found rapid growth early in life, followed a protracted period of synaptic pruning (Huttenlocher, 1978; Huttenlocher, 1990). The non-constant change of these processes is represented by improvised axes used to describe the number of synapses over developmental time. One strategy for increasing statistical power is to measure age on a categorical scale, where individuals of a similar age are binned together. Many low-throughput molecular studies (Martin et al., 1998; Williams et al., 1993) and studies of psychological development (Fischer et al., 1985; Flavell, 1982) have relied on this approach. The advent of technologies such as fMRI, EEG and MRS provided better windows into the living brain, and longitudinal analyses of brain structure and function blurred the divisions between developmental periods. Consequently, the dominant view today is that human developmental progresses along a continuum (Shaw et al., 2008; Grydeland et al., 2013; Mills et al., 2014; Fjell et al., 2015).

Recently, molecular tools have become integrated with developmental neuroscience and an abundance of research has focused on mapping the transcriptomic development of human cortex (e.g. Allen Human Brain Atlas, Human Connectome Project, PsychEncode). As sequencing technology has become cheaper and more efficient, transcriptomic stuides have been able to add more tissue samples to their analyses. With enough samples spanning from fetal development through to senescence, it becomes possible to model continuous change in gene expression throughout the human lifespan. The precision afforded by the inclusion of many individuals in functional and structural imaging studies (Gogtay et al., 2004) and sequencing stuides (Kang et al., 2011; Colantuoni et al., 2011) has biased our view of brain development towards a continuous vignette. Conversely, the current study asks whether discontinuous stages of development are represented in the human brain developmental transcriptome.

One approach to addressing this question is the identification of subgroups in developmental samples. *Unsupervised clustering* is an analysis method perfectly suited for this task. Clustering is the process of identifying distinct groups of objects in a dataset, and can be

described as either hierarchical or partitional. Hierarchical clustering involves calculating all the pair-wise similarities between samples and constructing a tree-like dendrogram by iteratively grouping the most similar gene pairs. The major drawbacks of this method are that it is sensitive to outliers, cannot undo a suboptimal previous step, and can be ambiguous to interpret (Duda, Hart, & Stork, 2012). Hierarchical clustering is most appropriate when a hierarchical organization is expected in the data, and the main goal is reconstructing that organization. On the other hand, partitional clustering involves optimizing a simple model to fit the data. The most widely used partitional algorithm is k-means, where groups are formed by minimizing the within-cluster sum of square distances between each point in the cluster and the cluster center. K-means has several disadvantages: it requires the number of clusters to be set *a priori,* is influenced by the starting seed, and has trouble scaling with the number of dimensions in the data (Tibshirani, Walther, & Hastie, 2001). In parallel with clustering, *dimensionality-reduction* methods aim to find features that can adequately characterize the original high-dimensional data in a lower dimensional space. Principal component analysis (PCA) describes the process of computing a series of orthogonal vectors that iteratively capture the greatest amount of variance in the data. Despite its utility, PCA cannot describe non-linear relationships (Jolliffe & Cadima, 2016).

Several transcriptomic (Kang et al., 2011; Colantuoni et al., 2011) and proteomic (Carlyle et al., 2017; Breen et al., 2018) developmental studies have used clustering and PCA to visualize the molecular differences between individual samples. However, none found good separation of developmental periods. This is possibly because many of the clustering algorithms used in these studies require hundreds or even thousands of independent data points to reliably capture global structure. Nevertheless, most studies of the human brain include less than 100 unique samples.

## 2.2 Robust and sparse k-means clustering

Witten and Tibshirani (2010) present *sparse clustering* as a solution to clustering data with relatively few observations, which is common in investigations that include rare samples. This approach was originally developed for the classification of cancer types using gene expression profiling (Tibshirani et al., 2001; Witten & Tibshirani, 2010). Sparse clustering leverages high-dimensional datasets with many *features* (e.g. genes) to enhance the precision of clustering *observations*. One challenge that arises from including many features is that not all

features contribute to clustering. Some features are expected to change from cluster to cluster, while other behave similarly across clusters. Standard K-means clustering is vulnerable to the influence of noise variables (Kondo et al., 2016). On the other hand, sparse k-means (SK-means) simultaneously finds clusters and clustering feature. A lasso-type penalty is used to assign weights to each feature and cluster according to a weight dissimilarity measure (Witten & Tibshirani, 2010). Kondo et al. (2016) extend the SK-means method to make it resistant to outliers by trimming a fixed proportion of observations in each iteration. These outliers are flagged both in terms of their weighted and unweighted distances to eliminate the effects of outliers in the selection of feature weights and the selection of partitions. The RSKC algorithm is implemented in the R package *RSKC* (Kondo et al., 2016) and is used extensively in the current study.

The capacity of RSKC to simultaneously identifies clusters and features benefits the current study in two ways. First, transcriptomic data have a large number of features (genes) relative to observations (samples) – in our case, we have 1,340 samples with 17,243 associated features. One expects that the underlying clusters in the data differ only with respect to a small fraction of features (Witten & Tibshirani, 2010). Sparse clustering is ideally suited to transcriptomic data as it utilizes a lasso-type penalty to adaptively select a subset of features that best cluster the data. Second, since we are clustering observations according to genes, we can use feature weights as a proxy for the contribution of each gene to the formation of developmental clusters. Its important to emphasize that unsupervised clustering algorithms such as RSKC classify objects without external classification (e.g. age). Developmental clusters are identified by gene expression alone.

## 2.3 Evaluation of clustering methods

Clustering is an unsupervised technique used to group together objects that are "close" in a multidimensional feature space (Brock et al., 2008). In the analysis of high-throughput transcriptomic data, clustering is often used to identify genes with similar expression patterns over a collection of samples. The present study used clustering to perform the inverse – identify groups of samples with similar patterns of gene expression across the genome. The resulting sample clusters serve as a proxy for a developmental "phenotype" that individuals may express

during a particular period of the lifespan. Therefore, the selection of a clustering algorithm is an important step that directly influences the formation of development clusters.

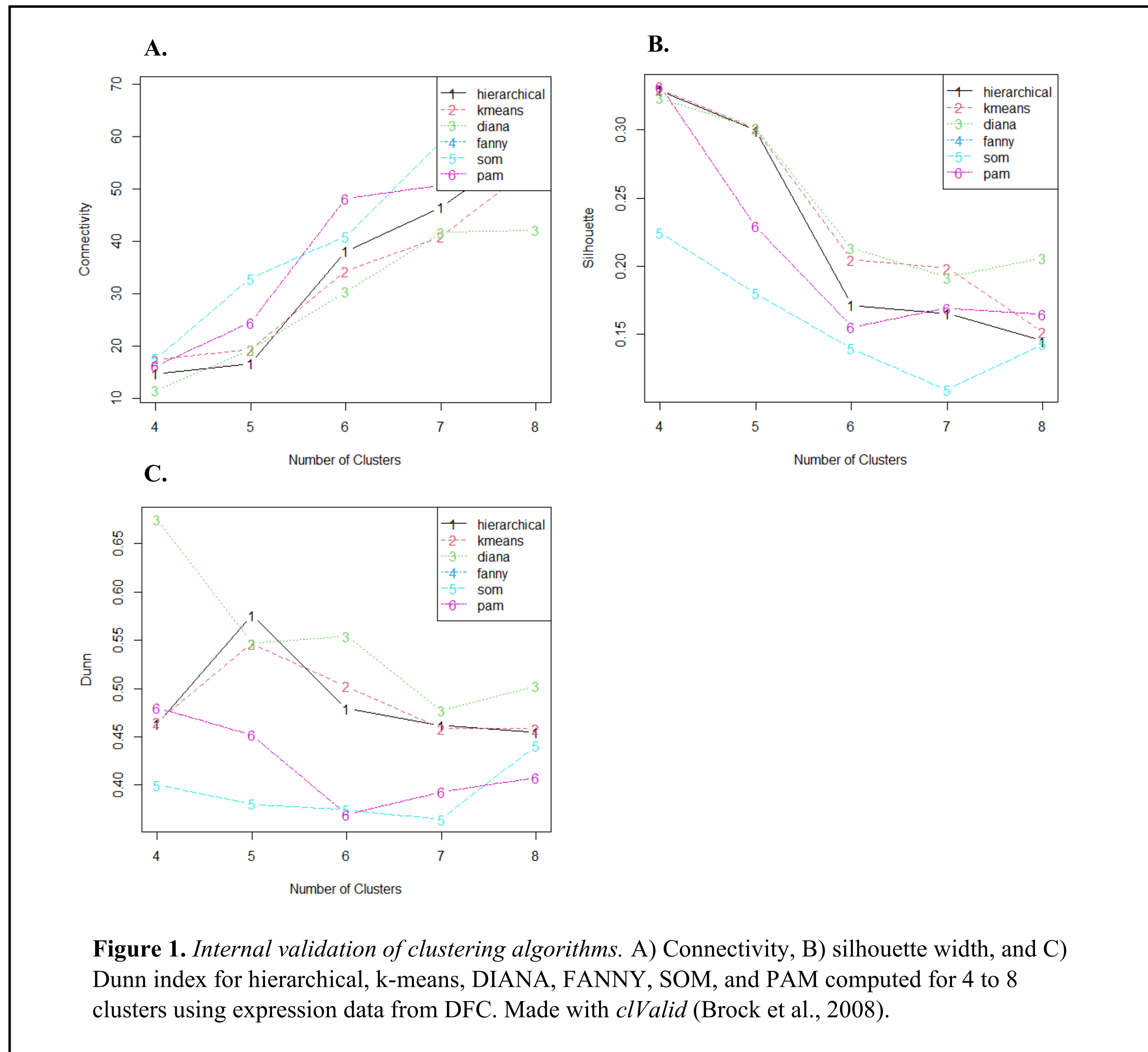


**Figure 1.** *Internal validation of clustering algorithms.* A) Connectivity, B) silhouette width, and C) Dunn index for hierarchical, k-means, DIANA, FANNY, SOM, and PAM computed for 4 to 8 clusters using expression data from DFC. Made with *clValid* (Brock et al., 2008).

There are a variety of measures that can be used to determine which clustering algorithm performs the best for a particular dataset or experiment. Internal validation measures reflect the compactness, connectedness, and separation of clusters (Brock et al., 2008). *Connectivity* is the extent to which observations are placed in the same clusters as their neighbours in feature space, and is a function that should be minimized (Handl et al., 2005). Compactness uses intra-cluster variance to assess homogeneity, and separation is measured by the distance between cluster

centroids. Both are combined in the *Dunn index* and *silhouette width* methods, which should be maximized (Dunn, 1974; Rousseeuw 1987).

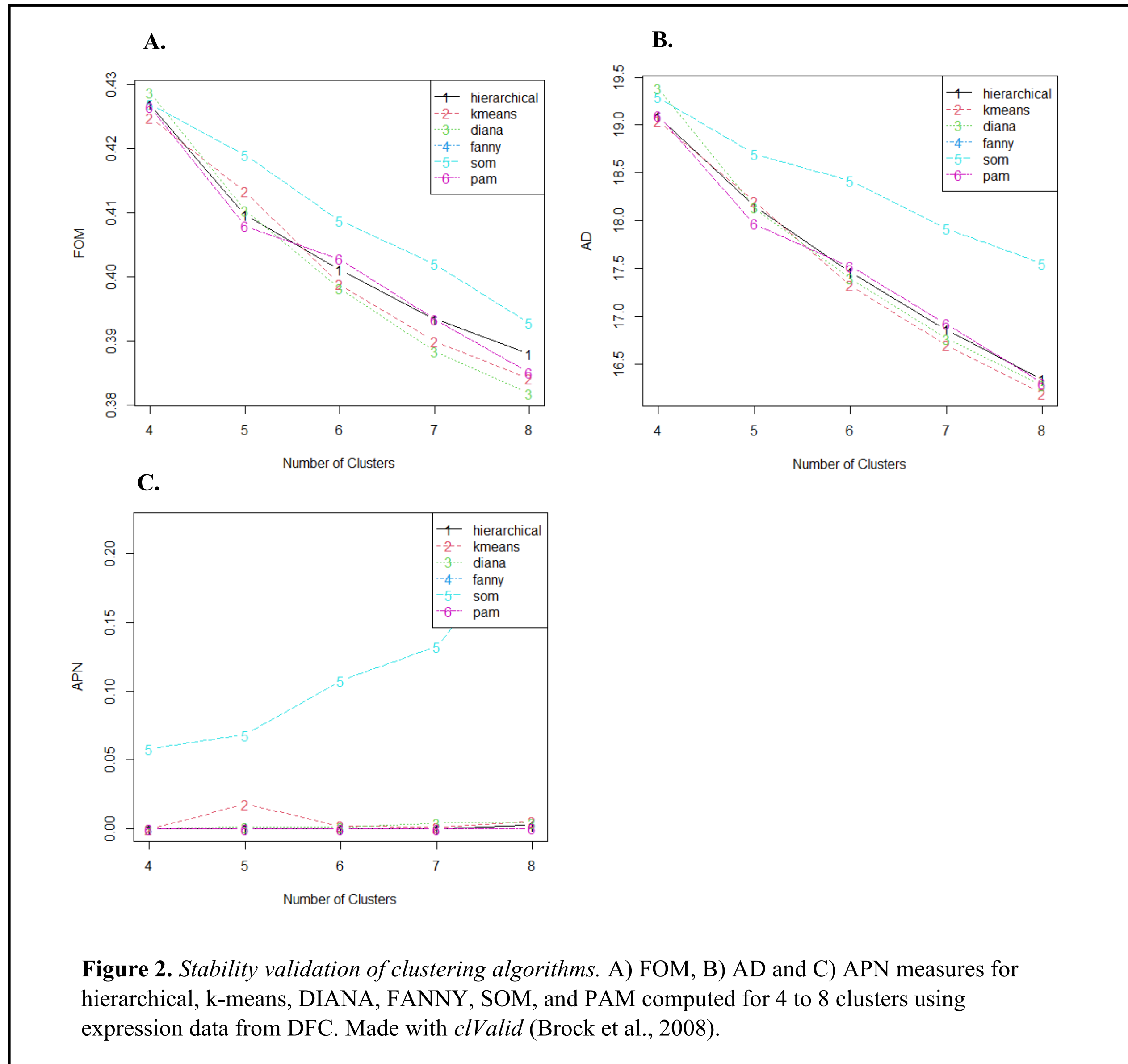


**Figure 2.** *Stability validation of clustering algorithms.* A) FOM, B) AD and C) APN measures for hierarchical, k-means, DIANA, FANNY, SOM, and PAM computed for 4 to 8 clusters using expression data from DFC. Made with *clValid* (Brock et al., 2008).

Here, expression data from DFC is used to assess the internal validity of various clustering algorithms across different numbers of clusters using the *clValid* package (Brock et al., 2008). The algorithms evaluated here are ward.D2 hierarchical clustering, k-means clustering, divisive analysis (DIANA), fuzzy clustering (FANNY), self-organizing maps (SOM), and partitioning around medoids (PAM). The results of internal validation indicate that DIANA

is superior across most numbers of clusters, followed closely by k-means and hierarchical clustering (Figure 1).

Clustering performance can also be validated using stability measures, which compare the results of clustering based on the full data verse clustering based on iteratively removing each sample (Datta & Datta, 2006). Measures of stability include the *average proportion of non-overlap* (APN), the *average distance* (AD), and the *figure of merit* (FOM). In all cases the average is taken over all the deleted columns, and all measures should be minimized (Brock et al., 2008). The stability of clusters produced by each algorithm over different numbers of clusters were assessed with these measures. The results suggest that DIANA, k-means, and hierarchical clustering have the best performance (Figure 2). Taken together, internal and stability validation measures indicate that gene expression in DFC is most competently clustered by the DIANA, k-means, and hierarchical algorithms. The previous steps were repeated for several brain regions with similar results.

The performance of robust and sparse k-means clustering (RSKC) was assessed in comparison to these three methods. Samples from each brain region were clustered by all four algorithms across different numbers of clusters, and performance was evaluated using distance-based statistics available in the *cluster.stats* function from the *fpc* package (Hennig, 2010). The statistics measured include: average distance between and within clusters; average silhouette width; Calinski and Harabasz index, which measures the ratio of the sum of between-clusters dispersion and of inter-cluster dispersion for all clusters, and should be maximized (Calinski & Harabasz, 1974); Dunn index, which measures minimum separation over maximum diameter, and should be maximized (Dunn, 1974); Dunn 2 index, which measures minimum average dissimilarity between clusters over maximum average within cluster dissimilarity, and should be maximized; entropy, which represents the uncertainty of splitting data into clusters, and should be minimized (Meila, 2007); minimum cluster size; Pearson gamma, which is the correlation between distances and a 0-1-vector where 0 means same cluster and 1 means different clusters, and should be minimized (Halkidi et al., 2001); separation index, which measures the magnitude of the gap between a pair of clusters, and should be maximized (Qiu & Joe, 2006); and the ratio of average within cluster distance over average between cluster distance, which should be mimimized.

Clustering validation revealed that hierarchical, k-means, DIANA, and RSKC perform similarly (Figure 3A). RSKC was superior in some measures, namely average distance within clusters, Calinski and Harabasz index, and Pearson gamma. In other measures, RSKC performed comparability to other algorithms or more poorly, such as in average distance between clusters,

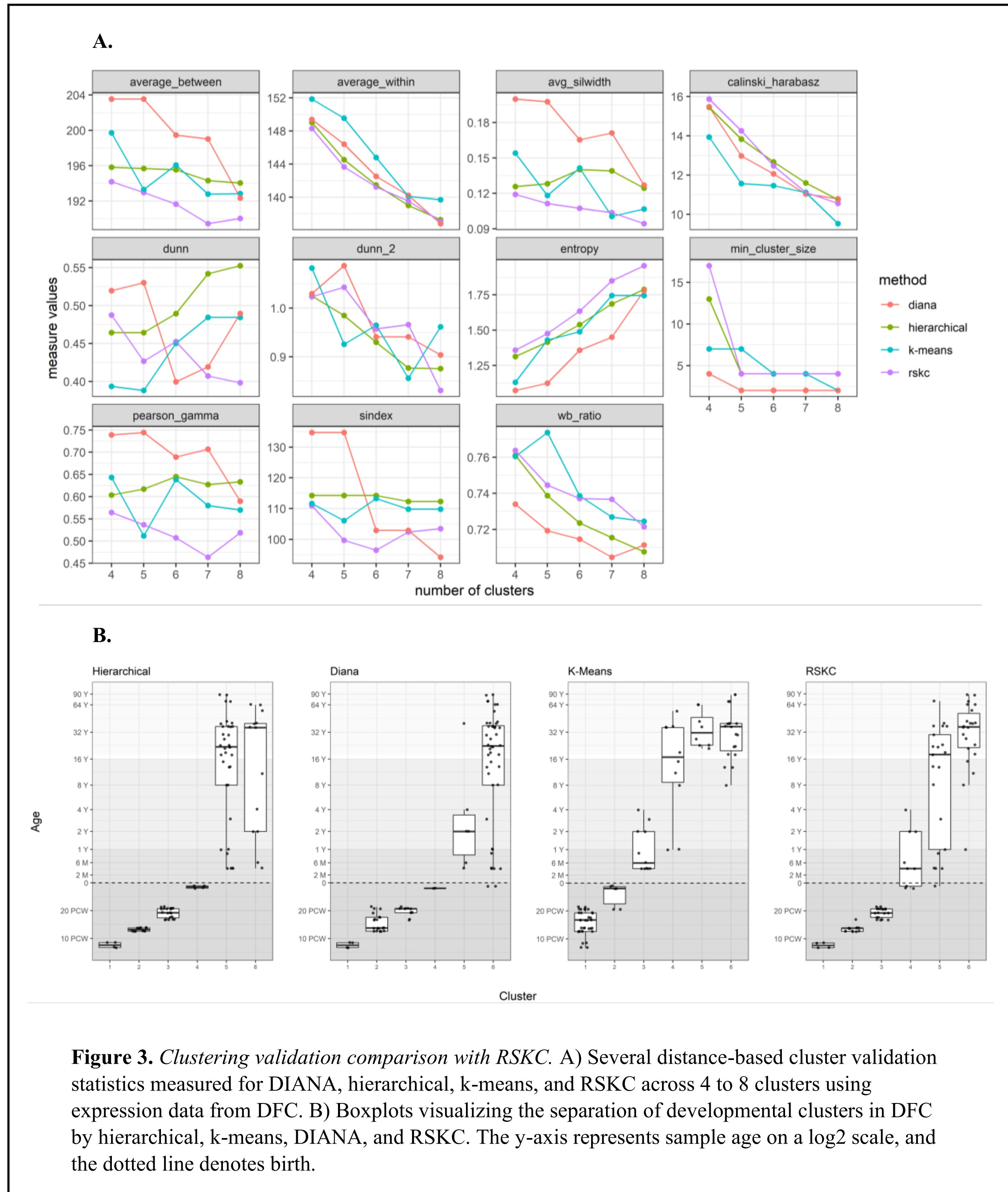


**Figure 3.** *Clustering validation comparison with RSKC.* A) Several distance-based cluster validation statistics measured for DIANA, hierarchical, k-means, and RSKC across 4 to 8 clusters using expression data from DFC. B) Boxplots visualizing the separation of developmental clusters in DFC by hierarchical, k-means, DIANA, and RSKC. The y-axis represents sample age on a log2 scale, and the dotted line denotes birth.

average silhouette width, entropy, and separation index. There are some important caveats to consider. Cluster validation statistics are not prescriptive, rather, they are tools for assessing the general performance of a clustering algorithm applied to a particular dataset. Clustering algorithms are also variable, and can produce different results over multiple runs. Lastly, some measures are linear and will favour certain clustering approaches over others. An important criterion for the current study that is not explicitly captured by validation statistics is the ability to identify developmental clusters that show progression across the lifespan. While hierarchical, k-means, and DIANA are able to identify developmental clusters, RSKC demonstrated the best separation of samples by age using gene expression data (Figure 3B). These observations were comparable across all brain regions.

## 2.4 Selection of *k*

An important step in k-means clustering is the selection of *k*, which denotes the number of groups to classify observation into. The correct choice of *k* is often ambiguous, as there are many different approaches for making this decision. Intuitively, an optimal *k* lies in between maximum generalization of the data using a single cluster and maximum accuracy by assign each observation to its own cluster. One of the most common heuristics for determining *k* is the elbow plot method, where the sum of squared distances of observations to the nearest cluster center are plotted for various values of *k*. As *k* increases, the sum of squared distances tends towards zero. The "elbow" occurs at the point of diminishing returns for minimizing the sum of squared distances, and the *k* value at this point is selected as the optimal number of clusters (Thorndike. 1953).

In order to tailor the selection of *k* to RSKC, we applied the elbow method to the weighted between sum of squares (WBSS), the objective function that is maximized by the algorithm. WBSS was calculated for various values of *k* and averaged over 100 iterations. The elbow was identified using the *elbowPoint* function in the *akmedoids* package (Genolini et al., 2015), which uses a Savitzky-Golay filter to smooth the curve and identifies the *x* where the curvature is maximized. This method indicated that $k = 6$ (Figure 4). It should be noted that the elbow plot can be an unreliable and subjective method of selecting *k*. In the future, another method should be used to corroborate these results.

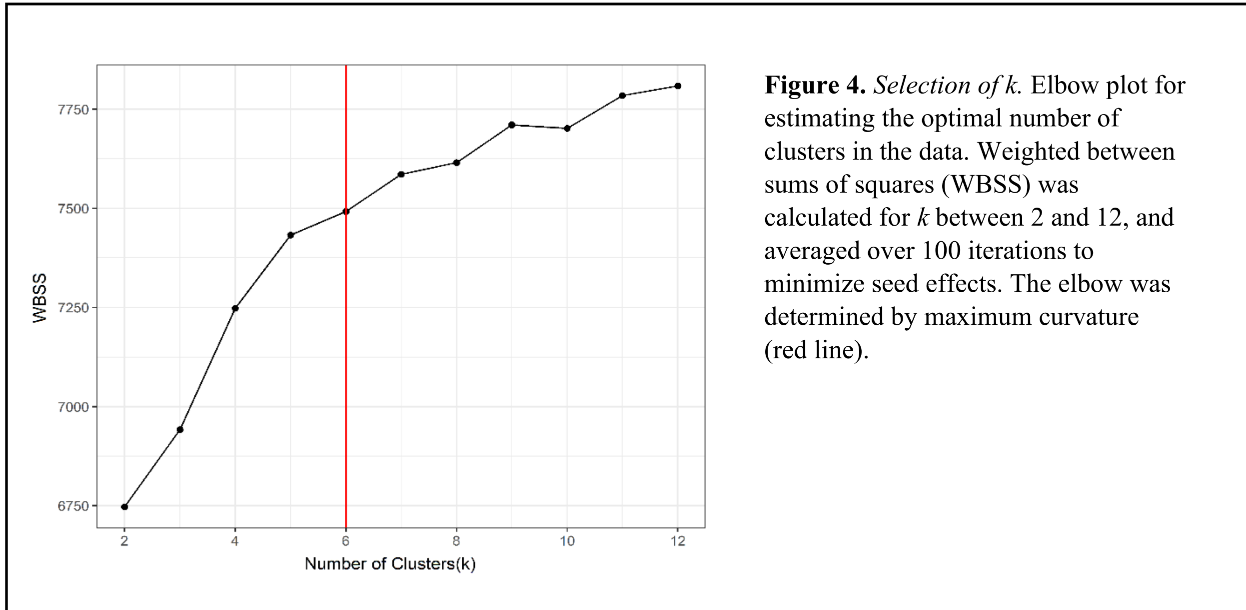


**Figure 4.** *Selection of k.* Elbow plot for estimating the optimal number of clusters in the data. Weighted between sums of squares (WBSS) was calculated for *k* between 2 and 12, and averaged over 100 iterations to minimize seed effects. The elbow was determined by maximum curvature (red line).

## 2.5 Minimizing biases

Before the expression data can be clustered by RSKC, some adjustments must be made. A principal finding of the original Kang et al. (2011) publication was that the majority of spatiotemporal differences in gene expression were detected before birth. While RSKC is capable of detecting subtle differences in the presence of larger ones, the selection of feature weights may become biased. In order the circumvent this issue, we use a double-layered approach. The data are first clustered to separate prenatal and postnatal samples, and then clustering is performed again on each subcluster. The resulting developmental clusters are comparable to performing RSKC once on all samples, and we obtain two sets of feature genes: one describing the separation of prenatal and postnatal samples, and another describing the separation of postnatal samples only. Features selected during the clustering of prenatal samples are not discussed here.

Another important consideration is the effect of random seed. When k-means clustering is performed, a randomly generated seed is used to determine the starting centroids of the clusters. If features perfectly match observations to groups, then the starting seed will not have an impact on the final cluster memberships. However, if there are many features that are evenly distributed, then cluster membership may be influenced by the initial random seed. To mitigate this source of error, each step of RSKC was iterated 100 times. The most common cluster for

each sample over all iterations was chosen as the final cluster membership. If a sample did not fall into a particular cluster for more than 50/100 iterations, indicating that cluster membership was split across multiple clusters, it was labelled "noisy." Feature weights were also averaged across 100 iterations.

## 2.6 Summary

An increase in the precision of investigative tools and the inclusion of more individuals in modern stuides has favoured continuous models of human brain development. Conversely, the current study asks whether the transcriptome can be used to describe development as a discontinuous process. One approach is the identification developmental subclusters using global patterns of gene expression. Previous clustering and dimensionality reduction approach do not find clear separation of developmental periods. A potential solution is sparse clustering, which is specifically suited to data with more features (genes) than observations (samples), and can simultaneously identify features that drive cluster separation. We validate the use robust and sparse k-means clustering (RSKC) against other methods, discuss the selection of *k,* and review steps for mitigating the biases of strong differential expression and seed effects. In the subsequent chapter, RSKC is used to identify developmental clusters and features in primary visual cortex.

# Chapter 3. Developmental Clusters in Primary Visual Cortex

## 3.1 Introduction

To demonstrate the steps of our analysis pipeline, we will focus on a single brain area – the primary visual cortex (V1C). V1C is the first cortical area that processes visual information, and has served as model of critical period plasticity for over 40 years (Hensch, 2005). The physiological changes that take place in V1C over normal and abnormal development have been studied extensively in animal models (Wiesel & Hubel, 1963; Hubel & Wiesel, 1970; Hubel & Wiesel, 1976), but less is known about the neurobiology of the human visual cortex.

While brain imaging studies can address the structural and functional development of the human visual cortex, the lack of information about molecular and cellular mechanisms has limited our ability to address developmental disorders such as amblyopia. Transcriptomic stuides provide a unique opportunity to address this gap. The aim of the current study is to identify and characterize developmental clusters in V1C that are driven by gene expression. This model may help align biological mechanisms with well-know anatomical and visual processing milestones.

## 3.2 Methods

### 3.2.1 Data acquisition and processing

Data for the current study was originally collected by Kang et al. (2011) and later integrated into the Human Brain Atlas (http://hbatlas.org). In brief, the authors measured genome-wide exon-level gene expression in 1,340 sample collected from 57 brains and 16 brains regions:

All tissue specimens were collected from clinically unremarkable donors without history or signs of neurological or neuropsychiatric illness or drug use. The Affymetrix Human Exon 1.0 ST array, which features 1.4 million probe sets, was used to assay exon expression across the entire genome. Outlier samples were detected using hierarchical clustering and Spearman correlation, and removed from further processing steps. Affymetrix CEL files were then processed using a standard approach, involving RMA background correction, quantile normalization, mean probe set summarization, and log2-transformation. A total of 17,656 main protein-coding genes were surveyed.

The preprocessed data from Kang et al. (2011) were downloaded directly from the Gene Expression Omnibus (GEO) under the accession number GSE25219 using the *GEOquery* package (Davis & Meltzer, 2007). All data processing and visualization were performed in the R statistical software. The exon-summarized expression data were extracted, and probe identifiers were matched to genes. If a gene was matched by two or more probes then probe expressions were averaged if highly correlated (Pearson correlation, $r \geq 0.9$).

### 3.2.2 Analysis of neuroanatomy

In the present chapter, we investigate the histological localization of *VAMP1* in coronal slices of the visual cortex. Briefly, full resolution *in situ* hybridization images of *VAMP1* in visual cortex were downloaded from the Allen Human Brain Atlas (Hawrylycz et al., 2012). Specimen H0-0069.01.02 (22 years, male, right hemisphere, African American) was selected for good staining as well as distinct gyral and sulcal landmarks. The three most anterior slices were imported into Photoshop CC (2015), where a 1500 x 4000 μm section spanning six layers of the neocortex, beginning at the pial surface and ending in the white matter, were sampled from the upper and lower bank of the calcarine fissure. A total of six samples were imported into ImageJ (Eliceiri, 2017) and processed through the following steps: 1) conversion to 8-bit format, 2) background subtraction using 50-pixel rolling window, 3) image threshold adjustment to include only stained objects, 4) processing into binary format, and 5) particles analysis using custom upper and lower area bounds. Bounds were set after several measurements of various cells (area = 15 - 800 $\mu m^2$). Particle analysis results were imported into Microsoft Excel (2020), and cells were counted in a rolling window of 75 μm along the y-axis (perpendicular to the pial surface). Measurements were separated by small ($\leq 30\ \mu m^2$) and large ($> 30\ \mu m^2$) cells in order to detect any distinct laminar patterns.

### 3.2.3 Differential gene expression analysis

To unpack which genes are over-expressed and under-expressed within clusters, we implemented a differential expression (DE) analysis using the *limma* package (Ritchie et al., 2015). *Limma* utilizes linear models to analyze microarray experiments. A linear approach at this step is appropriate as the non-linear structure in the data has already been captured by RSKC. Two important inputs for *limma* are the design matrix, indicating which groups samples belong to, and the contrast matrix, which specifies the comparisons to make between samples. In the

current study, the design was informed by RSKC cluster membership and the contrasts were set such that each cluster was compared to all other clusters pooled together. The automated steps in a *limma* analysis are as follows: 1) compute the pool variance by computing within sample variance for each gene and the correlation among samples within a cluster, 2) create a coefficient matrix for the contrasts, 3) compute the moderated contrast t-statistic for each gene, and 4) identify significant genes based on p-values after adjustment for multiple comparisons.

### 3.2.4 Cell-type analysis

Nearly all developmental transcriptomic stuides of the human brain use whole tissue, and RNA expression from all cells are analyzed *en masse.* Kelley et al. (2018) reasoned that when many intact samples are analyzed together, genes expressed with the greatest sensitivity and specificity in the same cell class should appear highly correlated, as their expression levels depend on the proportion of that cell class in the homogenate. Accordingly, the authors aggregated single-cell RNA-seq data from the adult human brain to create synthetic samples representing the heterogeneity of intact brain tissue, and used unsupervised gene co-expression analysis to identify modules that were enriched for published marks of astrocytes, oligodendrocytes, microglia, and neurons (Kelley et al., 2018). The first principal component (PC1) of these modules was used to estimate the relative abundance of each cell class over all samples and calculate $k_{ME}$, which is the Pearson correlation between the expression pattern of a gene and PC1. The $k_{ME}$ values for all significant cell-class modules across multiple datasets were combined to create a single z-score for each gene that measures its global expression fidelity for each cell class. Fidelity quantifies the extent to which a gene's expression level are correlated with the inferred abundance of a cell type, where a higher score indicates high sensitivity and specificity for that cell type

Scores generated using data from the occipital cortex were downloaded and subsetted to include only the genes available in the current study. For each developmental cluster, a one-tailed Fischer's exact test was used to identify a significant overlap of DE genes with genes in the top 95$^{th}$ percentile of fidelity scores per cell type. The Benjamini-Hochberg procedure was used to adjust p-values for multiple comparison within a cell-type.

## 3.3 Results

### 3.3.1 A small selection of genes separates prenatal and postnatal samples

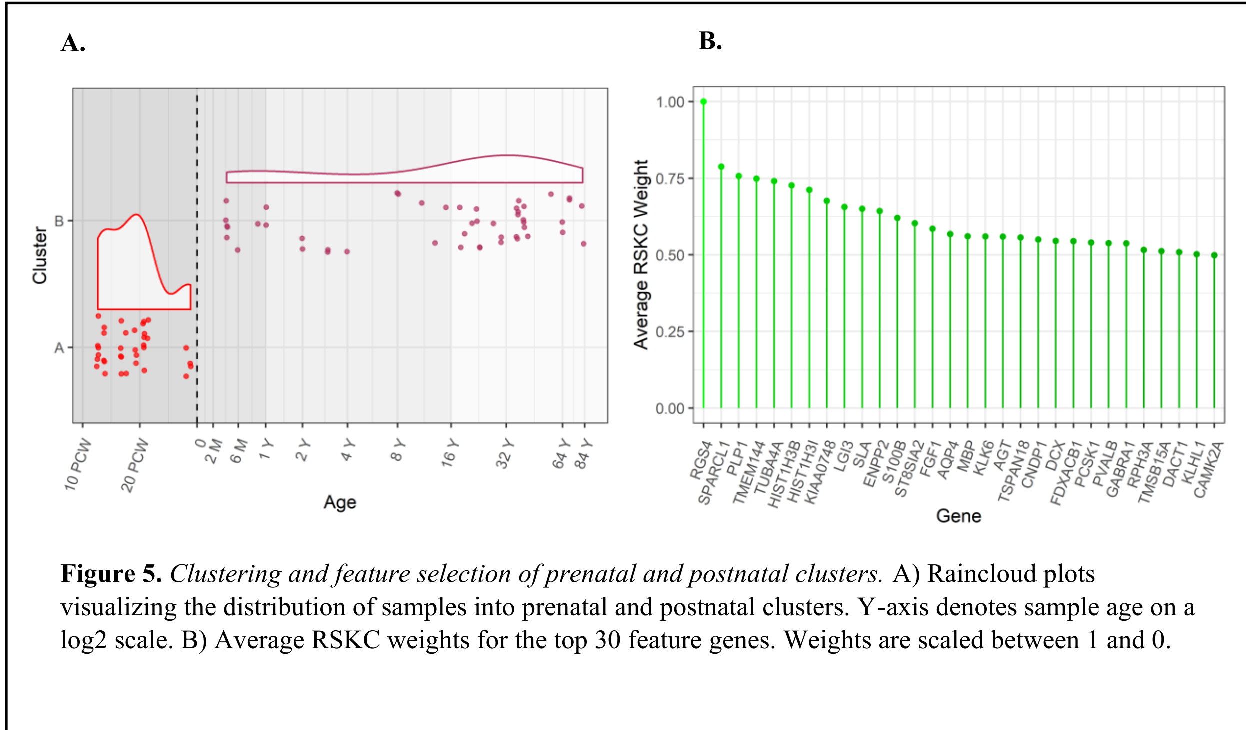


**Figure 5.** *Clustering and feature selection of prenatal and postnatal clusters.* A) Raincloud plots visualizing the distribution of samples into prenatal and postnatal clusters. Y-axis denotes sample age on a log2 scale. B) Average RSKC weights for the top 30 feature genes. Weights are scaled between 1 and 0.

Robust and sparse k-means clustering (see section 2.2) correctly separates all prenatal and postnatal samples based on gene expression (Figure 5A). Figure 5B visualizes the feature weights for the top 30 genes that drive the separation of prenatal and postnatal samples. Notably, *RGS4* (regulator of G protein signalling 4) has a transient phase of neuronal expression during embryonic development that overlaps substantially with *Phox2b,* and is thought to be involved in the type-specific program of neuronal differentiation (Grillet et al., 2003). *SPARCL1* (SPARC like 1), better known as *hevin*, is highly expressed during embryogenesis, where it regulates extracellular matrix (ECM) organization and cell-ECM interactions, which are crucial for cell migration, survival, and the inhibition of cell proliferation (Vincent et al., 2008). Intriguingly, both *RGS4* and *SPARCL1* exhibit greater expression postnatally than prenatally in the current study (Figure 6). This finding is more consistent with reports that RGS4 expression peaks in rat neocortex after P2 (postnatal day 2) and is dynamically regulated in the hippocampus and thalamus, suggesting roles in experience-dependent brain development (Ingi & Aoki, 2002).

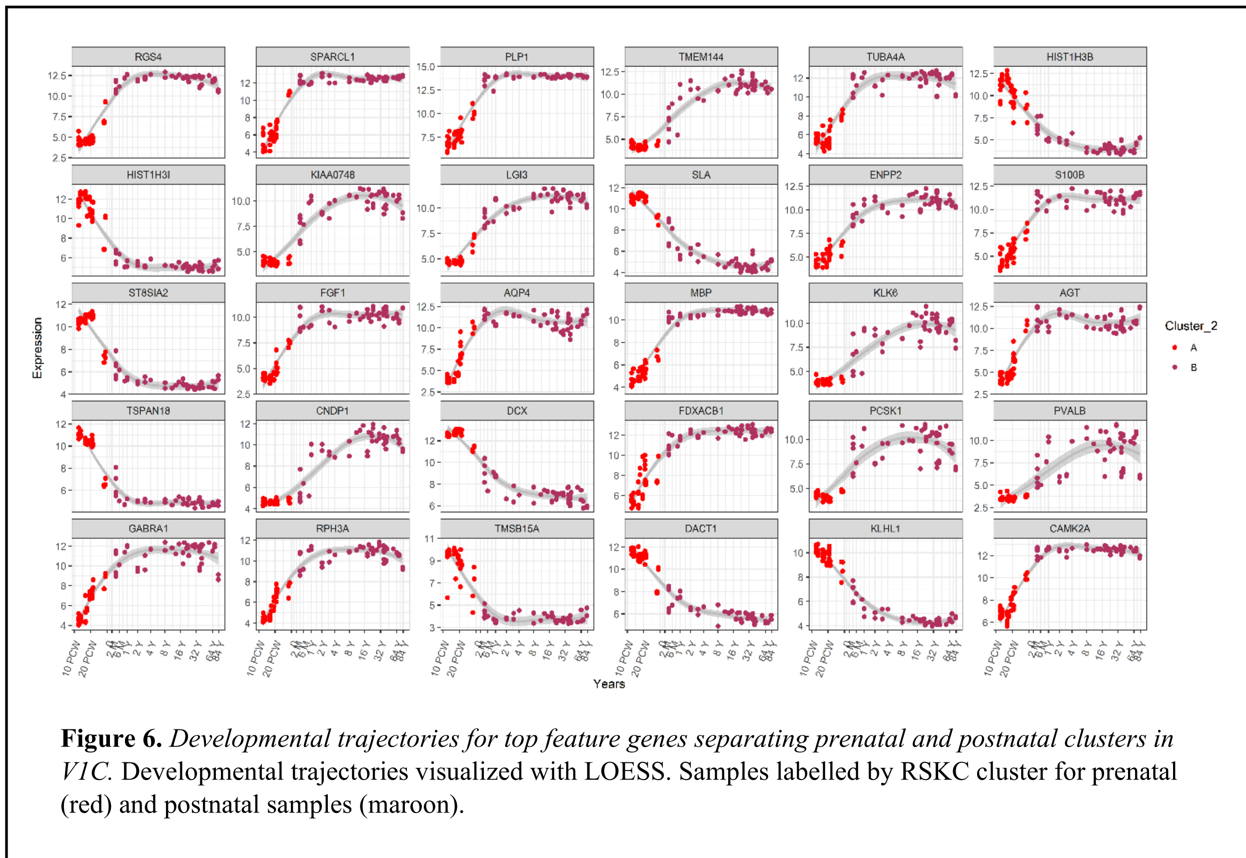


**Figure 6.** *Developmental trajectories for top feature genes separating prenatal and postnatal clusters in V1C.* Developmental trajectories visualized with LOESS. Samples labelled by RSKC cluster for prenatal (red) and postnatal samples (maroon).

Similarly, astrocyte-secreted hevin is required for the establishment of thalamocortical connections in the mouse cortex (Rischer et al., 2014).

A concomitant increase in *PLP1* (proteolipid protein 1) and *MBP* (myelin basic protein) expression was observed between prenatal and postnatal development (Figure 6). A developmental increase of intra-cortical myelin is thought to act as a structural brake on critical period plasticity in V1C (Bavelier et al., 2010). Myelin signaling inhibits experience-driven neurite growth through various myelin-associated proteins, including Nogo, MAG and OMgp (Wang et al., 2002; McGee et al., 2005; Akbik et al., 2012), and knockout of the Nogo receptor prolongs ocular dominance plasticity in V1C (McGee et al., 2005). Interestingly, expression of *PLP1* and *MBP* reach adulthood levels as early as one year of age in the current study. This is well before the peak expression of MBP protein at 42 years of age in human V1C (Siu et al., 2015), the protracted development of myelinated fibers (Miller et al., 2011), and the end of susceptibility to amblyopia at 7 years of age (Homes et al., 2011). One possible explanation is that *MBP* is expressed constantly through prenatal development and adulthood, and that

regulation of MBP occurs post-transcriptionally. This notion is supported by the fact that Classic- and Golli-MBP each have multiple isoforms and post-translation modification pathways (Harauz et al., 2009; Harauz & Boggs, 2013).

Other notable feature genes include *TUBA4A* (tubulin alpha 4a), which is essential for the generation, migration, and differentiation of neurons (Breuss et al., 2017). *SLA,* also known as *SLAP1,* is selectively expressed in neurons once they migrate to deep layers of the cortex, suggesting a role in axon guidance (Marton et al., 2015). *ENPP2* (ectonucleotide pyrophosphatase/phosphodiesterase 2) synthesizes lysophosphatidic acid, which is crucial for embryonic axis formation in zebrafish (Frisca et al., 2016). *S100B* (S100 calcium binding protein B) is a peptide produced by astrocytes that exerts paracrine and autocrine effects on neurons and glia. S100B stimulates neurite outgrowth and enhances the survival of neurons during development, while excess amounts simulate the expression of proinflammatory cytokines and induce apoptosis (Rothermundt et al., 2003). *ST9SIA2* (ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 2) synthesizes polysialic acid, which is implicated in multiple processes, including synaptic plasticity, migration of neural progenitors, and synaptogenesis (Ikegami et al., 2019). Although the embryonic and prenatal functions of many of these gene are the dominant topic in the literature, some show increased expression after birth, suggesting that further investigation of their function in the mature brain may be worthwhile. Additionally, other feature genes such as *TESPA1* (thymocyte expressed, positive selection associated 1), which is required for the development and maturation of T-cells, have not yet been functionally characterized in the brain. Altogether, features selected during the separation of prenatal and postnatal samples by RSKC appear to be biologically relevant. Some are implicated in general cortical development, while others, such as *PLP1* and *MBP,* may be more particular to changes in V1C.

### 3.3.2 Developmental clusters in primary visual cortex

The sub-clustering of prenatal and postnatal clusters revealed novel developmental dynamics in V1C (Figure 7A). Prenatal samples separated into two distinct clusters (1A and 1B), with median ages of 13 and 21 PCW (post-conception weeks) respectively. Conversely, postnatal samples separated into clusters with considerable overlap (2A, 2B, 2C, and 2D), although progression in median cluster age is observed (0.9, 2.0, 30.0, and 37.0 years). Cluster separation

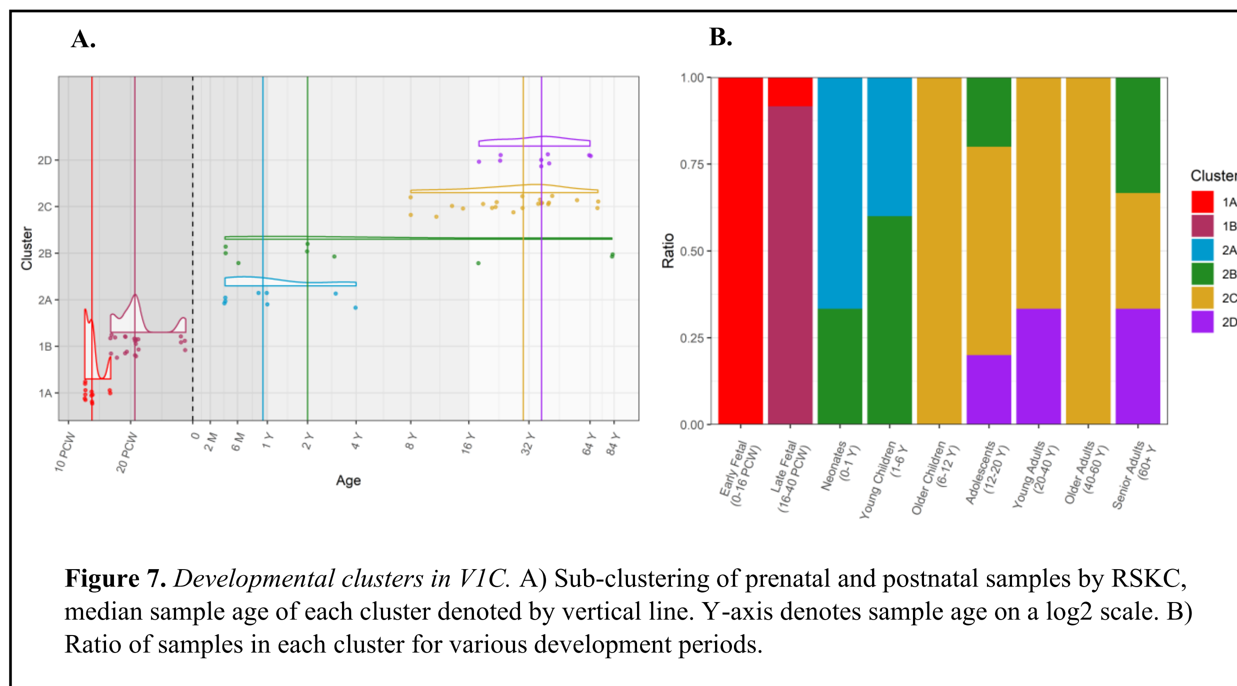


**Figure 7.** *Developmental clusters in V1C.* A) Sub-clustering of prenatal and postnatal samples by RSKC, median sample age of each cluster denoted by vertical line. Y-axis denotes sample age on a log2 scale. B) Ratio of samples in each cluster for various development periods.

was not explicitly driven by sample sex or hemisphere (Figure 8). Figure 7B demonstrates that traditional development periods, defined by anatomical and behavioural milestones (Kang et al., 2011), are not described by a single transcriptomic "phenotype." Instead, developmental periods may be defined by two or more patterns of gene expression across the transcriptome. Such patterns are not limited to a particular section of the lifespan, and may be recurring. For example, the transcriptional profile matching samples from cluster 2B describes early development at 120 days as well as ageing at 82 years. Colantuoni et al. (2011) similarly found that numerous changes in infancy were mirrored by changes in ageing. Samples in clusters 2C and 2D overlap considerably in age, yet have presumably distinct transcriptomic profiles. This may reflect the dynamic systems principle of multiple "solutions" being presented by the brain during late development and adulthood. Further unpacking of these developmental clusters is required.

Figure 9 lists the top genes that contributed to the clustering of postnatal samples and Figure 10 visualizes their respective expression patterns. The latter demonstrates the potential variability in gene expression between samples that are adjacent in developmental time. These differences are averaged in regression models such as LOESS (locally estimated scatterplot smoothing) (Figure 10A), whereas in RSKC they drive the separation of clusters (Figure 10B). An examination of the selected features will reveal whether the developmental variability resolved by RSKC has any biological significance.

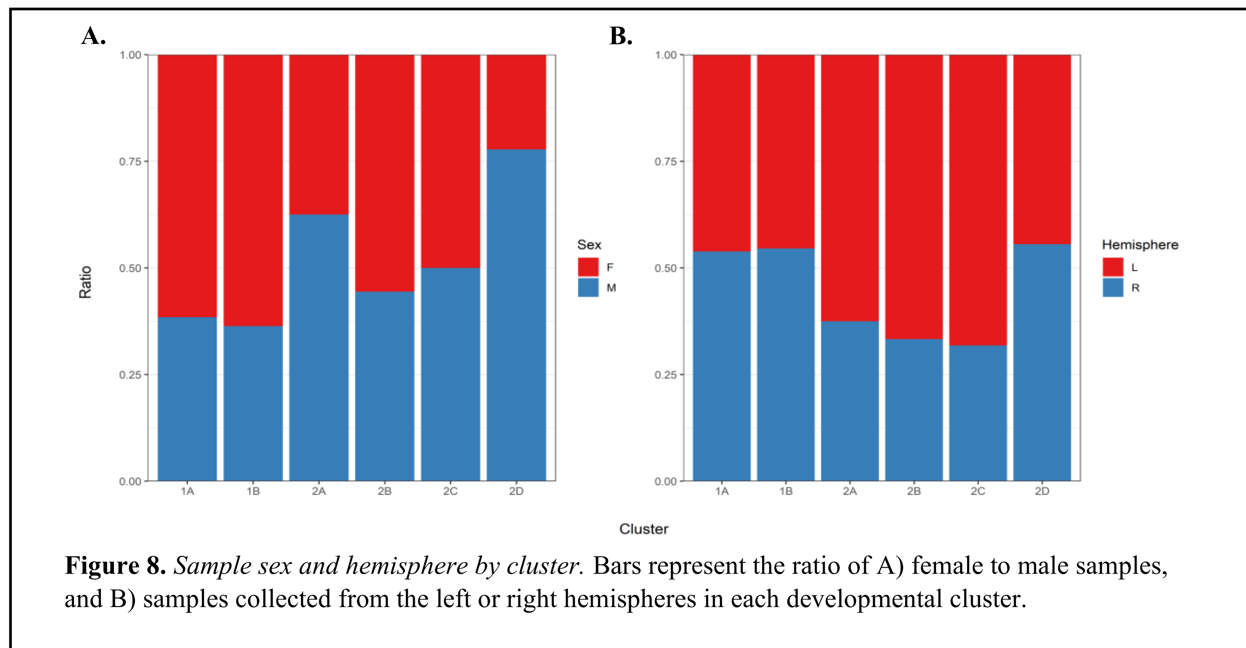


**Figure 8.** *Sample sex and hemisphere by cluster.* Bars represent the ratio of A) female to male samples, and B) samples collected from the left or right hemispheres in each developmental cluster.

Several notable genes were selected as features of postnatal V1C development. *PVALB* (parvalbumin) had the greatest weight among all genes measured. Parvalbumin (PV) is a calcium-binding protein and a marker of fast-spiking GABAergic interneurons, which are crucial in governing feedforward and feedback inhibition in cortical microcircuits, as well as the tight regulation of fast network oscillations (Hu & Jonas, 2014). In V1C, the onset of critical period plasticity is determined by $PV^+$ interneuron maturation (Fagiolini & Hensch, 2000). Moreover, the deposition of perineuronal nets (PNNs) onto $PV^+$ interneurons acts as a break on critical period plasticity (Bradshaw et al., 2018). Accordingly, the disruption of PNNs by chondroitinases restores the ocular dominance shift in adult rats (Pizzorusso et al., 2002). In the current study, *PVALB* expression is relatively low in prenatal clusters but is substantial in cluster 2A (sample ages: 160 days – 4 years). Cluster 2C begins at the end of the period of susceptibility to amblyopia (~8 years) and shows the greatest *PVALB* expression. Another feature gene, *SYT2* (synaptotagmin 2), is known to be co-expressed with *PVALB* (Sommeijer & Levelt, 2012). SYT2 has been identified as a functionally important $Ca^{2+}$ sensor at fast-spiking inhibitory synapses (Bouhours et al., 2017), and shows a similar pattern of expression to *PVALB* in the current study. *STT* (somatostatin) is another feature gene that codes for a neuropeptide marker of a major subtype of GABAergic interneurons. $STT^+$ interneurons are known to inhibit pyramidal neurons via GABA signaling, but can also disinhibit excitatory neurons via the inhibition of $PV^+$ interneurons (Pfeffer et al., 2013). In V1C, the release of STT improves visual perception by

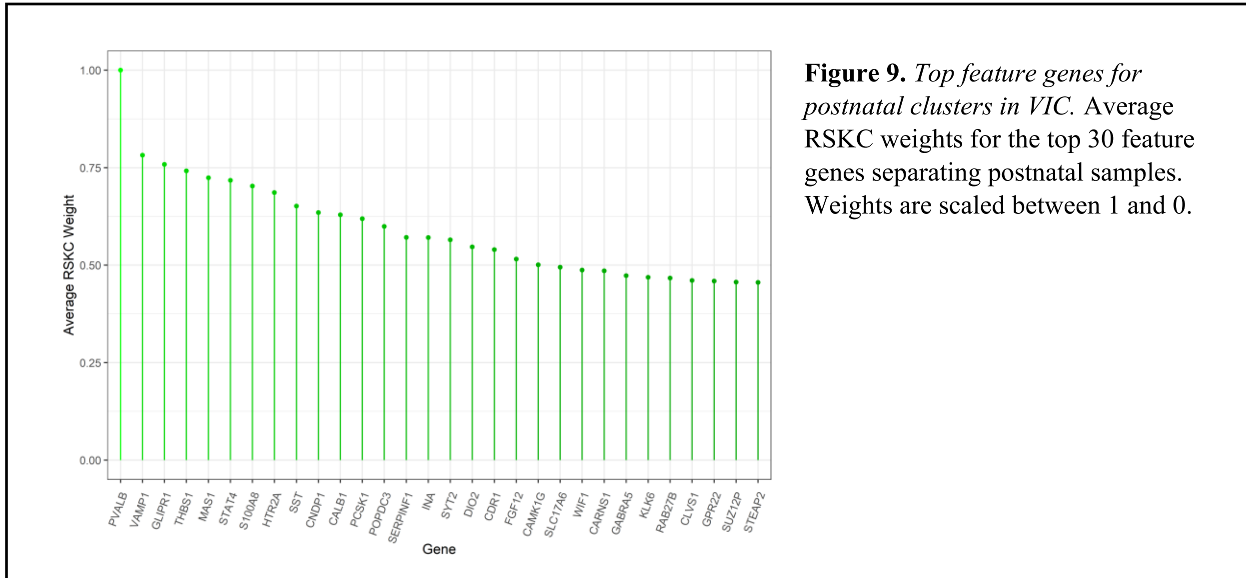


**Figure 9.** *Top feature genes for postnatal clusters in VIC.* Average RSKC weights for the top 30 feature genes separating postnatal samples. Weights are scaled between 1 and 0.

enhancing the gain of visually responsive neurons through the reduction of excitatory transmission to PV$^+$ interneurons (Song et al., 2020). STT$^+$ interneurons also contribute to the surround suppression of pyramidal cells (Adesnik et al., 2012). Studies of cross-modal reactivation have identified STT$^+$ interneurons as regulators of cortical plasticity following sensory loss. Scheyltjens et al. (2018) demonstrate that the stimulation of STT$^+$ interneurons in visual cortex prior to monocular enucleation (ME) decreased functional recovery through spared-eye potentiation and somatosensory activation in adult mice. In the current study, *STT* showed the greatest level of expression in cluster 2A, which spans critical period of visual acuity development in children (birth through 3-5 years of age) (Daw et al., 1998). *HTR2A* (5-Hydroxytryptamine Receptor 2A) codes for the serotonin-2A receptor, which is highly concentrated in the layer IV of macaque V1C and exhibits activity-dependent expression (Watakabe et al., 2009; Nakagami et al., 2013). The administration of HTR2A antagonists has been shown to supress cross-modal plasticity in visual cortex following ME (Lombaert et al., 2018). In the current study, *HTR2A* expression can be either relatively high (clusters 2A and 2C) or low (clusters 2B and 2D) at similar times during the lifespan.

Several feature genes did not have known specialized functions in V1C, but more general roles in the healthy and diseased brain. *GLIPR1* (GLI Pathogenesis Related 1) and *THBS1* (Thrombospondin 1) are both upregulated in glioblastoma (Murphy et al., 1995; Daubon et al., 2019). Specifically, THBS1 is an extracellular matrix (ECM) protein secreted mainly by astrocytes that is involved in regulating angiogenesis, synaptogenesis, and spine development

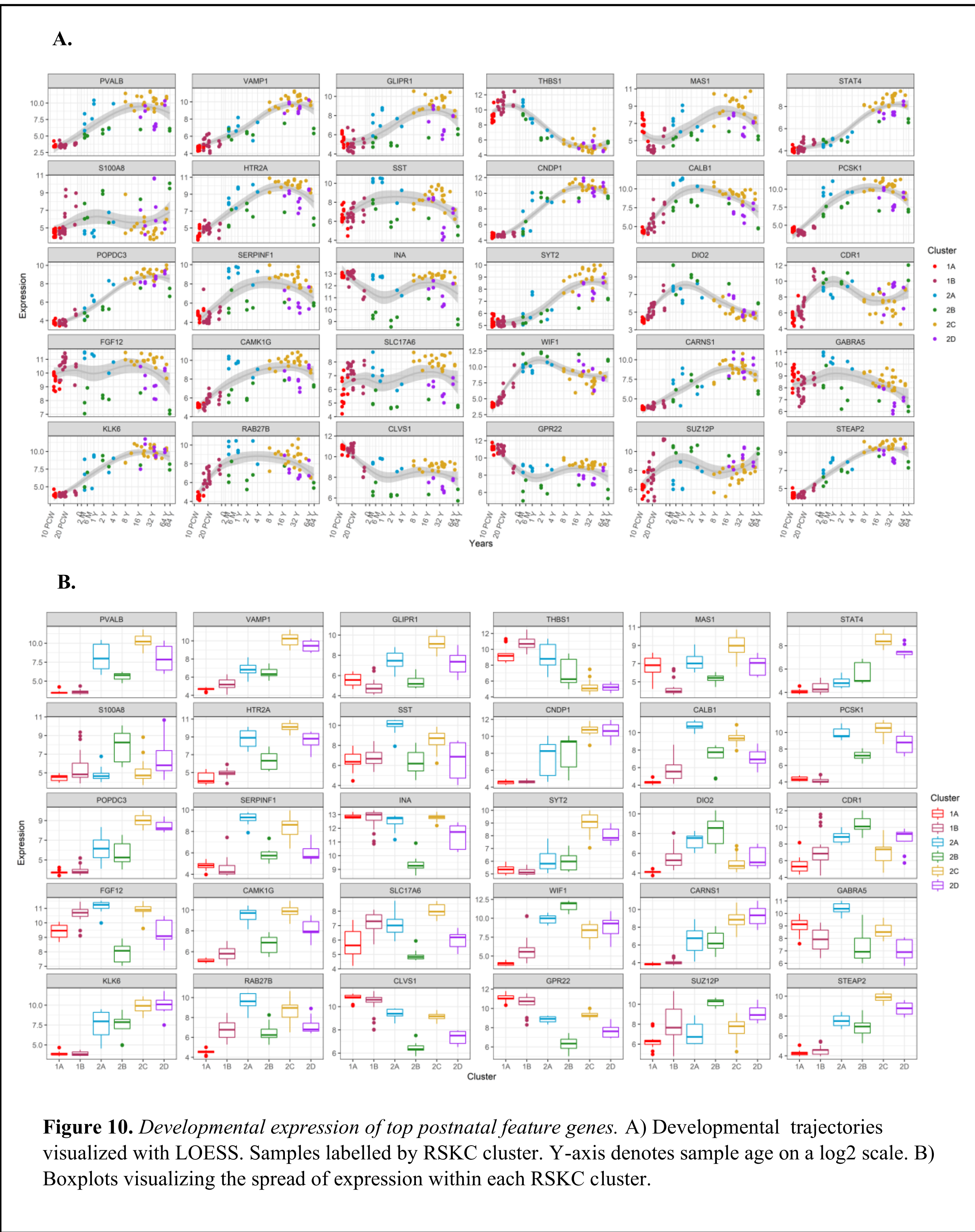


**Figure 10.** *Developmental expression of top postnatal feature genes.* A) Developmental trajectories visualized with LOESS. Samples labelled by RSKC cluster. Y-axis denotes sample age on a log2 scale. B) Boxplots visualizing the spread of expression within each RSKC cluster.

(Cheng et al., 2017; Cheng et al., 2016). *MAS1* (MAS1 Proto-Oncogene, G Protein-Coupled Receptor) is expressed in microglia and activates anti-inflammatory functions, as well as the regulation of angiogenesis (Foulquier et al., 2019). *S100A8* (S100 Calcium Binding Protein A8) codes for one of the most prominently expressed S100 proteins in the brain, and is upregulated by ageing and neuronal damage (Hagmeyer et al., 2019). S100 proteins play many roles in inflammation and immune response, however, S100A8/A9 are largely characterized by their ability to undergo amyloid oligomerization and fibrillation *in vitro*, linking them to the pathogenesis of Alzheimer's disease (AD) (Fritz et al., 2010). The gene product of *CALB1* (Calbindin 1) is one of the major calcium-binding and buffering proteins in the central nervous system (CNS), has critical roles in calcium homeostasis and neuroprotection, and is also a common marker of neuronal populations. Removal of CALB1 expression in transgenic mice susceptible to AD aggravated the pathogenesis of the disorder (Kook et al., 2014). Developmental stuides have found a significant loss of CALB1 in the hippocampus and striatum with ageing (Kishimoto et al., 1998). Additionally, immunocytochemical investigation of CALB1 in the mouse visual cortex revealed localization to stellate cells in layers II/II and V (Park et al., 2002). *INA* (Internexin Neuronal Intermediate Filament Protein Alpha), better known as α-Internexin, is the first neuronal intermediate filament (IF) expressed in neurons after they are committed to the neuronal lineage and then decreases as neurons mature (Lennarz & Lane, 2013). While the physiological roles of α-Internexin remain undefined, they are suspected to contribute to late-onset dementia characterized by inclusions of neuronal IF proteins (Yuan et al., 2006). *DIO2* (Iodothyronine Deiodinase 2) encodes a protein that catalyzes the prohormone thyroxine (3,5,3',5'-tetraiodothyronine, T4) to the bioactive thyroid hormone (3,5,3'-triiodothyronine, T3). Thyroid hormones such as T3 are important for brain development at multiple time windows influencing neurogenesis, neuronal migration, neuronal and glial differentiation, myelination, and synaptogenesis, and the amount of T3 the brain receives is regulated in part by DIO2 in astrocytes (Bernal, 2000). Interestingly, the inflow to T3 converted by DIO2 determines the start of the sensitive period for filial imprinting in chicks (Yamaguchi et al., 2012). *FGF12* (Fibroblast Growth Factor 12) is a member of the non-secreted intracellular FGF family. FGF12 forms a complex with voltage-gated sodium channels at a subcellular domain specialized for action potential initiation, and increases channel availability (Wildburger et al., 2015). Recently, a *de novo* mutation in FGF12 has been reported in some cases of epileptic

encephalopathies (Shi et al., 2017). Lastly, *GABRA5* (Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha5) encodes one of the 16 distinct $GABA_A$ receptors subunits. In visual cortex, $GABA_A$ receptor subunit composition regulates synaptic plasticity. Namely, juvenile synapses are dominated by more $GABA_A\alpha3$ containing receptors that allow for long-term potentiation (LTP) in excitatory synapses and slower kinetics through the receptors. More mature synapses are dominated by $GABA_A\alpha1$ containing receptors that allow for more long-term depression (LTD) in excitatory synapses and faster kinetics through the receptors (Pinto et al., 2010). The role $GABA_A\alpha5$ is not specifically mentioned in studies of visual cortex plasticity. However, it is expressed abundantly in the hippocampus, primarily populating extrasynaptic sites in CA1 pyramidal cells where it generates a tonic inhibitory conductance (Magnin et al., 2019). $GABA_A\alpha5$ is also a target for anesthetic drugs (Martin et al., 2009). Interestingly, $GABA_A\alpha5$ null mutant mice (*GABRA5*−/−) exhibit a normal behavioral phenotype for contextual learning (Cheng et al., 2006) and normal LTP (Martin, 2010).

Genes that drove the clustering of postnatal samples in V1C overlapped with several genes involved in the development and function of the visual cortex. The top weighted genes included markers of distinct inhibitory interneuron subpopulations, which are crucial for the regulation of critical period plasticity and the normal functioning of visual circuits. Other feature genes had known roles in inflammation and immune responses. Recent studies have demonstrated that several immune molecules are endogenously expressed in the CNS, where they are essential for the establishment, function, and modification of synaptic connections (Boulanger, 2009; Garay & McAllister, 2010). Lastly, some features were associated with cancer and neurodegenerative disorders. In sum, RSKC is presents a new way of identifying developmentally regulated genes of interest. Other approaches, such as co-expression analysis, simply detect groups of genes that change together – no group is privileged over another, and they must be manually selected for further analysis. Feature selection is built into the RSKC algorithm, and presents an unsupervised method of identifying genes with the greatest developmental variance without the linear assumption of other dimensionality reduction approaches (e.g. PCA).

### 3.3.3 Further investigation of feature genes: *VAMP1*

The release of neurotransmitters (NTs) in the brain is mediated by the fusion of synaptic vesicles with the plasma membrane at the presynapse. Three proteins from the SNARE (soluble N-ethylmaleimide-sensitive factor attachment receptor) family are responsible for membrane fusion. Two t-SNAREs anchored in the plasma membrane, syntaxin and SNAP-25 (synaptosomal-associated protein 25), form a four-α-helix bundle with one vesicle anchored v-SNARE, synaptobrevin (Sørensen et al., 2006). The major v-SNARE in the mammalian brain is synaptobrevin 2 (syb2), also known as VAMP2 (vesicle-associated membrane protein 2) (Zimmermann et al., 2014). In addition to its role in exocytosis, VAMP2 has been implicated in the insertion of GluA1 (glutamate ionotropic receptor AMPA type subunit 1) receptors into the spine plasma membrane, which is crucial for LTP (Hussain & Davanger, 2015). Trimble et al., (1990) reported that VAMP1 and VAMP2 have distinct, but slightly over-lapping, patterns in the rat brain. VAMP2 expression was more ubiquitous, and found in neurons associated with autonomic, sensory, and integrative roles. VAMP1 expression was largely restricted to the spinal cord and a number of neurons involved in somatomotor functions (Trimble et al., 1990). A recent study investigating cultures of hippocampal neurons from VAMP2 knockout mice found that the majority of cells were devoid of any evoked or spontaneous NT release and had no measurable readily releasable pool (Zimmermann et al., 2014). However, a small subpopulation of neurons expressing VAMP1 demonstrated NT release. Subsequent rescue experiments found that VAMP1 can substitute VAMP2, albeit with a lower efficiency in NT release probability. Moreover, NT release in a culture of VAMP2-deficient neurons was significantly reduced following a knockdown of VAMP1, suggesting that VAMP1 is responsible for the remaining release activity (Zimmermann et al., 2014).

The current study identified *VAMP1* as the second highest weighted gene contributing to the formation of postnatal developmental clusters in V1C (Figure 9). Furthermore, *VAMP1* expression increases gradually across the lifespan and across clusters (Figure 10). This v-SNARE has not been explicitly studied in the visual cortex. VAMP1 may represent a functionally relevant gene of interest, given the role of v-SNAREs in receptor cycling (Hussain & Davanger, 2015) and the importance of AMPA receptor trafficking to homeostatic synaptic plasticity in visual cortex (Murphy et al., 2012).

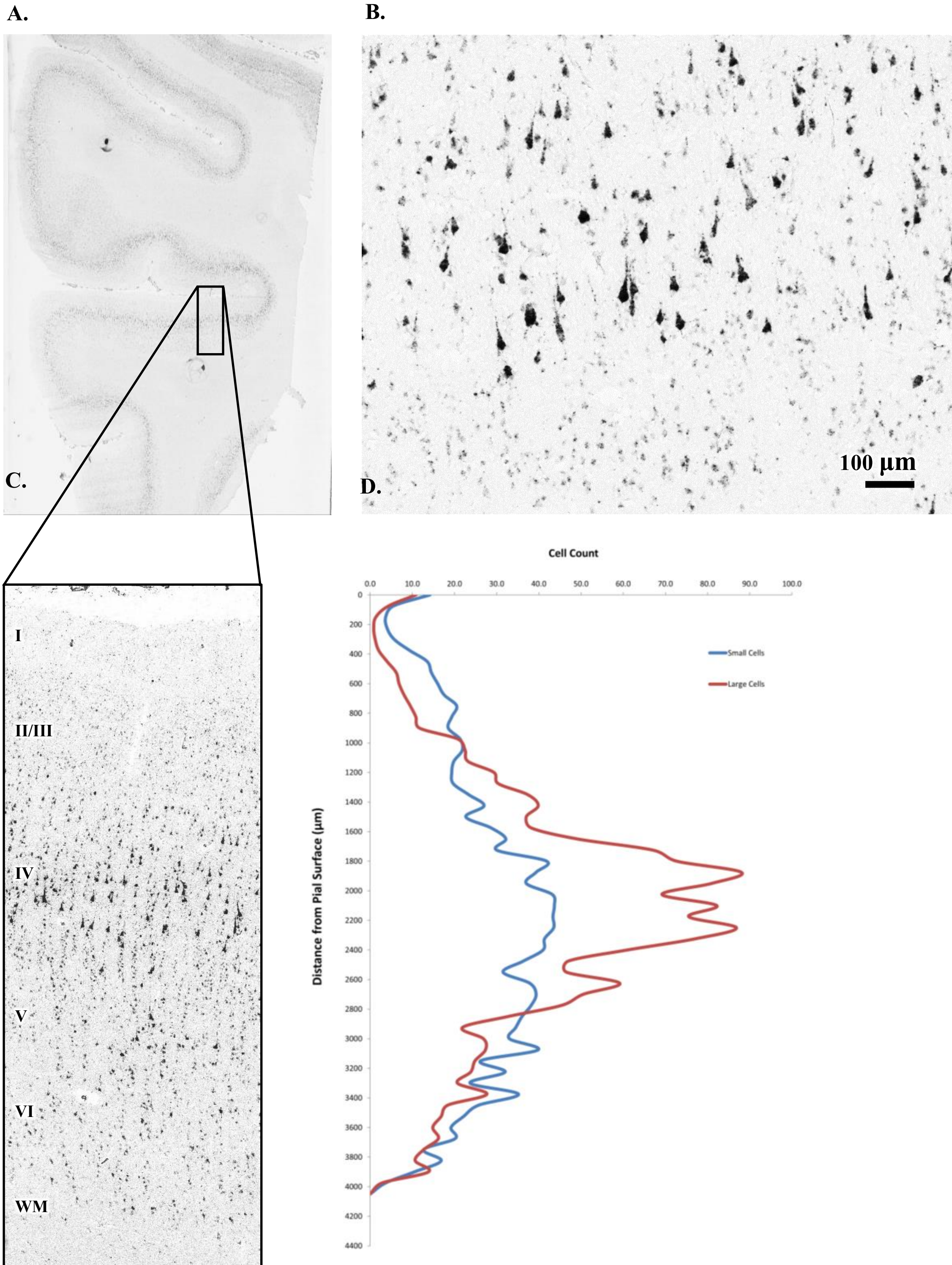


**Figure 11.** *In situ hybridization of VAMP1 in visual cortex.* A) Coronal slice of striate cortex obtained from Allen Human Brain Atlas (Hawrylycz et al., 2012). B) Cellular morphology revealed by *VAMP1* hybridization, scale indicated bottom right. C) Laminar pattern of expression, 4000 µm deep from pial surface. D) Cells counts over laminar depth, averaged over 6 samples for small (blue) and large (red) cells.

*In situ* hybridization validates the expression of VAMP1 in human V1C. Analysis of cell counts revealed a subpopulation of cells smaller than 30 $\mu m^2$ with a relatively constant presence across cortical layers. These may represent nonneuronal cells or neuronal spines on apical dendrites. The majority (52.3%) of large cells (> 30 $\mu m^2$) expressing VAMP1 were found in the granular layers (1650 – 2400 µm from pial surface). In the architecture of the sensory neocortex, layer IV is the main target of sensory inputs coming from the thalamus (Scala et al., 2019). The spiny morphology of these cells is consistent with excitatory pyramidal neurons (Figure 10B). Based on findings from Zimmermann et al. (2014) in hippocampus, VAMP1 may be responsible for NT release from pyramidal neurons in V1C. Intriguingly, VAMP1 does not appear to label non-spiny cells corresponding to inhibitory interneurons in layer IV. This may reflect the restricted expression of VAMP1 to specific subpopulations, as seen in other studies (Trimble, 1990). A more far-reaching possibility is the involvement of VAMP1 in GluA1 receptor trafficking (as seen in VAMP2), which is more specific to excitatory neurons and necessary for synaptic plasticity (Kooijmans et al., 2014; Malenka et al., 1989).

### 3.3.4 Differential gene expression analysis

Lists of significantly over-expressed (positive log-fold change relative to all other clusters) and under-expressed (negative log-fold change relative to all other clusters) genes were determined for each cluster (FDR adjusted p-value < 0.05). Cluster 1A exhibited by far the greatest number of DE genes (Figure 12), consistent with the Kang et al. (2011) finding that most changes in gene-expression occurred during early prenatal development. This large effect biases the analysis towards identifying genes that are differentially expressed relative to samples in cluster 1A. Accordingly, all two-way comparisons made between clusters included only those samples in the DE analysis.

Cluster 2A was used to validate the results of the DE analysis. The age of samples in this developmental cluster range from 160 day to 4 years, which encompasses the emergence and maturation of binocular fusion, stereopsis, spatial acuity, contrast selectivity, and orientation selectivity in children (Siu & Murphy, 2018). Figure 13A lists the top 40 over- and under-expressed genes by log-fold change (adjusted $p < 0.05$), and Figure 13B show the expression levels of the top genes by adjust p-value. The latter clearly demonstrates the success of *limma* for identifying genes that are differentially expressed relative to other developmental clusters. One

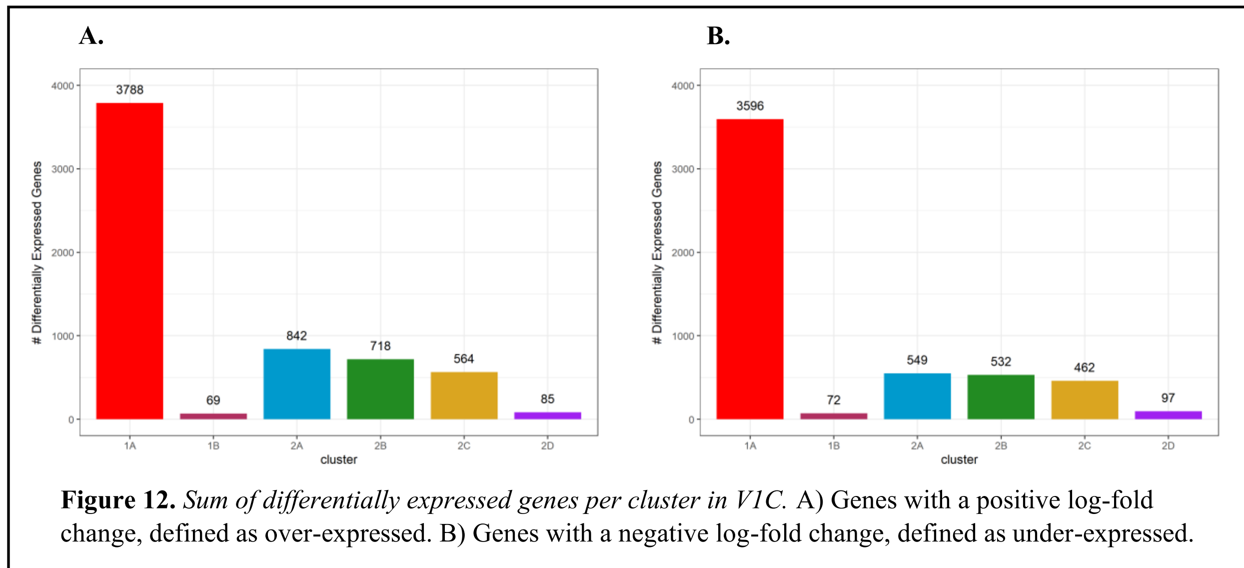


**Figure 12.** *Sum of differentially expressed genes per cluster in V1C.* A) Genes with a positive log-fold change, defined as over-expressed. B) Genes with a negative log-fold change, defined as under-expressed.

of the top over-expressed gene in cluster 2A, PDGFB (Platelet Derived Growth Factor Subunit B) (logFC = 0.7485, adjusted p = 0.0039), is a crucial neuromodulator. PDGFRB, a receptor for the PDGFB ligand, colocalizes with both presynaptic synaptophysin and post-synaptic density-95 (PSD-95) in the adult mouse hippocampus (Shioda et al., 2012). The application of PDGFB to slices from this area produces a long-lasting inhibition of NR2B-containing NMDA receptor currents and enhances LTD in an NR2B subunit-dependant manner (Funa & Sasahara, 2014). PDGFRB also binds to Na+/H+ exchanger regulatory factors (NHERFs), which are scaffold proteins distributed in dendritic spines and in axon terminals of hippocampal pyramidal neurons. Accordingly, PDGFRB crucially contributes to actin reorganization and may regulate dendrite spine morphogenesis and plasticity, both through PDGFB-dependent and independent manners (Svitkina et al., 2010; Funa & Sasahara, 2014).

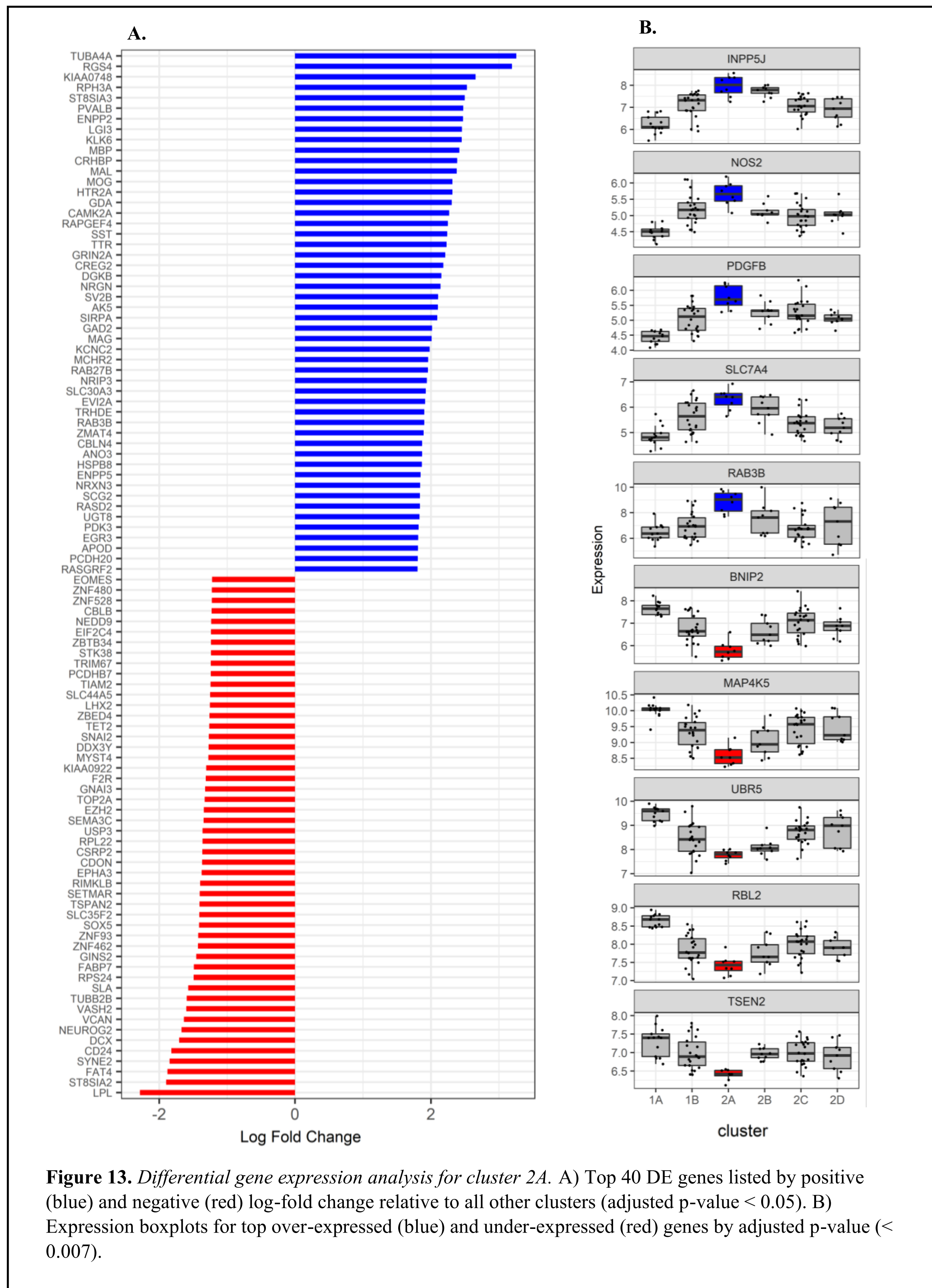


**Figure 13.** *Differential gene expression analysis for cluster 2A.* A) Top 40 DE genes listed by positive (blue) and negative (red) log-fold change relative to all other clusters (adjusted p-value < 0.05). B) Expression boxplots for top over-expressed (blue) and under-expressed (red) genes by adjusted p-value (< 0.007).

### 3.3.5 Cell-type analysis

Identifying DE genes within each cluster allows for the identification of biological features associated with those clusters. Brain tissue is highly heterogeneous, composed of diverse cell types characterized by distinct patterns of gene expression. The current study utilized *fidelity scores* from Kelley et al. (2018) to identify the enrichment of astrocytes, oligodendrocytes, microglia, and neurons in developmental clusters. Fidelity quantifies the extent to which a gene's expression level are correlated with the inferred abundance of a cell type, where a higher score indicates high sensitivity and specificity for that cell type.

Fidelity scores for all over-expressed and under-expressed genes in each developmental cluster were collected and their distributions visualized with violin plots (Figure 14). A cluster with a greater affinity for a particular cell type will have a top-heavy distribution, indicating more DE genes with high fidelity scores. In each cluster, a significant overlap between DE genes and genes with the highest fidelity for each cell type (top $95^{th}$ percentile) is represented with a grey box ($p < 0.05$). This denotes a significant enrichment of cell-type specific markers within the list of DE genes for that cluster. For example, we find that cluster 2B demonstrates an upregulation of astrocyte- and oligodendrocyte-specific genes. On the other hand, cluster 1A saw a significant downregulation of astrocyte- and oligodendrocyte-specific genes.

A striking outcome of the current cell-type analysis is the significant enrichment of genes specific to microglia in developmental clusters 1B and 2A (Figure 14A). Microglia are the resident immune cell of the CNS and have well-known roles in disease and response to injury, including inflammation and the clearance of cellular debris (Hanisch & Kettenmann, 2007). However, recent studies suggest that microglia may also play a role in synaptic sculpting and plasticity in the healthy brain. Microglial cells enter the developing brain early, preceding neurogenesis, neuronal migration, and gliogenesis (Bystron et al., 2008). Wakeselman et al. (2008) demonstrated that transgenic mice deficient in microglia-expressed molecules essential for cell adhesion and phagocytosis saw a significant reduction in neuronal apoptosis in neonatal hippocampus. Additionally, Cunningham et al. (2013) pharmacologically "inactivated" microglia with broad-spectrum tetracyclines and observed an increase in neuronal progenitor cells (NPCs). Both stuides suggest that microglia may actively regulate gestational neurodevelopment by inducing apoptosis of NPCs followed by phagocytosis (Schafer & Stevens, 2015). Similar to the

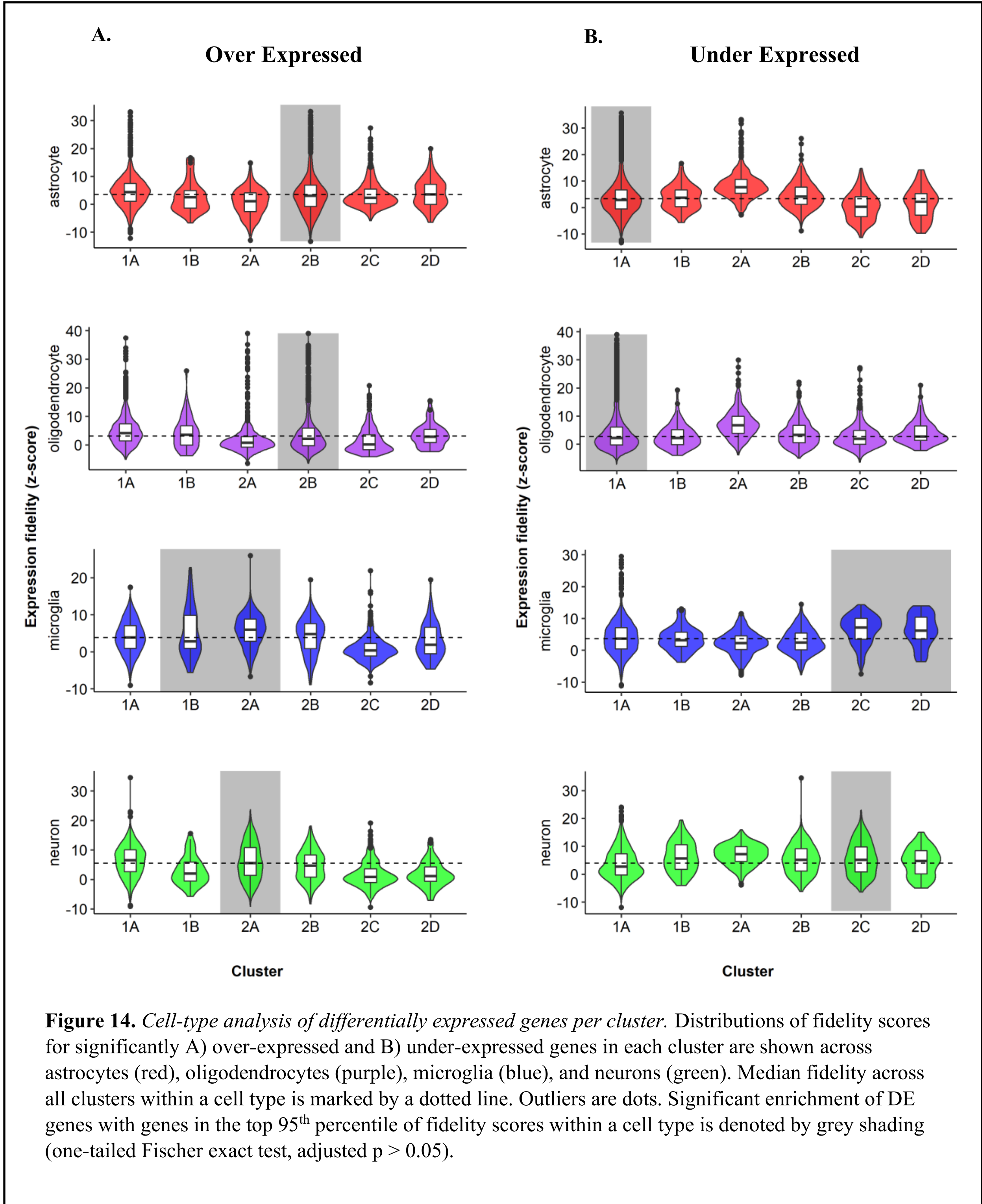


**Figure 14.** *Cell-type analysis of differentially expressed genes per cluster.* Distributions of fidelity scores for significantly A) over-expressed and B) under-expressed genes in each cluster are shown across astrocytes (red), oligodendrocytes (purple), microglia (blue), and neurons (green). Median fidelity across all clusters within a cell type is marked by a dotted line. Outliers are dots. Significant enrichment of DE genes with genes in the top 95$^{th}$ percentile of fidelity scores within a cell type is denoted by grey shading (one-tailed Fischer exact test, adjusted $p > 0.05$).

excess of neurons born in early development, many excess synaptic connections are pruned by phagocytic microglia, producing precise synaptic circuitry. Recent stuides also suggest that microglia are involved in experience-dependent synaptic plasticity. The first evidence came from observations of microglia with "active" morphologies in brain regions undergoing synaptic remodeling, including the cerebellum, hippocampus, and visual system (Maślińska et al., 1998; Dalmau et al., 1998; Schafer et al., 2012). Tremblay et al. (2010) combined three-dimensional electron microscopy and two-photo live imaging to observe microglia in mouse V1C during the critical period. They found that microglia appeared to contact synaptic spines – often those smaller in size and that were no longer visible in subsequent imaging sessions – suggesting that they were being selectively eliminated. Following a dark adaptation paradigm, microglia expanded their processes and increased frequency of contact with synaptic clefts (Tremblay et al., 2010). Schafer et al. (2012) further tested the hypothesis that microglia prune exuberant synaptic connections in the mouse retinogeniculate system. The authors fluorescently labeled presynaptic inputs within the dorsal lateral geniculate nucleus (dLGN) from each eye. During the height of activity-dependent remodeling, microglia contained fluorescent presynaptic cell inclusions within their processes and lysosomal compartments. When competition between the two eyes was enhanced by pharmacologically inhibiting one with tetrodotoxin (TTX), microglia preferentially phagocytized inputs from the eye in which neuronal activity was decreased (Schafer et al., 2012). Microglia can also influence synapses indirectly though modification of the extracellular matrix (ECM). Studies have demonstrated that ECM modifiers such as tissue plasminogen activator (tPA) are likely responsible for generating a permissive environment for enhanced spine motility during ocular dominance (OD) plasticity (Oray, Majewska, & Sur, 2004; Mataga, Mizuguchi, & Hensch, 2004). Crucially, this plasticity occurs in a lamina-specific manner and is limited to the binocular zone where competition between inputs occurs during the critical period. The cellular source of ECM modifiers and their ability to achieve spatiotemporal specificity remains unclear, but microglia provide a likely solution. By targeting specific synaptic locations with their processes and modifying extracellular geometry locally, microglia can enable the compartmentalization of molecules secreted by neuron, astrocytes, and microglia themselves (Trembley & Majewska, 2011). A prime example of multi-cellular interactions at this "tetrapartite" synapse is the secretion of TGF-β (Tumor Growth Factor-β) by astrocytes which induces the expression of classical complement cases protein C1q on neurons. This serves as an

"eat-me" signal to local microglia, which opsonize the synapse with C3 (complement component 3) and perform receptor-mediated phagocytosis via the C3R (C3 receptor) to eliminate the synapse (Cowan & Petri, 2018). In sum, the enrichment of microglia-specific genes in clusters 1B and 2A is consistent with the prenatal regulation of cell survival (cluster 1B) and experience-dependent synaptic plasticity (cluster 2A) by microglia in V1C.

Genes significantly under-expressed in cluster 1A were enriched for oligodendrocytes. This is reflective of the appearance of oligodendrocyte precursors at 22 PCW and commencement of myelination at 28 PCW succeeding the age of the oldest cluster member (16 PCW) (Rezaie & Male, 1999; Jakovcevski et al., 2009). An interesting finding was the overlap between microglia-specific genes and significantly depleted genes in the two oldest clusters, 2C and 2D (mean sample age: 32.1 and 39.2 years). While there is considerable heterogeneity of sample age in both clusters, they account for 83 % of samples older that 40 years (Figure 7B). Microglia are the cell population most associated with the protection of neurons in the healthy brain and their destruction in disease, signifying their potential to mediate age related changes (Angelova & Brown, 2019). As discussed previously, there is evidence that microglia constantly survey neuronal states (Trembley et al., 2010). Microglia sense neuronal activity through the release of ATP, glutamate, and other neurotransmitters (Hickman et al., 2013). This information is used maintain cellular homeostasis, either through physical means such as phagocytosis and remodeling of the extracellular matrix, or the release of signalling molecules (York et al., 2018). For example, in the absence of neuronal activity, microglia release TNF-α (Tumor Necrosis Factor-α) to induce rapid integration of AMPA glutamate receptor isoforms expressing GluA1 subunits and lacking GluA2 into the plasma membrane, leading to enhanced Ca2+ conductivity and excitatory activity (Pribiag & Stellwagen, 2014; Levin & Godukhin, 2017). If signs of damage are detected, microglia migrate to the site of injury and transition from their ramified surveillance state to an activated state characterized by thickened processes and secretion of cytokines and reactive oxygen species (ROS) (Colton, 2009). There is also evidence that microglia can support the re-myelination of damaged axons following injury (Lloyd & Miron, 2019). Microglia are very long-lived cells, and widespread changes in microglia are reported in ageing (Weinberg, 2008). These include reduced process motility, deramification, cytoplasmic fragmentation, and other physical abnormalities (Streit et al., 2004). When responding to injury, senescent microglia exhibit lower migration rates and have a sustained inflammatory response,

consistent with the low-grade inflammation typical of the ageing brain (Damani et al., 2011; Sparkman & Johnson, 2008). Aged microglia are also characterized by reduced phagocytosis and increased ROS production (Koellhoffer et al., 2017). Altogether, microglia fail to maintain neurons healthy in ageing and may also impair them in neurodegenerative disorders. The depletion of microglia-specific genes in older clusters is challenging to interpret. Several gene associated with normal microglia functions are downregulated in ageing, such as IDE (insulin-degrading enzyme) and SIRT1 (sirtuin 1), but many are upregulated, likely pro-inflammatory cytokines TNF-α and IL-β (interleukin 1- β) (Angelova & Brown, 2019). Nevertheless, a significant age-related change in the expression of core microglial genes is likely to represent a change function, which is potentially maladaptive.

There are some caveats of consider when interpreting the results of this cell-type analysis. First, while changes in fidelity score across developmental clusters are most parsimoniously explained by a change in the proportion of cells from that class, this cannot be concluded from the analysis alone. Second, Kelley et al. (2018) determined fidelity scores using single-cell data from adult cases (> 18 years of age), so cell-type markers may not be reflective of very young or very old individuals.

## 3.4 Summary

In this chapter, we introduced our developmental clustering methodology using transcriptomic data from human primary visual cortex (V1C). Choice of clustering algorithm and cluster number were validated and discussed. Robust and sparse k-means clustering (RSKC) revealed novel developmental dynamics in V1C, characterized by overlapping clusters that progressed across the lifespan. This model suggests that developmental periods may be defined by two or more distinct patterns of gene expression. Examination of the features driving the separation of samples into clusters identified genes with known roles in the development and function of the visual cortex, including markers of distinct inhibitory interneuron subpopulations. Other feature genes were involved in immune responses and disease. *In situ* hybridization analysis of feature gene *VAMP1* revealed localization to excitatory pyramidal neurons in layer IV, suggesting roles in neurotransmitter release or potential receptor trafficking in V1C. Differential expression (DE) analysis allowed for exploration of developmental clusters. When combined with an external database of cell-type markers, we were able to identify an up-

regulation of microglial-specific genes during perinatal development, and a subsequent down-regulation during ageing. In conclusion, developmental clustering pulled together several molecular mechanisms describing the development and ageing of the visual cortex and associated them with a more global pattern of expression which corresponded to known milestones in visual development.

# Chapter 4. Developmental Clusters in Multiple Brain Regions

## 4.1 Introduction

The brain is functionally organized into regions, which are distinguished by compositions of cells with distinct molecular profiles and laminal organizations, as well as region-dependent patterns of long-range and short-range connectivities (Keil et al., 2018). These anatomical, functional, and circuit-level differences are reflected in the transcriptome. In the original analysis of the current data, Kang et al. (2011) performed unsupervised hierarchical clustering across different periods of cortical development. They reported distinct and developmentally regulated clustering of neocortex (NCX; combination of 11 areas), AMY and HIP, with CBC exhibiting the most distinctive gene expression profile. Transcriptional differences between regions were particularly pronounced during early development, but became increasingly correlated with age. Notably, most spatially regulated genes were also temporally regulated (Kang et al., 2011). In another large-scale study, Carlyle et al. (2017) performed an in-depth transcriptomic and proteomic survey of regions of postnatal human brain, ranging in age from early infancy to adulthood. The authors jointly modeled protein changes over time and between brain regions using regression and ANOVA, identifying 1,840 proteins with differential expression between one or more regions. Unsupervised hierarchical clustering of these proteins revealed groups with substantial differences between regions and specific functional enrichments. For example, the largest CBC-enriched cluster contained several proteins involved in mRNA processing, possibly reflecting the higher density of nuclei in the cerebellar granular layers (Carlyle et al., 2017). Principal component analysis (PCA) of RNA expression in the same samples showed a clear separation of CBC, but not other regions. In sum, transcriptomic and proteomic data capture anatomical differences in brain tissue.

Nearly all stuides investigating the developmental dynamics of brain structure and anatomy are performed with brain imaging technologies. Several show *nonlinear* regional changes in gray matter (GM) density during childhood and adolescence (Giedd et al., 1999; Sowell et al., 2001; Jernigan & Tallal, 1990). GM density captured by magnetic resonance imaging (MRI) is a proxy for the complexity of neuronal and glial networks. Studies of GM maturation show a loss of density over time, correlating with periods of increased synaptic pruning during adolescence and early adulthood (Huttenlocher et al., 1982). Gogtay et al. (2004) used brain-mapping techniques to investigate cortical GM development in 13 healthy children

(4-21 years old) scanned with MRI every 2 years for 8-10 years. They found that development had significant regional heterogeneity and appears to follow a functional maturation sequence. Frontal lobe matured in a back-to-front direction, starting in the primary motor cortex, and propagating anteriorly over the superior and inferior frontal gyri, ending with the prefrontal cortex. In the posterior half of the brain, maturation began in the primary sensory areas and spread laterally over the parietal lobe. The frontal and occipital poles matured earliest, and the lateral poles matured last (Gogtay et al., 2004). The molecular and neurobiological events underlying this nonlinear, hierarchical pattern of brain development remain unknown.

The current study presents a unique approach to exploring this gap. Unsupervised clustering of developmental samples using transcriptomic data allowed for the identification of stable patterns of gene expression across the lifespan. Data from 16 brain regions were independently surveyed using this analysis pipeline, revealing spatially distinct patterns of developmental clusters, which may shed light on the heterogeneous maturational sequence of the human brain. Further, features driving the developmental separation of samples was compared between regions. Clustering of brain regions with similar patterns of gene expression was used to search for common developmental patterns. Interestingly, region clusters roughly charted anatomical separation, and clusters enriched for frontal areas exhibited sample separation by sex.

## 4.2 Methods

Data for the current study was originally collected by Kang et al. (2011). In brief, the authors measured genome-wide exon-level gene expression in 1,340 sample collected from 57 brains and 16 brains regions:

1. Neocortex
    - 1.1. Frontal cortex
        - 1.1.1. Orbital prefrontal cortex (OFC)
        - 1.1.2. Dorsolateral prefrontal cortex (DFC)
        - 1.1.3. Ventrolateral prefrontal cortex (VFC)
        - 1.1.4. Medial prefrontal cortex (MFC)
        - 1.1.5. Primary motor cortex (M1C)
    - 1.2. Parietal cortex

1.2.1. Primary somatosensory cortex (S1C)

1.2.2. Posterior inferior parietal cortex (IPC)

1.3. Temporal cortex

1.3.1. Primary auditory cortex (A1C)

1.3.2. Posterior superior temporal cortex (STC)

1.3.3. Inferior temporal cortex (ITC)

1.4. Occipital cortex

1.4.1. Primary visual cortex (V1C)

2. Hippocampus (HIP)
3. Amygdala (AMY)
4. Striatum (STR)
5. Mediodorsal nucleus of the thalamus (MD)
6. Cerebellar cortex (CBC)

All tissue specimens were collected from clinically unremarkable donors without history or signs of neurological or neuropsychiatric illness or drug use. The Affymetrix Human Exon 1.0 ST array, which features 1.4 million probe sets, was used to assay exon expression across the entire genome. Outlier samples were detected using hierarchical clustering and Spearman correlation, and removed from further processing steps. Affymetrix CEL files were then processed using a standard approach, involving RMA background correction, quantile normalization, mean probe set summarization, and log2-transformation. A total of 17,656 main protein-coding genes were surveyed.

The preprocessed data from Kang et al. (2011) were downloaded directly from the Gene Expression Omnibus (GEO) under the accession number GSE25219 using the *GEOquery* package (Davis & Meltzer, 2007). All data processing and visualization were performed in the R statistical software. The exon-summarized expression data were extracted, and probe identifiers were matched to genes. If a gene was matched by two or more probes then probe expressions were averaged if highly correlated (Pearson correlation, $r \geq 0.9$). Regions with less than 30 developmental samples were removed.

## 4.3 Results

### 4.3.1 Developmental clustering within brain regions

The analysis steps used in Chapter 2 to identify developmental clusters in V1C were applied to transcriptomic data from 16 distinct brain regions. The Kang et al. (2011) dataset includes samples from all four lobes of the neocortex (frontal, parietal, temporal, and occipital) and several sub-cortical structures, including the hippocampus (HIP), amygdala (AMY), striatum (STR), and mediodorsal nucleus of the thalamus (MD) (see section 2.2.1 for complete list of regions). Not all regions had the same number of developmental samples, and only those with at least 30 unique cases (two hemispheric samples per case) were selected for further analysis. Just as before, double-layered robust and spare k-means clustering (RSKC) was used to cluster pre- and post-natal samples, and subsequently identify subclusters. The results are visualized in Figure 15.

Many regions exhibited clusters of samples that progressed across the lifespan. A1C, ITC, M1C, MFC, OFC and STC followed a pattern of developmental progression similar to the one previously discussed in V1C, with clear separation of prenatal clusters and significant overlap of postnatal clusters. Specifically, cluster 2B spanned nearly all of postnatal life, while clusters 2A, 2C, and 2D were relatively restricted to early development and ageing, respectively. Perhaps this reflects a system where several stable patterns of gene-expression are available to the neural substrate at particular points during the lifespan. Other regions, such as DFC, IPC, and S1C, had much more distinct developmental clusters with less extensive overlap. This may be conceptualized as cortical tissue with more consistent gene expression in various developmental periods. The most striking finding was that subcortical regions AMY, CBC, HIP, and STR had very poor separation of postnatal samples following RSKC. This suggests that age is not a significant driver of clustering, or that gene expression following early development is mostly homogenous and no meaningful subclusters exist within the data.

Next, we asked whether differential patterns of lifespan gene expression across region reveal anything about the hierarchical maturation of the cortex as seen in imaging stuides (Gogtay et al., 2004; Giedd et al., 1999; Sowell et al., 2001; Jernigan & Tallal, 1990). To explore this question, the median age of samples in each cluster were plotted for all regions and grouped

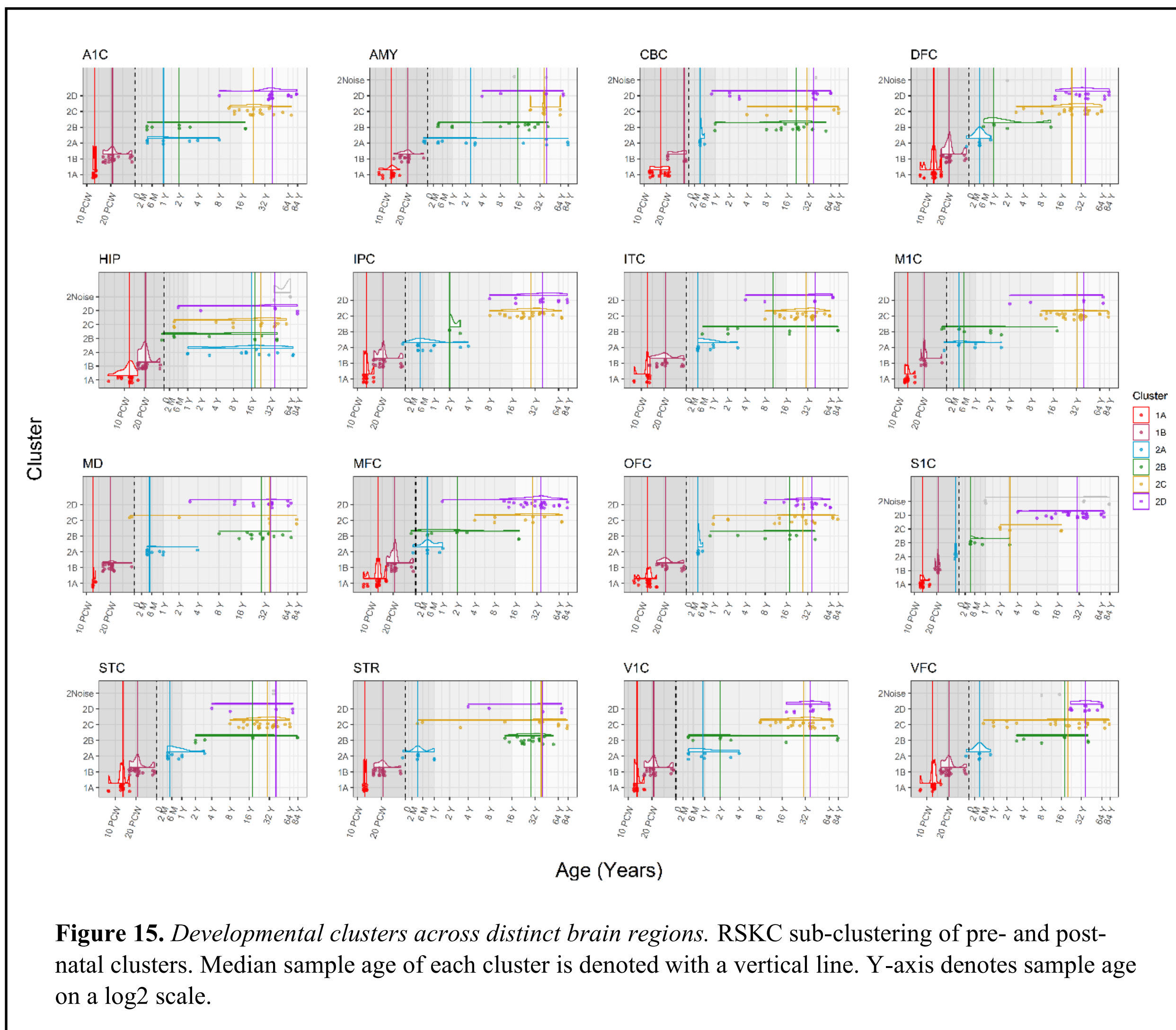


**Figure 15.** *Developmental clusters across distinct brain regions.* RSKC sub-clustering of pre- and post-natal clusters. Median sample age of each cluster is denoted with a vertical line. Y-axis denotes sample age on a log2 scale.

roughly into subcortical structures, primary sensorimotor areas, prefrontal cortex, and integration areas. This analysis must be interpreted carefully. We find that subcortical structures in general jump quickly to an older median sample age by the first or second postnatal cluster (Figure 16A). This may perhaps be interpreted as early maturation of the transcriptome. Gogtay et al. (2004) reported that the phylogenetically oldest cortical regions (e.g. entorhinal cortex and posterior piriform cortex) matured the earliest, suggesting that development follows evolutionary sequence to some degree. As the deep structures of the brain are considered to be evolutionarily more ancient, the current results appear to agree. Nevertheless, we must keep in mind that separation of prenatal samples in these regions did not appear to be driven by age. Primary sensorimotor areas have a more gradual increase in median sample age across clusters (Figure 16B). While

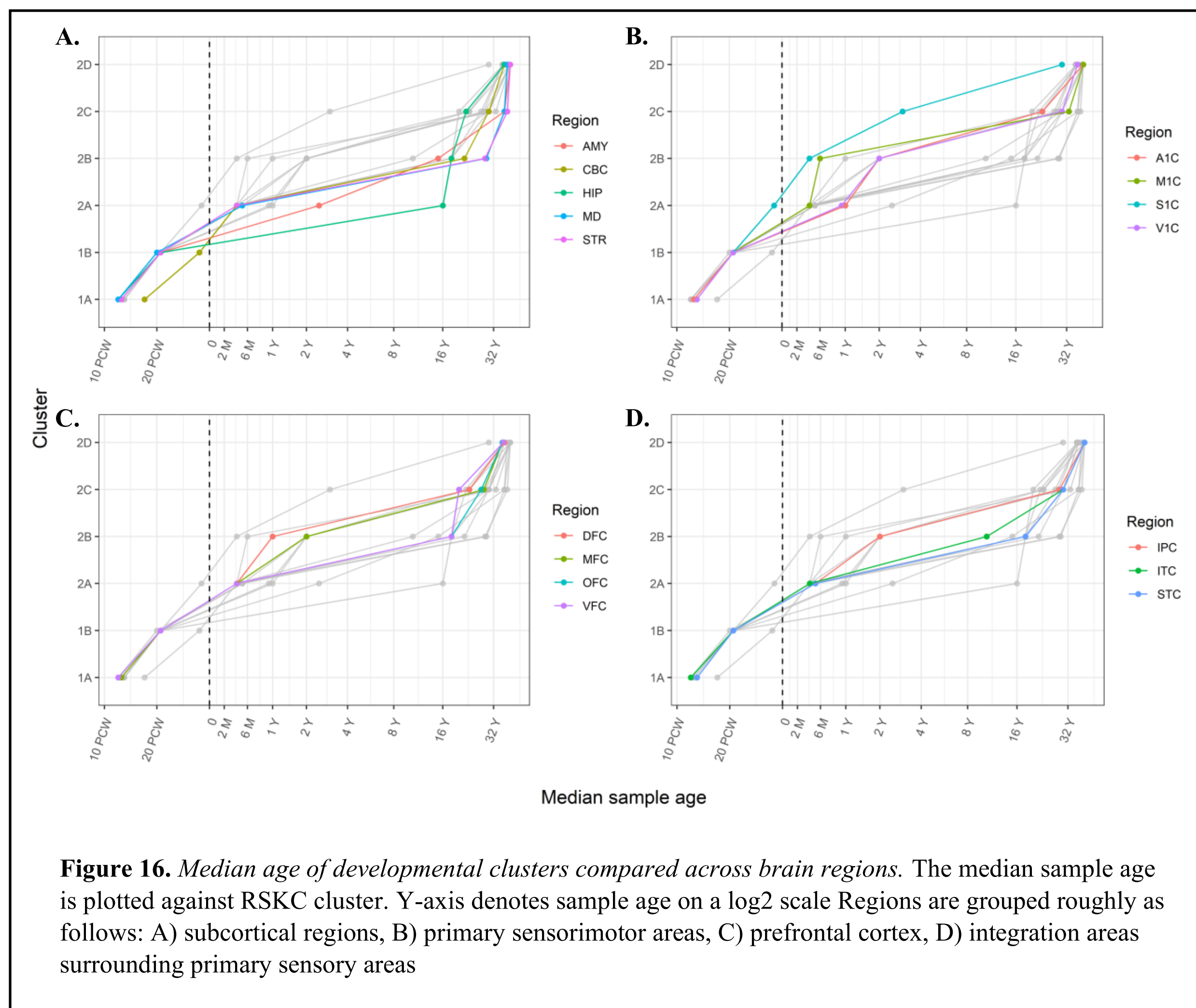


**Figure 16.** *Median age of developmental clusters compared across brain regions.* The median sample age is plotted against RSKC cluster. Y-axis denotes sample age on a log2 scale Regions are grouped roughly as follows: A) subcortical regions, B) primary sensorimotor areas, C) prefrontal cortex, D) integration areas surrounding primary sensory areas

this might be interpreted as slow maturation; it only reflects a greater number of distinct gene expression patterns that map onto early development. Ultimately, this analysis cannot be used to understand differential maturation rates across regions, but it does suggest that differences in when transcriptomic patterns emerge is related to anatomical factors.

### 4.3.2 Comparing developmental features across brain regions

Within each region, RSKC was used to identify two sets of feature genes: 1) those driving the separation of samples into pre- and postnatal clusters, and 2) those driving the separation of postnatal samples only (see section 2.3.1 for methodological justification). Weights were assigned to each of 17,243 genes, averaged over 100 iterations of clustering, and scaled between 1 and 0 for cross-region comparisons. Figure 17A visualizes the weights for the top thirty feature genes separating pre- and postnatal clusters. Several were observed to be shared

between regions. To summarize this information, the top 50 feature genes from each region were pooled then ranked by the number of other regions they appeared in. The weights for the top 30 most common features were visualized using a dot plot, and patterns of feature weights between regions were organized by hierarchical clustering. Figure 17B reveals considerable homogeneity in feature weights driving separation of pre- and postnatal clusters, however hierarchical clustering is still able to separate neocortical areas from subcortical structures, with MD and CBC exhibiting the most distinct patterns overall.

Two of the highest weighted genes across regions were *HIST1H3B* (H3 clustered histone 2) and *HIST1H3C* (H3 clustered histone 3). Both are sequence variants of histone H3, which is one of four core histones (H2A, H2B, H3, and H4) found in eukaryotic cells. Two core histones form an octamer, around which approximately 146 bp of DNA is wrapped in repeating units, call nucleosomes. The regulation of transcription through histone modifications and chromatin remodelling controls fundamental aspects of cell diversity and adaptation. Neurons in particular exhibit remarkable specialization and plasticity, which is mediated in part by activity-dependent changes in gene expression. Maze et al. (2015) demonstrate that *HIST1H3C* expression accumulates in neuronal and glial chromatin with age and remains highly dynamic throughout the lifespan. Manipulation of HIST1H3C dynamics in both embryonic and adult neurons confirmed its essential role in regulating cell-type specific gene expression programs and physiological plasticity (Maze et al., 2015). Another interesting common feature is *ELAVL4* (ELAV like RNA binding protein 4), better known as *HUD*. Precise regulation of mRNA processing and abundance are critical for proper spatiotemporal gene expression. These regulatory events in neurons are executed by several *trans*-acting factors, notably micro-RNAs (miRs) and RNA-binding proteins (RBPs), which bind to specific cis-acting elements or structures within mRNAs (Bonacci & Jasmin, 2013). The best characterized neuronal RBP is HUD, which has been demonstrated to govern the fate of many neuronal mRNAs through the modification of mRNA translation, processing, and stability. Accordingly, HUD is considered a "master-regulator" of various neuronal processes. Studies of HUD knockout (-/-) in mice revealed that HUD decreases self-renewal of neuronal progenitors and promotes cell-cycle exit, leading to the differentiation and commitment of cells. Adult knockdown increased neurogenesis in the subventricular zone (SVZ) (Akamatsu et al., 2005). HUD is also crucial for dendritic and axonal outgrowth, as demonstrated by the ectopic expression of HUD in forebrain neurons under

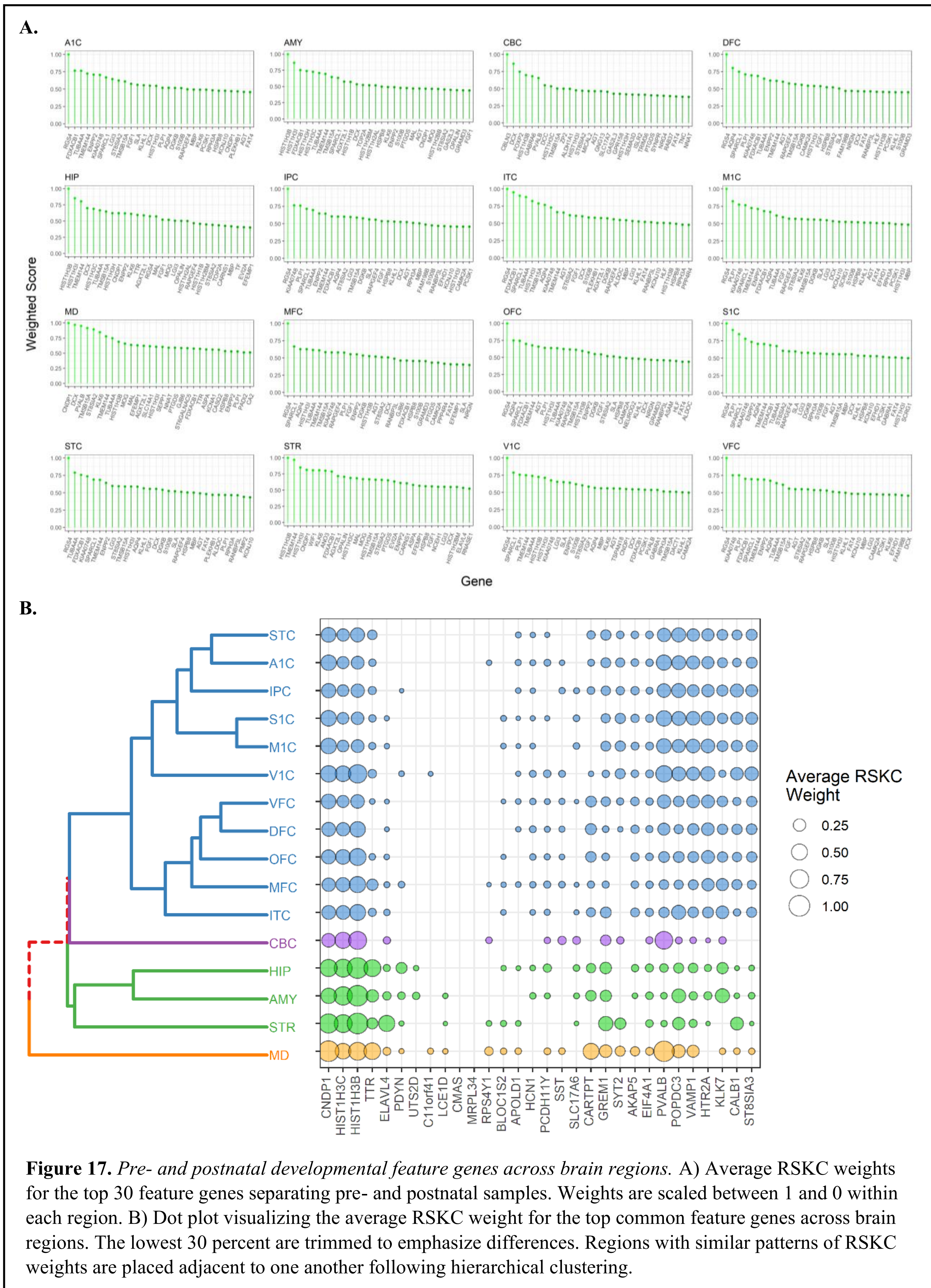


**Figure 17.** *Pre- and postnatal developmental feature genes across brain regions.* A) Average RSKC weights for the top 30 feature genes separating pre- and postnatal samples. Weights are scaled between 1 and 0 within each region. B) Dot plot visualizing the average RSKC weight for the top common feature genes across brain regions. The lowest 30 percent are trimmed to emphasize differences. Regions with similar patterns of RSKC weights are placed adjacent to one another following hierarchical clustering.

the control of the CAMKIIα promoter (Bolognani et al., 2007). Lastly, HUD appears to be involved in plasticity following nerve injury (Anderson et al., 2003) as well as hippocampal learning and memory (Bolognani et al., 2006).

Figure 18B visualizes feature weights for common genes driving the separation of postnatal clusters. There is considerably more heterogeneity in weights during postnatal development, and hierarchical clustering revealed a distinct organization of regions relative to the separation of pre- and postnatal samples. This adds an interesting perspective to the original observation by Kang et al. (2011) of increasing correlation between regions with age. While transcriptional differences between regions might decrease after birth, the influence (weight) of genes driving their development grows more varied. One interesting feature gene is *PDYN* (prodynorphin), which is proteolytically processed to form the secreted opioid peptides, including dynorphin. Expression of dynorphin increases with age and has been associated with memory impairments in rats (Kotz et al., 2004) and Alzheimer's disease in post-mortem tissue (Yakovleva et al., 2007). Ménard et al., (2013) reported that *PDYN* knockout (-/-) in mice results in significant changes in Group 1 metabotropic glutamate receptor (mGluR). Compared with age-matched wild-type (WT) littermates, mGluR1α and mGluR5 levels were elevated in the hippocampus and cortex of old, but not young, $PDYN^{(-/-)}$ mice. Increased mGluR expression in aged $PDYN^{(-/-)}$ mice was associated with enhanced LTP, implicating PDYN in the age-related decline of plasticity (Ménard et al., 2013). Additionally, individuals with a greater number of tandem repeat polymorphisms of the PDYN allele demonstrate decreased cognitive flexibility as measured by fMRI during reversal learning (Votinov et al., 2015). Another notable feature gene is *HCN1* (hyperpolarization activated cyclic nucleotide gated potassium channel), which present in both cardiac muscle cells (DiFrancesco, 1993) and in neurons (Pape, 1996). In neurons, HCN channels are found in dendrites and presynaptic axon terminals, where they regulate synaptic transmission. Specifically, HCN channels drive depolarizing currents in pacemaker neurons that generate rhythmic activity, which synchronizes the activity of nearby neurons (Santoro & Baram, 2003). The HCN1 isoform is expressed in mouse hippocampus during embryonic and postnatal development, and colocalizes with parvalbumin positive interneurons in the dentate gyrus (Seo et al., 2015). HCN ion channels are also associated with epilepsy (DiFrancesco & DiFrancesco, 2015), PTSD (Ni et al., 2020), and depression (Kim & Johnston, 2018). Lastly, AKAP5 (A-kinase anchoring protein 5) is a scaffolding protein found in dendritic spines that recruits PKA

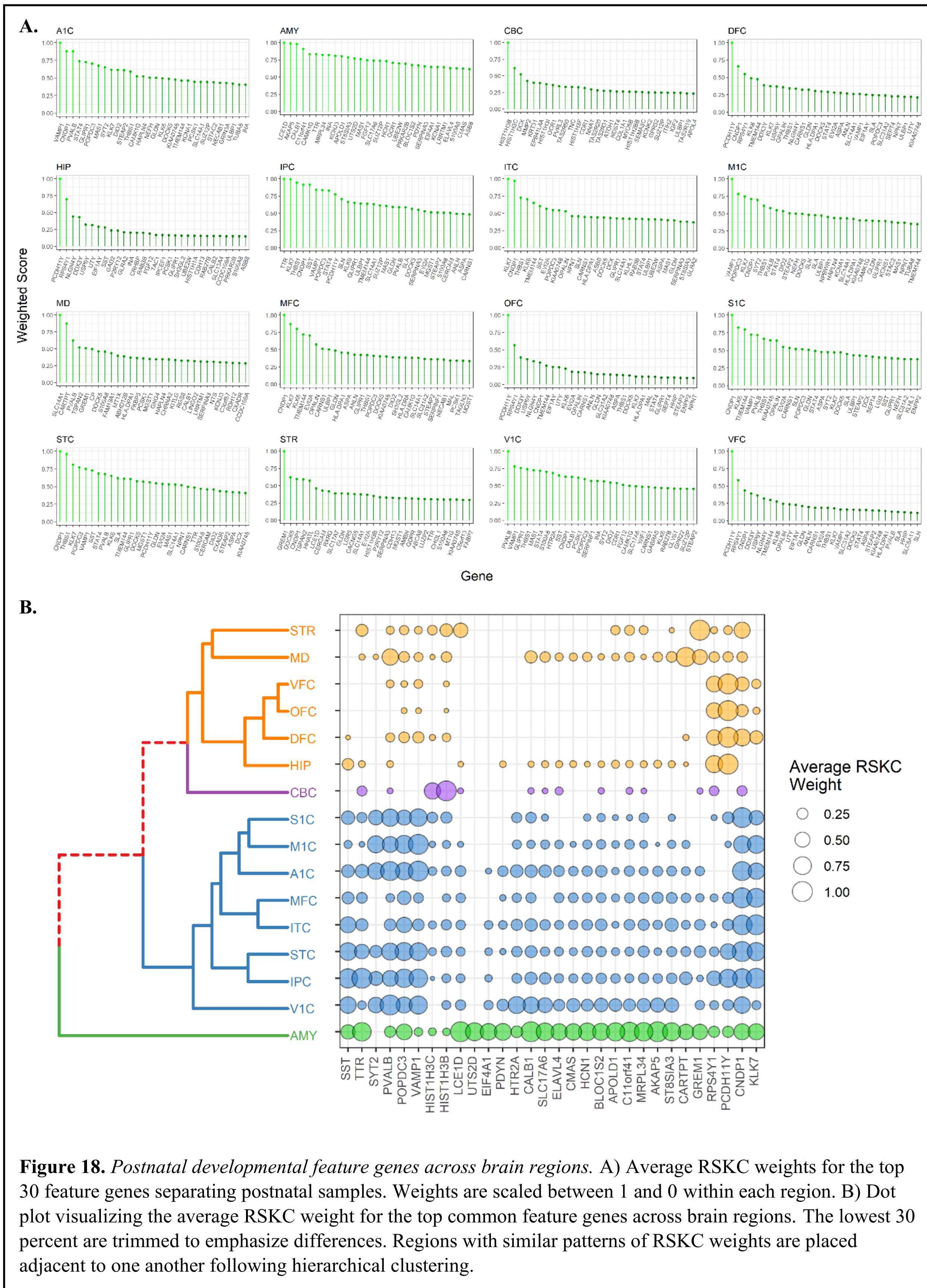


**Figure 18.** *Postnatal developmental feature genes across brain regions.* A) Average RSKC weights for the top 30 feature genes separating postnatal samples. Weights are scaled between 1 and 0 within each region. B) Dot plot visualizing the average RSKC weight for the top common feature genes across brain regions. The lowest 30 percent are trimmed to emphasize differences. Regions with similar patterns of RSKC weights are placed adjacent to one another following hierarchical clustering.

(cAMP-dependent protein kinase) and CaN (protein phosphatase 2B-calcineurin) to membrane-associated AMPA receptors to control receptor phosphorylation and synaptic plasticity (Robertson et al., 2009). Overexpression of AKAP5 in cultured hippocampal neurons increases the number of dendritic filopodia and spines, as well as AMPAR postsynaptic localization and activity (Robertson et al., 2009). A knockout of AKAP5 (-/-) results in severe deficits in synaptic plasticity and operate learning, as demonstrated by electrophysiological and behavioural analyses (Weisenhaus et al., 2010). Interestingly, a functional genetic polymorphism (Pro100Leu) in human AKAP5 contributes to individual differences in aggression and anger control (Richter et al., 2011).

In sum, feature selection allows for the identification of genes that contribute the most distinct transcriptomic patterns across the lifespan. The influence of feature genes can be compared across regions and combined to distinguish those that drive brain development in general. Some, such as *C11orf41*, code for proteins with no known function in the brain and may represent putative genes of interest for future studies of neurodevelopment.

### 4.3.3 Clustering of brain regions

Hierarchical clustering of feature weights revealed that some brain regions are more closely related than others. To explore this possibility further, RSKC was used to cluster regions by gene expression. The input for RSKC requires a matrix of observations (n) x variables (p). Each region (n) has several samples, each with an expression value for 17,243 genes. Accordingly, the data were organized such that each variable (p) corresponded to the expression of a specific gene for a specific sample (henceforth referred to as a gene-sample). There are an uneven number of samples across regions so any gene-samples missing data for 4 or more regions were removed, and the remaining gaps were imputed using nearest neighbour averaging. Reformatting of the data resulted in an input table with millions of columns, which was not computationally feasible. To circumvent this issue, the top 500 feature genes separating postnatal clusters most common across regions were used as a shortlist. This is viable because any sizable subset of the original data should still capture differences between regions.

As other studies have demonstrated, the most significant transcriptomic differences are found between neocortical and subcortical regions (Kang et al., 2011; Carlyle et al., 2017). Accordingly, RSKC was performed twice: 1) to separate between neocortical and subcortical

regions, and 2) to separate neocortical regions, which have more subtle differences in gene expression. Prior to clustering, the elbow method was used to estimate the optimal number of clusters, and $k = 4$ was selected (see section 2.2.4). As the results in Table 1 demonstrate, RSKC correctly separated all subcortical structures. The remaining neocortical regions were assigned into three clusters. Intriguingly, region clusters roughly charted anatomical separation. Cluster 1 consisted of the primary visual cortex and several areas of the temporal lobe (A1C, IPC, ITC, and STC). Cluster 2 consisted of areas on the rostral surface of frontal cortex (DFC, MFC, and OFC), while cluster 3 included areas along the lateral sulcus (M1C, S1C, and VFC). Lastly, cluster 4 contains all subcortical structures (HIP, AMY, STR, MD, and CBC).

Performing a clustering analysis on brain regions presents the unique opportunity to explore the features that drive their separation. Recall that the genes selected as inputs for clustering regions were also those assigned more weight during postnatal cluster separation across regions. In other words, any features selected by the current analysis drive *both* the separation of postnatal development as well as brain regions. While the same analysis is possible with the separation of neocortical and subcortical areas, the current section will focus on features identified during the clustering of neocortical areas. Weights for each gene were averaged over gene-samples and 100 iterations of RSKC. To get a broader perspective on this list of features, a Gene Ontology (GO) over-representation test was performed using *clusterProfiler* (Yu et al., 2012). Biological process terms enriched for the top 1000 feature genes were queried for common annotation categories, visualized in Figure 19. These included neurotransmitter transport (GO:0006836), regulation of ion transmembrane transport (GO:0034765), regulation of membrane potential (GO:0042391), regulation of neurotransmitter levels (GO:0001505), and synaptic vesicle cycle (GO:0099504). Remarkably, many of the genes underlying differential expression in postnatal development and between neocortical regions are associated with vital neuronal functions, especially neuronal communication. This finding agrees with the diversity in neuronal populations seen across cortical regions (Duath et al., 2017), and the functional maturation of neurons over development and ageing (Burke et al., 2020; Petralia et al., 2014).

| Region | RSKC 1 | RSKC 2 |
|---|---|---|
| A1C | 1 | 1 |
| AMY | 2 | |
| CBC | 2 | |
| DFC | 1 | 2 |
| HIP | 2 | |
| IPC | 1 | 1 |
| ITC | 1 | 1 |
| M1C | 1 | 3 |
| MD | 2 | |
| MFC | 1 | 2 |
| OFC | 1 | 2 |
| S1C | 1 | 3 |
| STC | 1 | 1 |
| STR | 2 | |
| V1C | 1 | 1 |
| VFC | 1 | 3 |

**Table 1.** *Results of region clustering.* RSKC was performed twice to group brain regions by gene expression. RSKC 1 successfully separated all neocortical and subcortical structures. RSKC 2 identified three clusters of neocortical regions. Cluster designation is indicated by a number.

### 4.3.4 Developmental clustering within region clusters

The identification of brain regions with similar patterns of gene expression allows us to update our developmental analysis. By pooling together samples from similar regions and increasing the amount of data points, RSKC can theoretically refine the boundaries of development clusters. The same analysis steps described previously were applied to the pooled sample data, and the results are presented in Figure 20. The development dynamics summarize those observed when regions were clustered independently (Figure 15). Cluster 1 (A1C, IPC, ITC, STC, V1C) demonstrate considerable overlap of postnatal clusters, whereas early and late postnatal development are more distinct for cluster 2 (DFC, MFC, OFC) and cluster 3 (M1C, S1C, VFC). Most striking result was the complete lack of developmental clustering in subcortical structures (cluster 4).

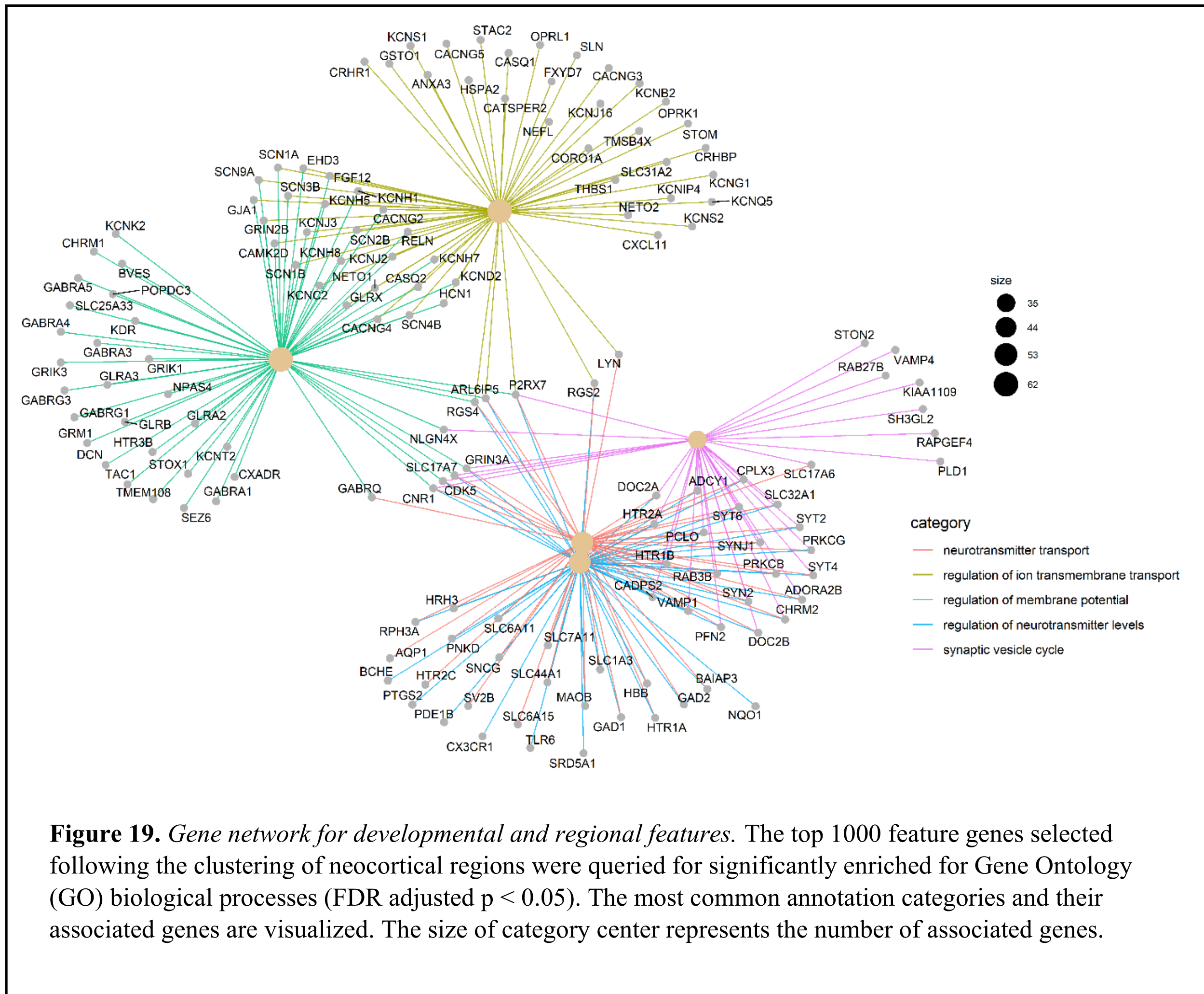


**Figure 19.** *Gene network for developmental and regional features.* The top 1000 feature genes selected following the clustering of neocortical regions were queried for significantly enriched for Gene Ontology (GO) biological processes (FDR adjusted $p < 0.05$). The most common annotation categories and their associated genes are visualized. The size of category center represents the number of associated genes.

To identify variables other than sample age that might drive cluster separation, the region, hemisphere, and sex of samples in each cluster were examined (Figure 21). All regions included in clusters 1, 2, and 3 had equal representation in developmental clusters. In cluster 4, there was a significant relationship between sample region and RSKC cluster (chi-squared test of independence, $p = 0.0072$). In other words, the differences in gene expression are better described by sample region than sample age. Cluster 2B was dominated by samples from CBC, 2C was dominated by samples from MD, and 2D by samples from STR. Examining the ratio of samples from each hemisphere revealed equal contribution from the right and left (Figure 21B).

The most interesting results comes from the analysis of sample sex. Postnatal samples in cluster 2 (DFC, MFC, OFC) and cluster 3 (M1C, S1C, VFC) appeared to separate into two sets of overlapping clusters spanning early and late development, respectively. The separation of

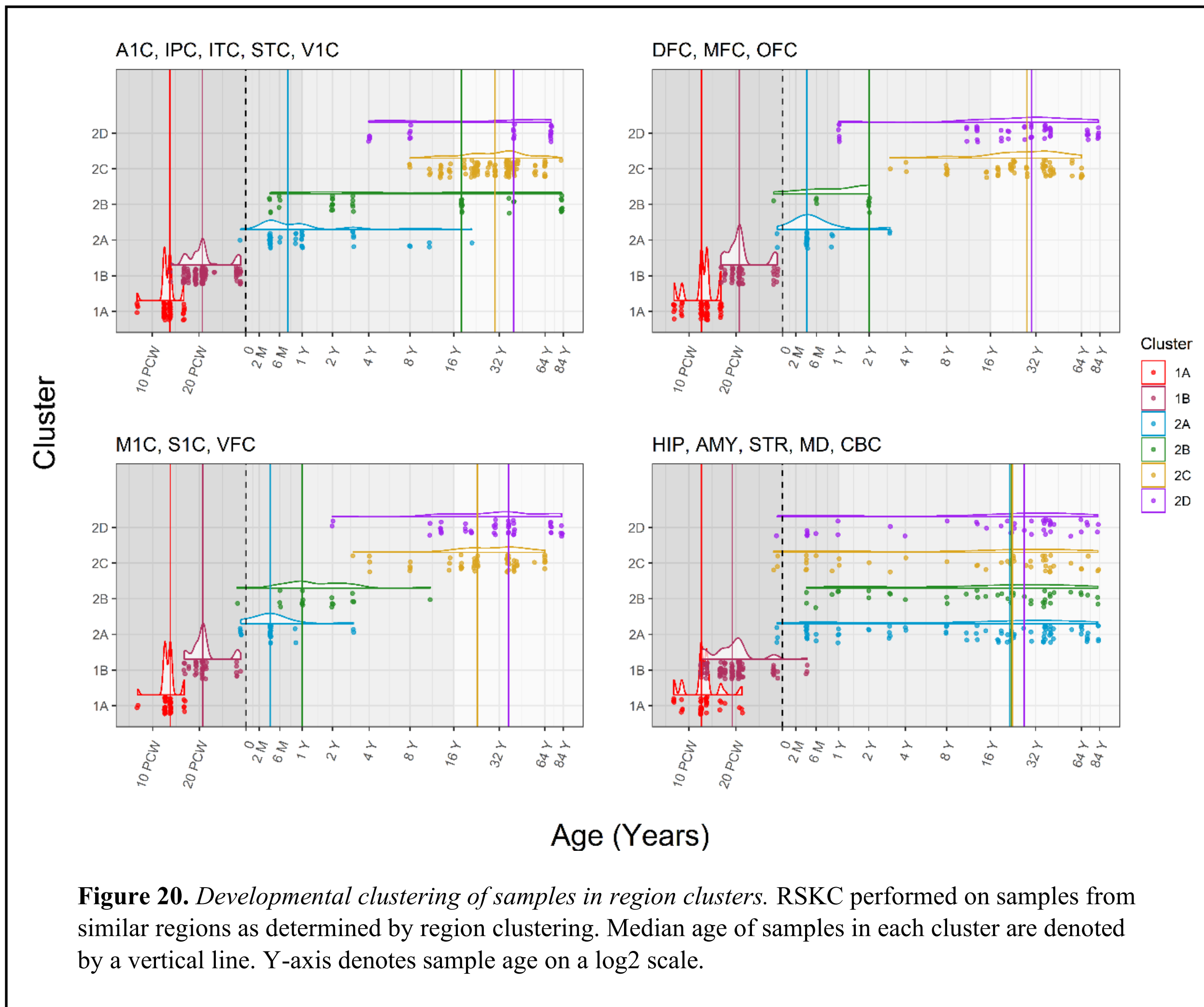


**Figure 20.** *Developmental clustering of samples in region clusters.* RSKC performed on samples from similar regions as determined by region clustering. Median age of samples in each cluster are denoted by a vertical line. Y-axis denotes sample age on a log2 scale.

overlapping clusters at similar periods of the lifespan is driven entirely by sample sex, as demonstrated in Figure 22. Specifically, clusters 2A and 2C consist of entirely males whereas cluster 2B and 2D consist entirely of females. This suggests that clustering is driven jointly, where the greatest variance in transcriptomic expression is explained first by large differences in age, and second by the sex of the individual. Moreover, this phenomenon appears to be specific to areas from the frontal and temporal lobes.

Differential expression (DE) analysis generates lists of genes that are significantly over- and under-expressed between groups. Here, DE analysis was used to make two-way comparisons between clusters separated by sample sex at similar periods during the lifespan (i.e. clusters 2A vs. 2B, and clusters 2C vs. 2D). To better characterize the results and place them in a biological context, genes lists were functionally annotated using the Gene Ontology (GO) (Ashburner et al.,

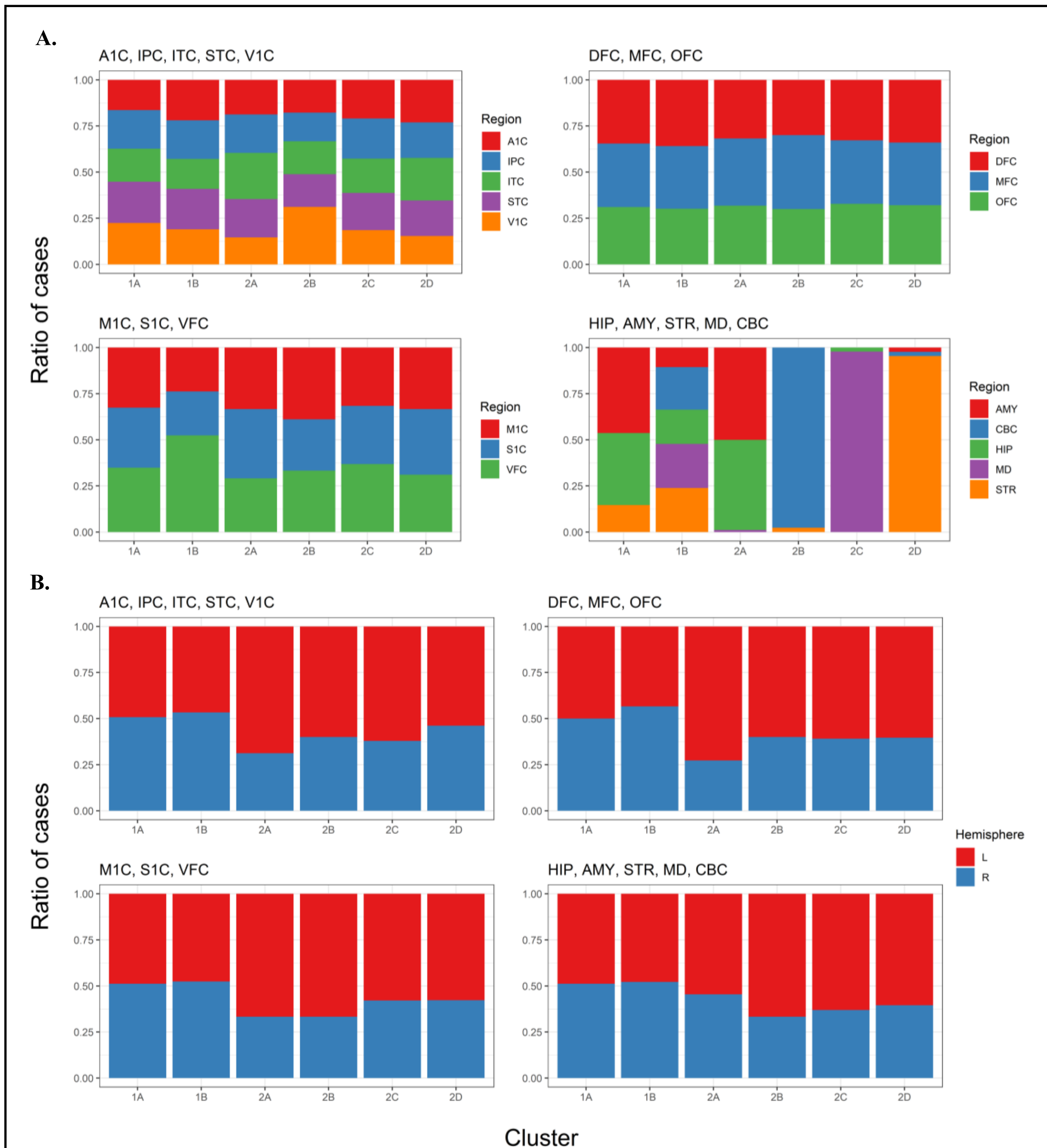


**Figure 21.** *Sample brain region and hemisphere per cluster.* The ratio of sample from different A) brain regions and B) hemispheres in each developmental cluster.

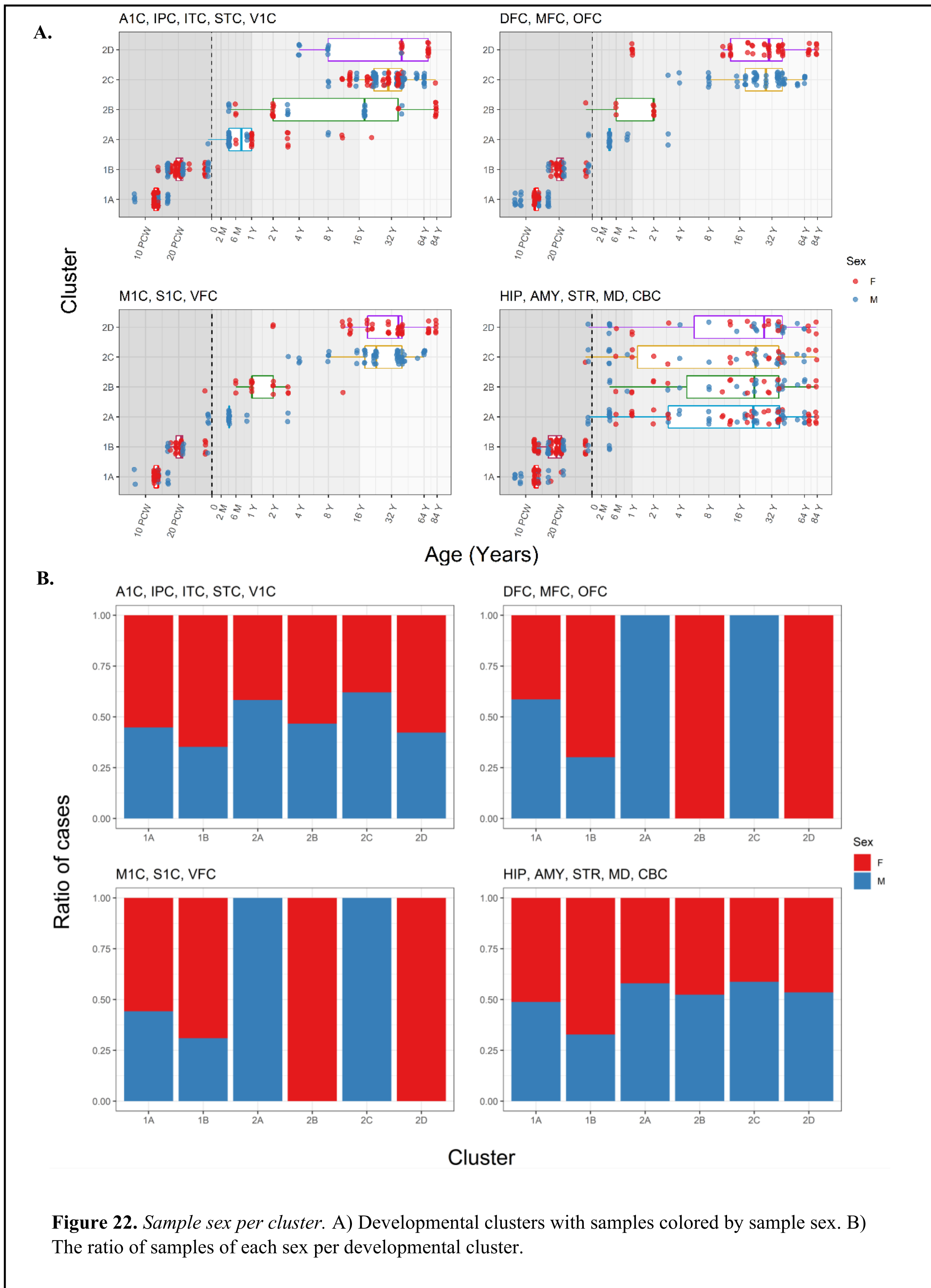


**Figure 22.** *Sample sex per cluster.* A) Developmental clusters with samples colored by sample sex. B) The ratio of samples of each sex per developmental cluster.

2000). Briefly, *limma* contrasts were adjusted such that only the clusters of interest were compared. Log-fold change between clusters was calculated for each gene and p-values were adjusted by false-discovery rate (FDR). Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) was performed on a list of all genes ranked by log-fold change, and significantly enriched biological terms were selected (FDR adjusted p-value $< 0.05$). Redundant terms were reduced using a similarity cut-off = 0.7. The top 20 terms with a positive normalized enrichment score (NES), representing enrichment in the male cluster, and top 20 terms with negative NES, representing enrichment in the female cluster, were plotted.

Figure 23A visualizes the enriched GO biological terms between clusters 2A (male, median age = 0.34 years) and 2B (female, median age = 2.0 years) for region cluster 2 (DFC, MFC, OFC). DE genes in cluster 2A are involved in axon development, neuron morphogenesis, synaptic transmission, and regulation of neurogenesis. The positive and negative regulation of neurodevelopment in cluster 2A relative to 2B is potentially due to younger samples in the cluster. Enrichment of processes related to the perception of smell in cluster 2B (female) is interesting given that women's olfactory abilities outperform men's (Sorokowski et al., 2019) and the sexual dimorphism of brain areas involved olfaction in humans (Oliveria-Pinto et al., 2014). These findings are also surprising given that areas in the temporal lobe are specialized for the perception of smell, rather than prefrontal cortex (PFC).

A greater number genes in clusters 2C (male, median age = 28.0 years) and 2D (female, median age = 30.0) were differentially expressed relative to the comparison between younger clusters. Several biological processes enriched in cluster 2C are involved in cell division, including DNA replication, sister chromatid segregation, g1/S transition of mitotic cell cycle, and nuclear division. This is surprising given that virtually all neuronal and glial cells in the frontal cortex are quiescent and permanently differentiated. Other processes are more pertinent to ageing, such as those involved in DNA repair. Genes differentially expressed in cluster 2D are largely involved in neuronal communication, represented by neuropeptide signalling pathway, action potential, ion transport, neurotransmitter transport, and regulation of neurotransmitter levels. This perhaps reflects greater activation of PFC areas in women to a range of stimuli (Stevens & Hamann, 2012; Goldstein et al., 2005) and increased neuronal activity in the

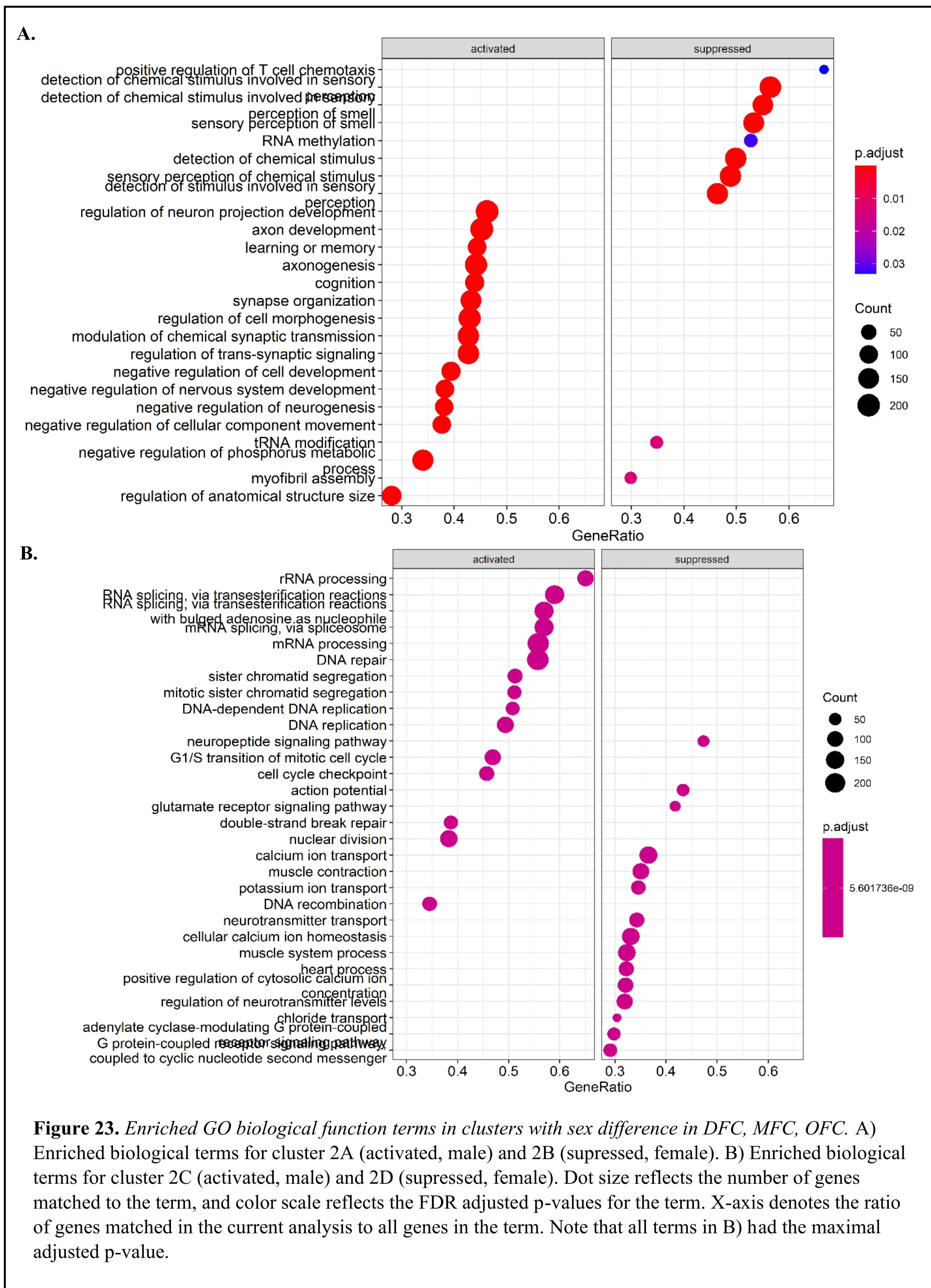


**Figure 23.** *Enriched GO biological function terms in clusters with sex difference in DFC, MFC, OFC.* A) Enriched biological terms for cluster 2A (activated, male) and 2B (supressed, female). B) Enriched biological terms for cluster 2C (activated, male) and 2D (supressed, female). Dot size reflects the number of genes matched to the term, and color scale reflects the FDR adjusted p-values for the term. X-axis denotes the ratio of genes matched in the current analysis to all genes in the term. Note that all terms in B) had the maximal adjusted p-value.

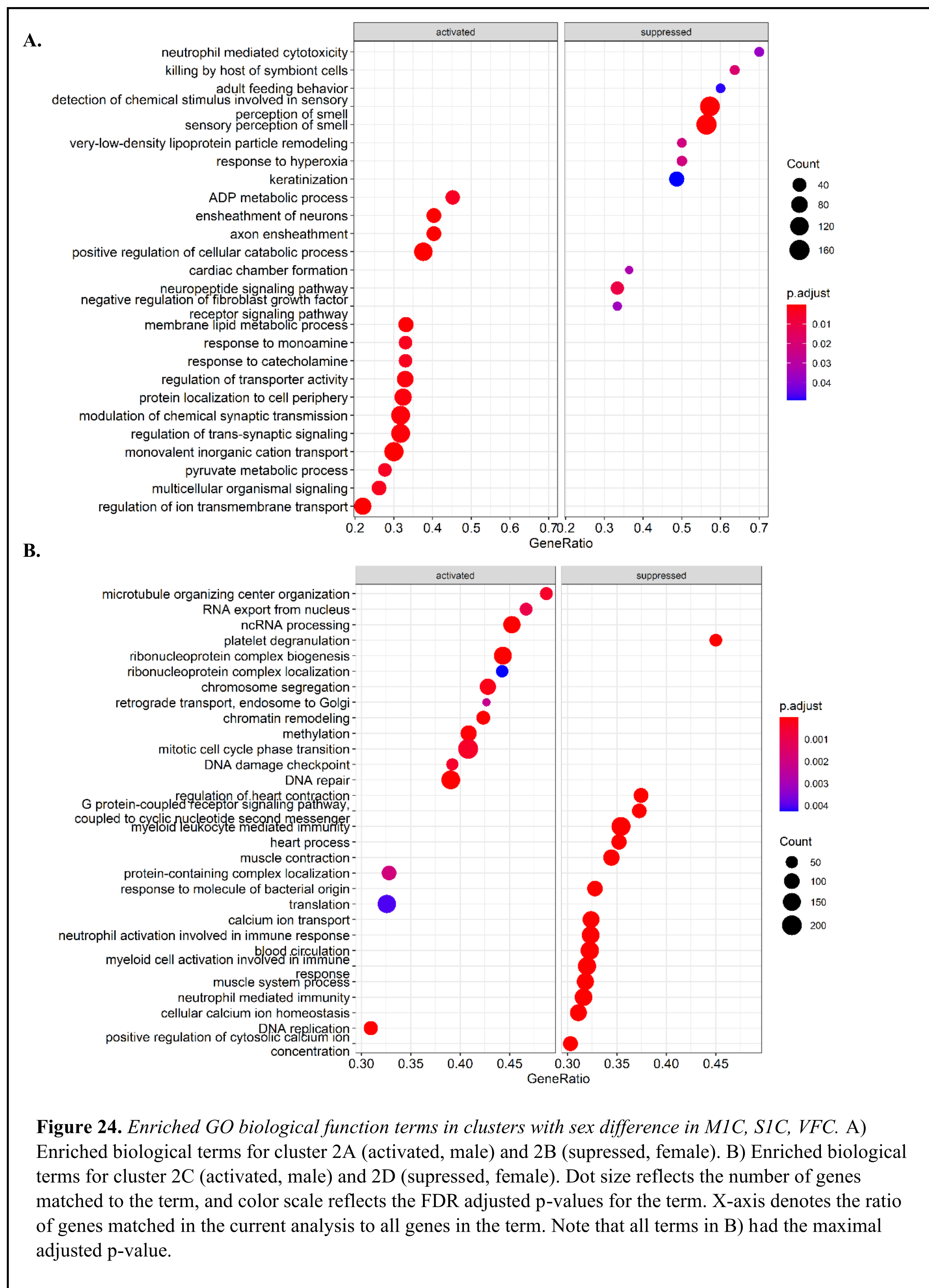


**Figure 24.** *Enriched GO biological function terms in clusters with sex difference in M1C, S1C, VFC.* A) Enriched biological terms for cluster 2A (activated, male) and 2B (supressed, female). B) Enriched biological terms for cluster 2C (activated, male) and 2D (supressed, female). Dot size reflects the number of genes matched to the term, and color scale reflects the FDR adjusted p-values for the term. X-axis denotes the ratio of genes matched in the current analysis to all genes in the term. Note that all terms in B) had the maximal adjusted p-value.

presence of estrogen (Zhou et al., 2005). Nevertheless, these biological processes are not overtly related to known sex-differences in the brain.

Developmental clusters 2A (male, median age = 0.33 years) and 2B (female, median age = 1.0 years) were also compared for region cluster 3 (M1C, S1C, and VFC). The male-dominated cluster was enriched for biological processes involved in axon ensheathment, metabolic processes, and chemical synaptic transmission. A log-fold increase in genes related to myelination may reflect the bias of younger samples, or potentially the greater overall white matter volume in males (Cerghet et al., 2009; Ritchie et al., 2018). It is interesting that sensory perception of smell was enriched again for the female-dominated cluster (2B). Similarly, comparing the two older clusters reproduced the puzzling up-regulation of cell cycle processes in the male-dominated cluster (2C). In both female clusters, there was an increase in immune-related processes. It should be noted that GO annotations are biased towards processes in the peripheral system, and terms describing peripheral immune activation may still reflect immune functions in the brain. We have previously discussed the critical role of microglial, the brain's local immune cell, in neurodevelopment (see section 2.3.5). In addition, crosstalk between the peripheral immune system and the brain is an important mechanism by which sex differences in the brain are established (Arambula & McCarthy, 2020). Rilett et al. (2015) demonstrated that mice lacking T-cell receptor β and δ chains exhibit sex differences in behaviour as well as sexual dimorphisms in several brain regions.

## 4.4 Summary

In this chapter, transcriptomic data from 16 brain regions were independently surveyed using our analysis pipeline, revealing spatially distinct patterns of developmental clusters. Notably gene expression in subcortical structures does not separate samples into postnatal developmental clusters. While the relationship between the transcriptome and the maturational sequence of the brain seen in MRI studies remains to be clarified, the varying patterns of lifespan gene expression seen between regions suggests a link. The visualization of shared features across regions identified several that are crucial for normal brain development and function. Moreover, the hierarchical clustering of regions based on feature weights demonstrated that while transcriptional differences between regions might decrease after birth (Kang et al., 2011), the influence of genes driving their development grows more varied. RSKC was also used to cluster

brain regions with similar patterns of transcriptomic expression. Interestingly, region clusters roughly charted anatomical separation, and provided a more nuanced separation of neocortical regions compared to the original analysis of this data (Kang et al., 2011). The feature genes driving both postnatal development and separation of neocortical areas were highly enriched for biological processes related to neuronal communication, agreeing with the diversity in neuronal populations seen across cortical regions, and the functional maturation of neurons over development and ageing. Samples within each regional cluster were pooled in order to refine our developmental analysis. The results surprising demonstrated that developmental clustering in DFC, MFC, OFC and M1C, S1C, VFC is driven jointly by sample age and sample sex. GO analysis of sex-specific clusters revealed the enrichment of interesting biological processes, although none were overtly related to known sex-differences in the human brain.

# Chapter 5. Clustering Bipolar Cases with Developmental Controls

## 5.1 Introduction

We have discussed the utility of stages for describing the discontinuous nature of development, particularly the spontaneous re-organization of function, structure, and behaviour. Staging is also an indispensable heuristic for clinicians. Bipolar disorder (BD) is a severe, chronic mood disorder characterized by recurrent episodes of mania, hypomania, and depression, punctuated by euthymic intervals (Belmaker, 2004). Since the 1920s, physicians have noted that a significant proportion of BD patients experience increasing frequency of episodes coupled with decreasing levels of cognitive and psychosocial functioning (Kraepelin, 1921). There is also evidence that patients at early stages of the illness have far better clinical outcomes and responses to treatments such as lithium, compared to those who have experienced multiple episodes (Schuepbach et al., 2008). Consequently, various staging models have been proposed that are intended to indicate where an individual sits on a succession from "at risk" but asymptomatic to an "end-stage" with poor prognosis (Kapczinski et al., 2004). Some define stages by the occurrence, remission, and recurrence of mood episodes (Berk et al., 2007), while others emphasize the assessment of cognitive and psychosocial functioning in between episodes (Kapczinski et al., 2009; Fries et al., 2012). All staging models, however, assume an underlying pathophysiological process that is associated with anatomical changes, loss of cellular fidelity, and cognitive decline. This pathological rewiring of the brain is referred to as *neuroprogression*. Neuroprogression is thought to be both multifactorial and interactive, putatively involving the dopaminergic system, inflammation, oxidative stress, mitochondrial stress, and changes in neurotrophins (Grande et a l., 2019). Several theoretical frameworks have attempted to explain the biological changes observed in severe psychiatric disorders. One nascent model is accelerated ageing (AA).

Ageing refers to the decline of an organism's fitness owning to internal physiological degeneration, and in humans, involves physical, neuropsychological, and social changes (Gladyshey et al., 2016). Early evidence for AA in severe mental disorders was the high prevalence and earlier onset of age-related medical conditions, including metabolic imbalances, cardiovascular disease, autoimmunity disorders, and cancer (Goldstein et al., 2009; Crump et al., 2013). There is also a well-recognized link between severe mood disorders and dementia (da Silva et al., 2013). Several biological changes seen in normal ageing are mirrored in neuro-

psychiatric disorders. These include a loss of myelinated fibers and white matter lesions (Raz & Rodrigue, 2006; Allen et al., 2005; Sassi et al., 2003), shortening of telomeres (Blackburn, 1991; Colpo et al., 2015), and increased oxidative stress (Şimşek et al., 2016; Brown et al., 2014). Another major topic in the literature is immunosenescence, which refers to changes in the function of the immune system that occur in with ageing. While older individuals are able to mount successful innate immune responses, their vulnerability to sustained low-grade inflammation may cause adaptive, anti-inflammatory networks to become overwhelmed and eventually fail, resulting in exposure to damaging agents (Franceschi et al., 2007). Individuals with BD demonstrate an inability to reduce inflammation after acute stimuli compared to healthy controls, as indicated by the lower percentage of regulatory T cells compared to healthy controls (Wieck et al., 2013). Several stuides also report significant changes in inflammatory markers in the bipolar brain (Kauer-Sant'Anna et al., 2009; Modabbernia et al., 2013).

While evidence for theories of neuroprogression like AA is growing, it remains challenging to collect data across various physiological domains in a special population. Many stuides have turned to transcriptomic resources to address this limitation (Pfaffenseller et al., 2016). In the standard praxis, gene expression is compared between age-matched groups of controls and BD cases. As the current work suggests, however, healthy brain samples of similar age may be represented by more than one distinct transcriptomic pattern. One solution is cluster BD samples together with developmental controls. This approach allows us to indirectly test the hypothesis of AA. If robust and sparse k-means clustering (RSKC) groups samples together based on similar expression across the transcriptome, we would expect BD samples to cluster with controls of a significantly higher physiological age, if at all. While this hypothesis was not supported by the results, we present a novel way to compare expression data from disease groups to data from controls.

## 5.2 Methods

### 5.2.1 Data acquisition

Two different datasets were used in the current chapter. The data for controls was originally collected and analyzed by Colantuoni et al. (2011). Using two-color arrays, gene expression was quantified in 269 human brain samples from prefrontal cortex (PFC) spanning

gestational week 14 through 80 years of age. This dataset was selected because it has substantially more samples from PFC than the Kang et al. (2011) study, and there are proportionally more adult samples. Additionally, using a different developmental dataset will allow us to validate our developmental clustering approach. The preprocessed data were downloaded directly from the Gene Expression Omnibus (GEO) under the accession number GSE5392 using the *GEOquery* package (Davis & Meltzer, 2007).

The data for BD samples were originally published by Ryan et al. (2006). The authors used microarray technology to quantify the expression of 22,000 mRNA transcripts in 30 bipolar and 31 control subjects for the dorsolateral prefrontal cortex and 10 bipolar and 11 control subjects for the orbitofrontal cortex. The raw files were downloaded from GEO under the accession number GSE5392 and processed in the statistical language R. Briefly, .CEL files were read using the *affix* package (Gautier et al., 2004), normalized via RMA (Robust Multichip Average), and probes were aligned for the Illumina Human 49K Oligo array (HEEBO-7 set).

### 5.2.2 *In silico* data merging

Both stuides were combined before downstream analyses using the workflow described by Taminau et al. (2012) for merging microarray gene expression datasets. First, the data was subsetted to included only genes common to both arrays (10081 in total). Quantile normalization was then applied to normalize expression between studies while maintaining meaningful differences. The density distributions for each study are visualized before and after normalization in Figure 25A-B. Subsequently, the *ComBat* function from the *sva* package (Leek et al., 2014) was used to remove batch effects across studies. Briefly, *ComBat* is an Empirical Bayes (EB) method where the mean and variance (L/S) parameters that represent batch effects are estimated by “pooling information” across genes in each batch to “shrink” the batch effect parameter estimates toward the overall mean of the batch effect estimates (across genes). These EB estimates are then used to adjust the data (Johnson et al., 2007). *ComBat* was supplied with the batch number for each sample from the metadata in each study. The effects of *ComBat* on the global structure of the data were visualized with tSNE (t-distributed stochastic neighbour embedding) (Figure 25C). Samples from both studies became integrated into the same feature space while retaining intra-study differences.

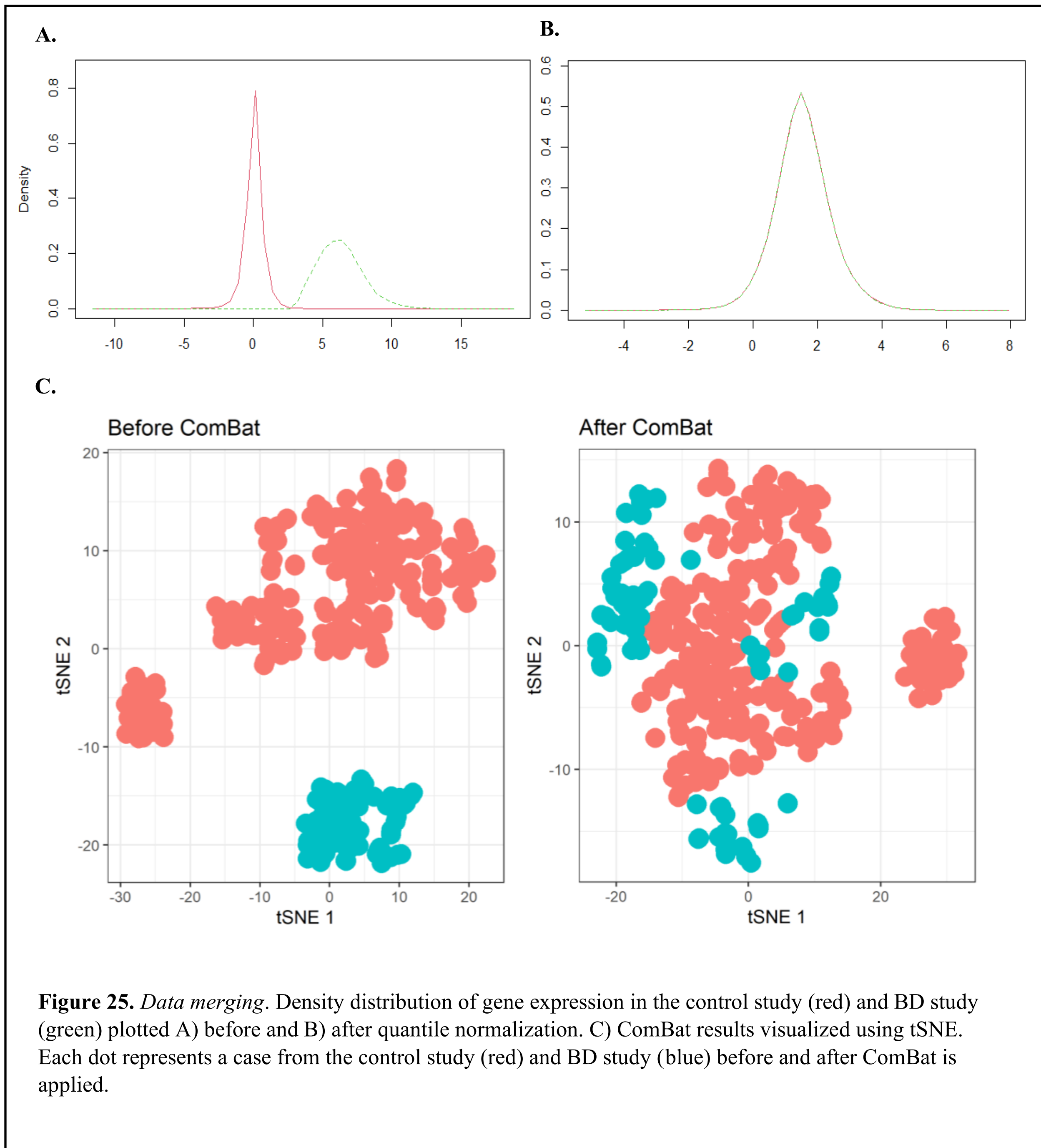


**Figure 25.** *Data merging*. Density distribution of gene expression in the control study (red) and BD study (green) plotted A) before and B) after quantile normalization. C) ComBat results visualized using tSNE. Each dot represents a case from the control study (red) and BD study (blue) before and after ComBat is applied.

## 5.3 Results

### 5.3.1 Clustering control and BD samples

RSKC clustering of samples pooled from Colantuoni et al. (2011) and Ryan et al. (2006) produced some unexcepted results (Figure 26). First, there was significant overlap of postnatal clusters with respect to sample age. This is in stark contrast to the minimally intersecting developmental clusters that formed between samples from prefrontal areas collected by Kang et al. (2011) (Chapter 3). One possible explanation is a 10-fold increase in the number of cases included the current analysis (330 compared to 37) and the greater proportion of samples from adult brains. While no distinct clusters formed during early development (0 – 4 years of age) there was still a moderate progression in cluster age across the lifespan (median sample age: 17.3, 30.5, 39.4, 46.2 years, respectively).

Even more surprising was the clustering of BD samples to different control-dominated clusters. One would expect the pathophysiology of the disorder to be broadly reflected in the transcriptome, resulting in a unique pattern of gene expression that would cluster BD samples together. However, the current results suggest that there is heterogeneity in the transcriptional profile of BD tissue, and that these differences are explained in part by heterogeneity seen in healthy individuals during normal development and ageing. Perhaps the sub-clustering of BD cases is not surprising considering the interindividual differences observed across several clinical characteristics, including age of onset (Bellivier et al., 2003), cognitive ability (Aminoff et al., 2013; Burdick et al., 2014), risk of suicidality (Cavazzoni et al., 2007), and frequency of mood episodes (Papadimitriou et al., 2005), among others. Moreover, while the etiology of BD is considered highly polygenic, GWAS (genome-wide association studies) have only identified a handful of genes with significant associations (Charney et al., 2017). Several meta-analyses also show discordant levels of biomarkers measured in serum and cerebrospinal fluid (CSF) of BD individuals (Barbosa et al., 2014; Wang & Miller, 2018). These findings suggest interindividual differences in the expression of genes implicated in the pathophysiology of BD. Ultimately, if only a subset of genes contribute to BD, and these genes vary between BD individuals, developmental genes likely have a larger effect on clustering.

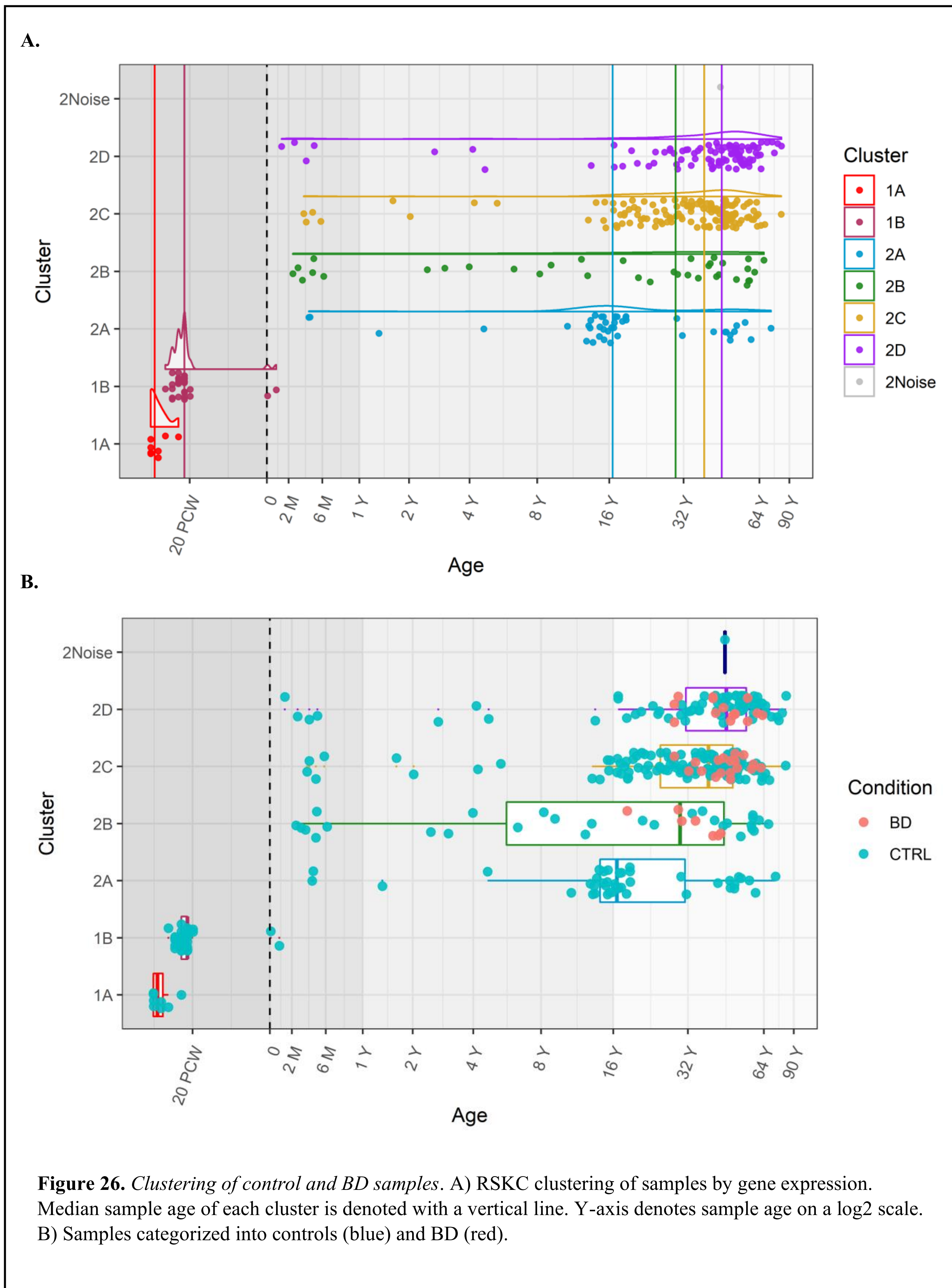


**Figure 26.** *Clustering of control and BD samples*. A) RSKC clustering of samples by gene expression. Median sample age of each cluster is denoted with a vertical line. Y-axis denotes sample age on a log2 scale. B) Samples categorized into controls (blue) and BD (red).

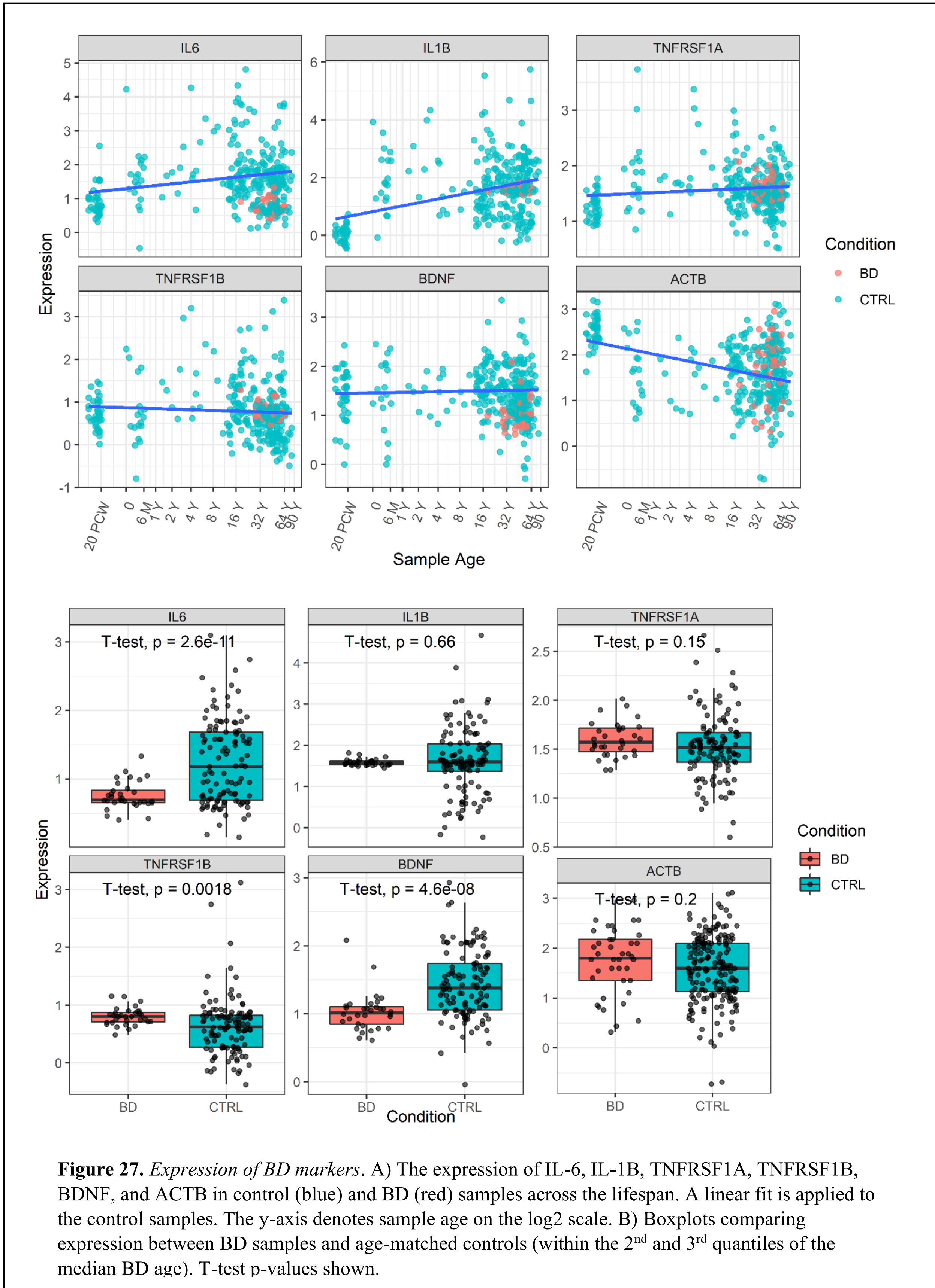


**Figure 27.** *Expression of BD markers*. A) The expression of IL-6, IL-1B, TNFRSF1A, TNFRSF1B, BDNF, and ACTB in control (blue) and BD (red) samples across the lifespan. A linear fit is applied to the control samples. The y-axis denotes sample age on the log2 scale. B) Boxplots comparing expression between BD samples and age-matched controls (within the 2nd and 3rd quantiles of the median BD age). T-test p-values shown.

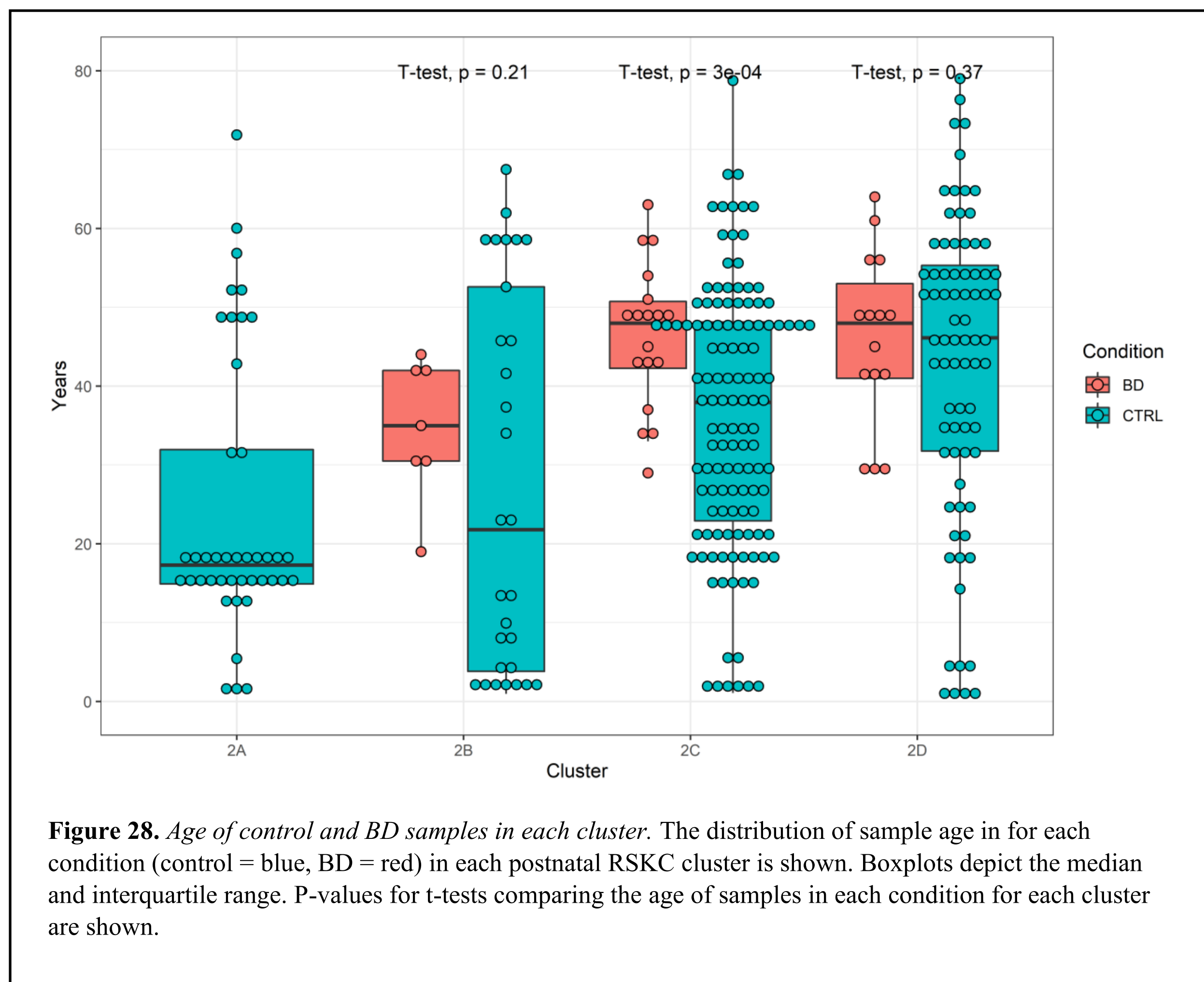


**Figure 28.** *Age of control and BD samples in each cluster.* The distribution of sample age in for each condition (control = blue, BD = red) in each postnatal RSKC cluster is shown. Boxplots depict the median and interquartile range. P-values for t-tests comparing the age of samples in each condition for each cluster are shown.

To make sure that differences in gene expression between control and BD samples were not improperly minimized by data merging, the expression of five markers genes were plotted (Figure 27). A number of studies have reported increased serum levels of inflammatory cytokine interleukin (IL)-6 in BD (Kauer-Sant'Anna et al., 2009; Barbosa et al., 2014). However, the current study finds a significant downregulation of IL-6 relative to age-matched controls ($p = 2.6 \times 10^{-11}$). This is actually in line with IL-6 measurement from CSF in bipolar individuals (Söderlund et al., 2011), which is a closer proxy for the brain's environment. Differences in IL-1B and TNFRSF1A expression were not observed, however TNFRSF1B expression was significantly greater in BD samples relative to controls ($p = 0.0018$). The peripheral levels of this cytokine in BD are similar to controls in some stuides (Kapczinski et al., 2011) and elevated in others (do Prado et al., 2013). BDNF (brain-derived trophic factor) was significantly depleted in BD samples ($p = 4.6 \times 10^{-8}$), which is consistent with multiple findings (Kauer-Sant'Anna et al.,

2009; Grande et al., 2010). Lastly, the expression of ACTB (actin beta), a common house-keeping gene, was compared between controls and BD samples. No differences were found, as expected. It should be noted that some well-characterized markers (e.g. tumour necrosis factor alpha; TNF-α) were not included in both datasets, and could not be represented here.

To take a closer look at the relationship between control and bipolar samples, the ages of each sample were plotted for each condition across postnatal clusters (Figure 28). A t-test was performed in each cluster to test whether there was a significant difference in age between conditions. Within the model of accelerated ageing (AA), we would expect concordance between the expression profile of BD samples and ageing controls, resulting in designation to the same cluster. Accordingly, the average age of control samples relative to BD samples should be higher in the same cluster. Yet, the results do not support this prediction. The only cluster with a significant difference in mean sample age between conditions was 2C (t-test, $p = 0.0003$), and in this case BD samples were older than controls. Notably, the younger BD samples matched to a relatively younger control cluster.

### 5.3.2 Comparing BD transcriptional subgroups

If an older transcriptional profile is not driving the separation of BD samples into different clusters, what is? To answer this question, the subject and clinical characteristic of cases in each cluster were examined (Table 2). There was no over-representation of one sex over the other in any cluster. Age of illness onset was 20.9 % younger in cluster 2B relative to the other clusters. The shorter duration of illness in this cluster likely reflects the younger median age of samples. Suicide was 37.4 % more common in cluster 2C relative to other clusters. No individuals in cluster 2B were prescribed lithium, but Valproate usage was consistent across clusters. Electroconvulsive therapy (ETC) had been administered to 4 individuals from cluster 2C. Despite evidence that antipsychotics and other interventions can reduce ageing-related changes seen in severe mood disorders, such as oxidative stress (Machado-Vieira et al., 2007), no obvious trends across multiple clinical factors are apparent between clusters.

Another approach to unpacking the differences between bipolar subgroups is to perform a differential gene expression (DE) analysis. Two-way comparisons between conditions within a cluster did not reveal many significant DE genes (< 50), as expected. Accordingly, both control and BD samples in each cluster were included in the analysis. Clusters 2B, 2C, and 2D were

| Cluster | N | Female (percent) | Median Age | Age of Onset | Duration | Suicide (percent) | Lithium (percent) | Valproate (percent) | ECT (percent) |
|---|---|---|---|---|---|---|---|---|---|
| 2B | 7 | 57.1 | 35 | 19.3 | 15.4 | 42.9 | 0.0 | 28.6 | 0.0 |
| 2C | 18 | 44.6 | 48 | 24.3 | 21.7 | 55.6 | 33.3 | 33.3 | 11.1 |
| 2D | 15 | 46.6 | 48 | 23.3 | 22.7 | 33.3 | 33.3 | 33.3 | 0.0 |

**Table 2.** *Subject and clinical variables for BD cases.* A summary of several variables describing the subject and clinical characteristic of BD samples in each postnatal cluster.

individually compared against pooled samples from the two remaining clusters to identify genes that were significantly over-expressed (log-fold increase in average expression) or under-expressed (log-fold decrease in average expression) in each. Cluster 2B had by far the most significant DE genes (4928, FDR adjusted p-value < 0.05). Consequently, Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) was performed to identify Gene Ontology (GO) biological processes that were activated and supressed in this cluster. The results are visualized in Figure 28. Cluster 2B was enriched for in myeloid leukocyte activation (GO:0002274), detoxification of inorganic compounds (GO:0061687), and stress response to metal ion (GO:0097501). As mentioned previously, GO annotations are biased towards processes in the peripheral system, and terms describing peripheral immune activation may still reflect immune functions in the brain. For example, leukocyte activation is characterized generally by the synthesis and release of inflammatory molecules. Several studies have linked the pathophysiology of BD with inflammation and changes in immune function. BD is frequently comorbid with autoimmune diseases such as multiple sclerosis, thyroiditis, and diabetes (Rosenblat &McIntyre, 2015). Acute mood episodes have been characterized as pro-inflammatory states based on increased peripheral levels of interleukin (IL)-6 and tumour necrosis factor alpha (TNF-α) during depression, and IL-2, IL-16, IL-4, and TNF-α during mania (Oritz-Dominguez et al., 2007; Brietzke et al., 2009). Traditionally, the brain has been considered an immune-privileged site. However, this doctrine has been contested by recent evidence of communication between the peripheral immune system and the brain (Garay & McAllister, 2010), and the effect of systemic inflammation on brain function (D'Mello & Swain,

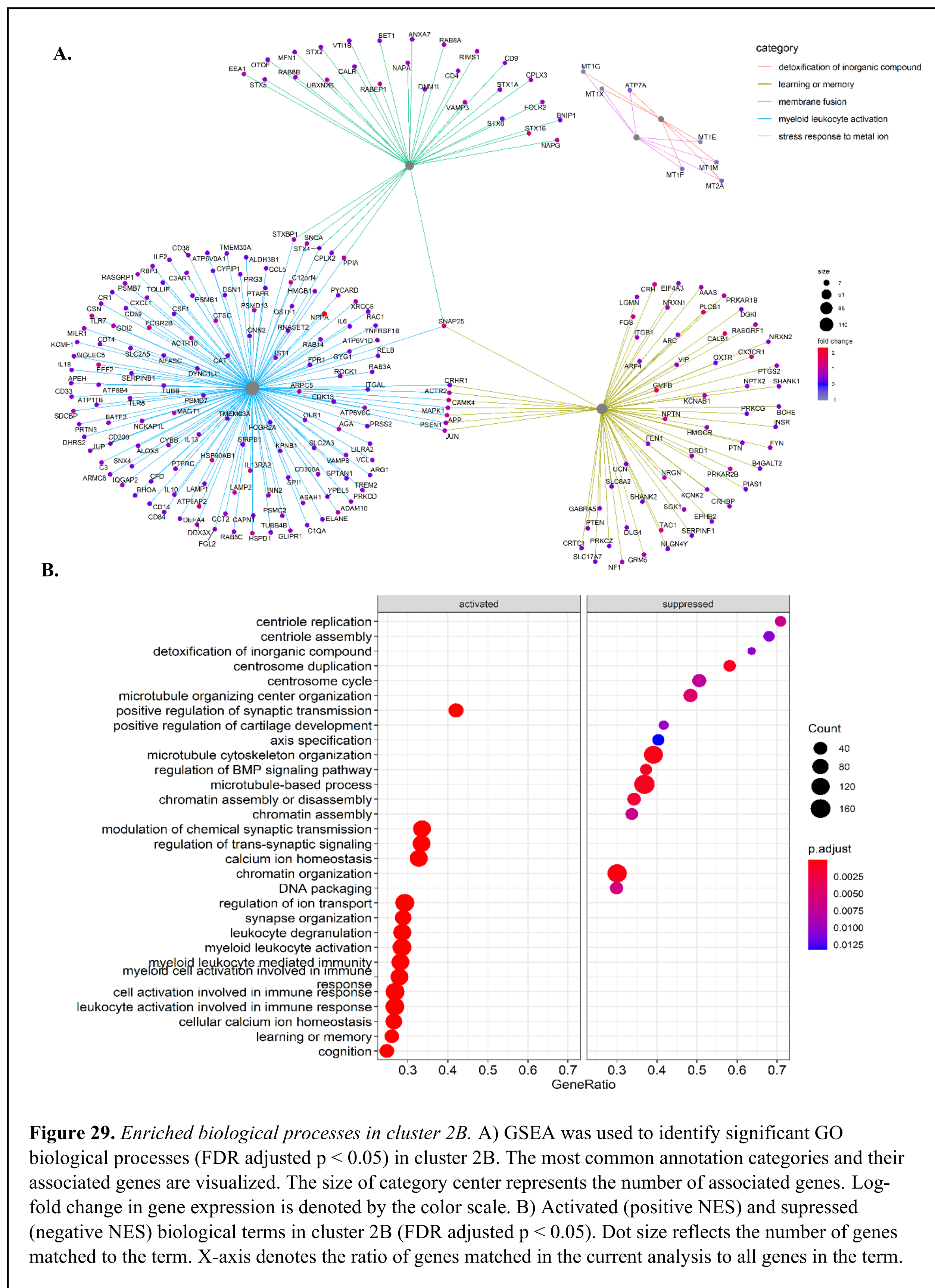


**Figure 29.** *Enriched biological processes in cluster 2B.* A) GSEA was used to identify significant GO biological processes (FDR adjusted p < 0.05) in cluster 2B. The most common annotation categories and their associated genes are visualized. The size of category center represents the number of associated genes. Log-fold change in gene expression is denoted by the color scale. B) Activated (positive NES) and supressed (negative NES) biological terms in cluster 2B (FDR adjusted p < 0.05). Dot size reflects the number of genes matched to the term. X-axis denotes the ratio of genes matched in the current analysis to all genes in the term.

2016). Consequently, chronic inflammation can influence several brain systems, including serotonic transmission. In a pro-inflammatory state with elevated levels of IL-1, IL-6 and TNF-α, the activity of the tryptophan- and serotonin-degrading enzyme 2,3-dioxygenase is upregulated leading to a decrease in serotonin and substrates for the synthesis of serotonin (Müller & Schwarz, 2007). Moreover, there are several repercussions of inflammation on synaptic plasticity, DNA integrity, myelination, and neurotropic support of neurons (Kapczinski et al., 2008). The upregulation of processes related to the detoxification of inorganic compounds and stress response to metal ions is relevant to neuroprogression. The brain has a high content of metal ions (e.g. calcium, copper, and iron) that act as oxidizing agents (Muneer, 2016). During acute mood episodes, a substantial release of neurotransmitters floods the synaptic space, exceeding the capacity of enzyme degradation and eventually leading to cellular toxicity. Dopamine auto-oxidizes in the presence of $Fe^{2+}$, forming hydroxyl ions (-OH), which are among the most reactive free radicals (Halliwell, 1992). Glutamate can promote Ca2+ influx into neurons and the activation of nitric oxide synthase, which forms the powerful oxidate peroxynitire (ONOO-) (Lafon-Cazal et al., 1993). Together, imbalances in neurotransmitters and their interaction with metal ions contribute to oxidative stress in cells, leading to mitochondrial dysfunction and ultimately apoptosis.

Despite having the youngest median sample age (30.5 years), the greatest proportion of samples under 10 years of age (38.9 %) among the three clusters compared, and BD samples with the shortest mean illness duration (15.4 years), cluster 2B still demonstrated the greatest enrichment of biological processes associated with late-stage impairments in BD. There are a couple possible explanations. Microglia are one of the major sources of cytokines in the central nervous system (CNS). Post-mortem studies have described decreased size and numbers of microglia in late-stage BD individuals (Watkins et al., 2014). In order to reconcile this finding with reports of increased cytokine production, Streit (2006) hypothesizes that persistent microglial activation in BD wears out their cellular machinery, resulting in cellar death or senescence in the long term. Similarly, serum assays evaluating the production of cytokines following lymphocyte simulation with lipopolysaccharide (LPS) reported lower production of interferon gamma (IFN-γ) in BD patients compared with controls (Liu et al., 2004; Hope et al., 2009), with no difference in IL-4 and IL-1β (Mota et al., 2013; Knijff et al., 2007). The apparent paradox of chronic low-grade inflammation and reduced cytokine production in response to

acute stimulation in BD can be explained within the AA framework. Namely, it is a symptom of an overstimulated immune system that is not able to respond to additional stimuli (Rizzo et al., 2014). Taken together, these findings suggest that genes regulating immune activation are over-expressed earlier on in BD, while their maladaptive effects propagate over the course of the disorder and eventually impair the cells that produce them. This is one potential reason why immune processes are enriched in cluster 2B relative to older clusters. In Chapter 2, we found that microglia were upregulated in early development of primary visual cortex (V1C) and subsequently downregulated in ageing. The importance of microglia and their immune-related functions in sculpting premature circuits and mediating synaptic plasticity were also discussed. Perhaps younger BD samples from individuals with a shorter average duration of illness matched with younger control samples because of an overlap between adaptive and maladaptive upregulation of immune-related genes.

It should be noted that cluster 2C and 2D are enriched for other biological processes potentially implicated in neuroprogression (Figure 29). Specifically, over-expressed genes in 2C are linked to mitochondrial processes and metal ion homeostasis. Besides cellular energy production, mitochondria are crucial for regulating inflammatory response. It is hypothesized that oxidative stress, in which metal ions are active players, leads to mitochondrial dysfunction and the activation of the NLRP3 (Nod-like receptor, pyrin domain-containing 3) inflammasome (Gurung et al., 2015). Cluster 2D was enriched for several processes involved in cellular division. While it is unlikely that cellular division is taking place in the brain, terms such as DNA replication, DNA repair, and DNA conformation change may be relevant. It is well known that oxidative stress is associated with DNA damage, including telomere shortening, which has been reported in peripheral cells for multiple psychiatric disorders (Ridout et al., 2015). Interestingly, the processes upregulated in each successively older cluster loosely follows a pathophysiological progression: immune activation, followed by mitochondrial dysfunction, followed by DNA damage and turnover.

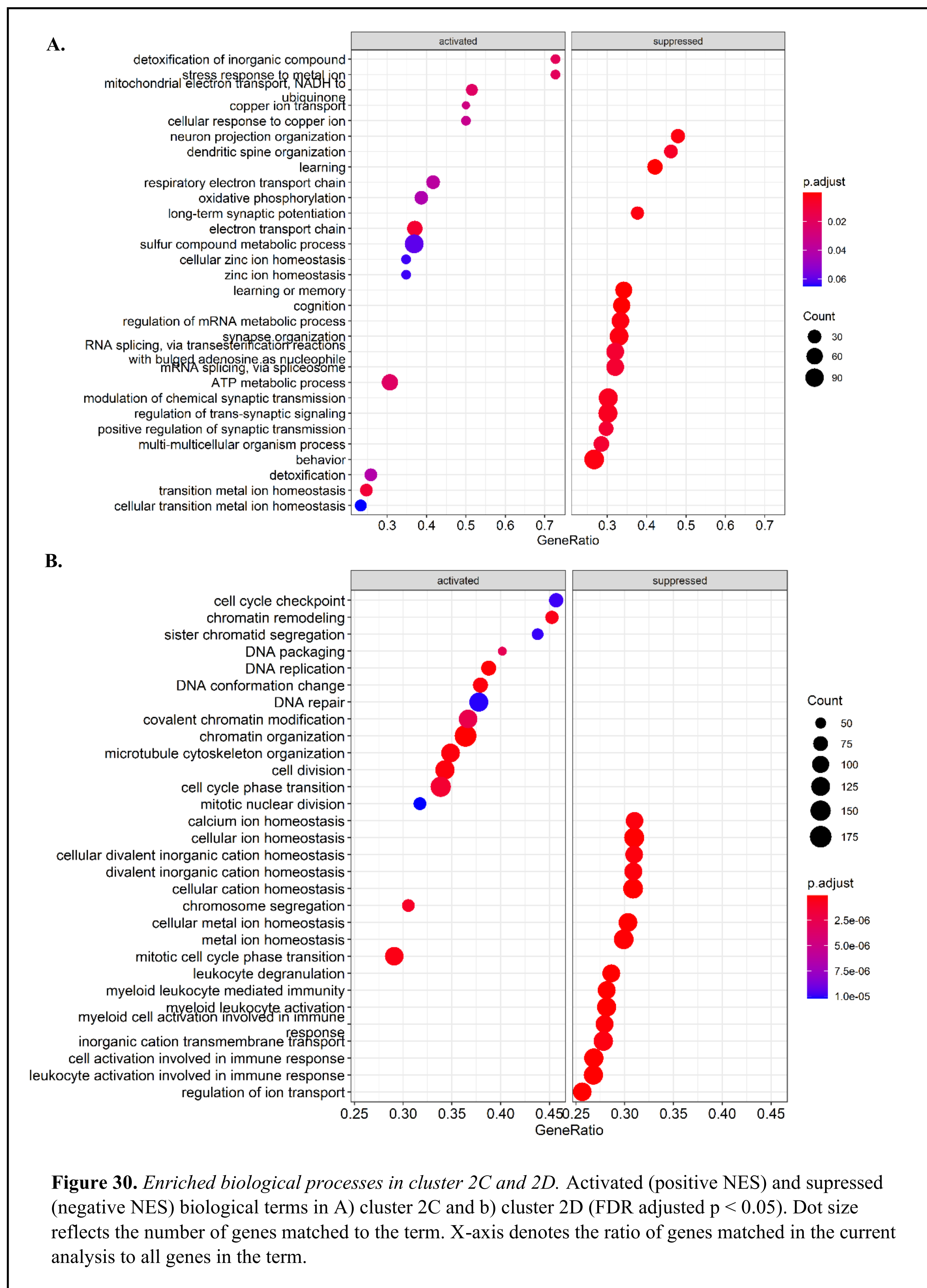


**Figure 30.** *Enriched biological processes in cluster 2C and 2D.* Activated (positive NES) and supressed (negative NES) biological terms in A) cluster 2C and b) cluster 2D (FDR adjusted p < 0.05). Dot size reflects the number of genes matched to the term. X-axis denotes the ratio of genes matched in the current analysis to all genes in the term.

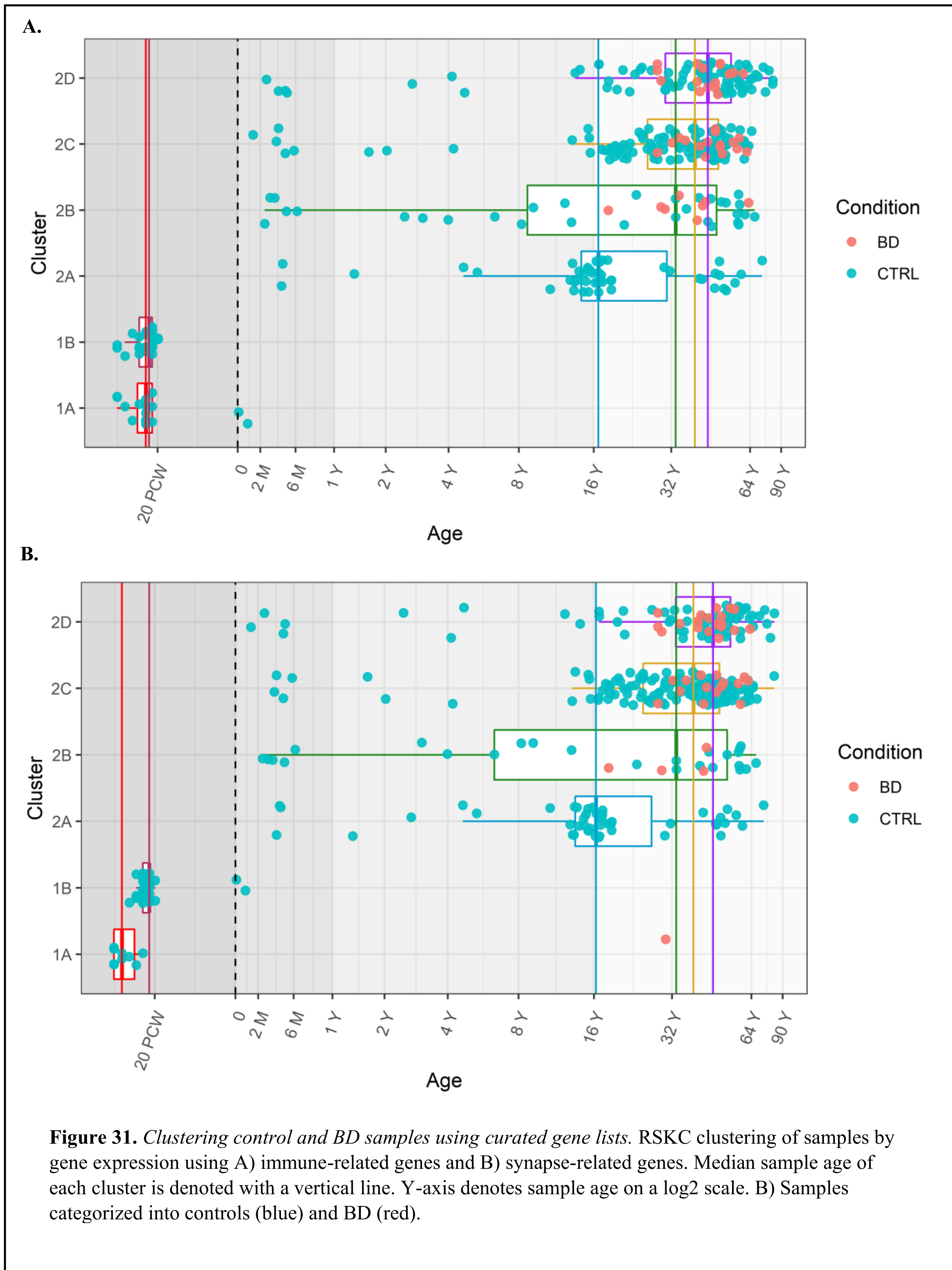


**Figure 31.** *Clustering control and BD samples using curated gene lists.* RSKC clustering of samples by gene expression using A) immune-related genes and B) synapse-related genes. Median sample age of each cluster is denoted with a vertical line. Y-axis denotes sample age on a log2 scale. B) Samples categorized into controls (blue) and BD (red).

### 5.3.3 Clustering control and BD samples using curated gene lists

To test whether the transcriptomic pattern that would distinguish BD samples from controls is getting lost in "noise" of other genes, RSKC was performed using two curated lists of genes. The first list includes genes involved in immune responses and was taken from the GO annotated biological process (GO:0006955) (Ashburner et al., 2000). This list intersected with 1996 genes in the current study and was used to cluster control and BD samples together. Figure 31A demonstrates no significant differences from clustering with the full set of 10082 genes. The second list consisted of genes involved in synapse structure and function, acquired from the SynGO knowledgebase. Several post-synaptic genes in this curated list have significant associations with BD as identified in GWAS data (Koopmans et al., 2019). Still, no major changes in BD cluster designation were observed, with the exception of some BD samples switching from cluster 2B to clusters 2C and 2D. These results suggest that clustering of BD samples is driven by nuanced patterns of gene expression that likely span the transcriptome.

## 5.4 Limitations and future directions

Not performing batch-correction during data preprocessing resulted in the formation of a BD-only cluster following RSKC. This is the result one might generally expect when clustering disease and control samples. However, this is also the result that would be expected from clustering expression data from two separate studies, with differences in tissue selection protocol, RNA extraction method, microarray platform used, data cleaning methods, etcetera. For instance, the control samples included in the BD study also clustered with BD samples when batch-correction was omitted. The current study used batch-correction via *ComBat* because batch effects are one of the main sources of variation which hinders *in silico* data merging (Lazar et al., 2013). One limitation was absence of BD samples in both studies, meaning the results of *ComBat* could not be evaluated for clustering by condition and absence of clustering by study (Taminau et al., 2012). Nevertheless, Figure 27 suggests that meaningful differences in gene expression between conditions are preserved,

There is an experimental design that avoids the pitfalls of data merging and can be implemented in a future study. Each dataset would be clustered by RSKC separately, and then the average expression for genes across samples within a clusters would be compared to clusters

in the other condition, with the goal of identifying clusters with the most similar expression patterns. This could then be followed by a comparison of sample age between control and BD subgroups. One limitation is that this approach would likely use a linear method such as Pearson's correlation to make comparisons, meaning some high-dimensional information would be lost.

## 5.5 Summary

The purpose of this chapter was to demonstrate the prospective utility of developmental clustering for addressing a hypothesis-based research question. Individuals with bipolar disorder (BD) suffer from physiological and cognitive impairments that mirror the decline seen in ageing. These observations have been formalized in the conceptual framework of accelerated ageing (AA). To test the hypothesis that the transcriptomic profile of the bipolar brain is comparable to an aged brain many years older, bipolar and developmental controls were clustered together according to similar patterns of gene expression, with the expectation that bipolar samples would cluster with older controls. This prediction was not supported; however, bipolar samples did cluster with developmental controls of various ages. To unpack the differences between the three bipolar subclusters, differential expression analysis was used to identify genes uniquely over- and under-expressed in each, and Gene Set Enrichment Analysis (GSEA) was used to identify biological processes associated with these genes. The youngest cluster containing BD samples was enriched for immune-related functions relative to the older clusters. It was hypothesized that this reflects the earlier production of inflammatory molecules in BD, followed by maladaptive effects that eventually impair the cells that produce them. The other two clusters were enriched for processes that may indicate oxidative stress, mitochondrial dysfunction, and DNA damage.

# Chapter 6. General Discussion

## 6.1 Summary of Main Findings

### 6.1.1 Developmental clusters in primary visual cortex

The current study used modern, data-driven clustering to identify groups of developmental controls with similar patterns of transcriptomic expression. Robust and sparse k-means clustering (RSKC) separated individuals into clusters that exhibited developmental progression across the lifespan. In primary visual cortex (V1C), postnatal clusters had considerable overlap within respect to sample age. RSKC also identified weighted features that drove the separation of developmental clusters. Features separating prenatal and postnatal samples included *PLP1* (proteolipid protein 1) and *MBP* (myelin basic protein). The myelination of neurons by oligodendrocytes in known to increase precipitously following birth (Williamson & Lyons, 2018), and the developmental increase in intra-cortical myelin is thought to act as a structural break on critical period plasticity in V1C (Bavelier et al., 2010). Features extracted during clustering of postnatal samples highlighted markers of distinct inhibitory interneuron subpopulations, such as *PVALB* (parvalbumin) and *STT* (somatostatin). Both of these neuronal populations play critical roles in circuits underlying vision, and undergo age-related changes associated with critical period plasticity and visual decline (Fagiolini & Hensch, 2000; Bradshaw et al., 2018; Scheyltjens et al., 2018). Differential expression (DE) analysis was used to unpack genes that were over-expressed and under-expression in each developmental cluster. When combined with an external database of cell-type markers (Kelley et al., 2018), we were able to identify an up-regulation of microglia-specific genes during perinatal development, and a subsequent down-regulation during ageing. This is consistent with the literature, which demonstrates the role of microglia in perinatal synaptic refinement and experience-dependent synaptic plasticity, as well as the contribution of microglial dysfunction and senescence in neurodegeneration. Overall, clustering pulled together several molecular mechanisms describing the development and ageing of the visual cortex and associated them with a novel global pattern of expression which corresponded to known milestones in visual development.

### 5.1.2 Developmental clusters in multiple brain regions

Transcriptomic data from 16 brain regions were independently surveyed using our analysis pipeline, revealing spatially distinct patterns of developmental clusters. Notably, gene

expression in subcortical structures did not separate samples into postnatal developmental clusters. This suggests that age is not a significant driver of transcriptional profiles in subcortical structures following prenatal development. There was considerable variation in median cluster age across regions, implying that the appearance of transcriptional patterns also varies from region to region. This finding may be connected to the heterogeneous maturational sequence of the human brain observed in MRI studies (Gogtay et al., 2004). Several feature genes were assigned a high weight across regions, suggesting that they are common drivers of change during brain development. Features genes separating prenatal and postnatal clusters included those that regulate transcription, such as *HIST1H3B* (H3 clustered histone 2), *HIST1H3C* (H3 clustered histone 3), and *ELAVL4* (ELAV like RNA binding protein 4). On the other hand, common features of postnatal development were involved in synaptic signalling, like *PDYN* (prodynorphin), *HCN1* (hyperpolarization activated cyclic nucleotide gated potassium channel), and AKAP5 (A-kinase anchoring protein 5). Hierarchical clustering of regions based on feature weights demonstrated that while transcriptional differences between regions might decrease after birth (Kang et al., 2011), the influence of genes driving their development grows more varied. RSKC was also used to cluster brain regions with similar patterns of transcriptomic expression. Interestingly, region clusters roughly charted anatomical separation, and provided a more nuanced separation of neocortical regions compared to the original analysis of this data (Kang et al., 2011). The feature genes driving both postnatal development and separation of neocortical areas were highly enriched for biological processes related to neuronal communication, agreeing with the diversity in neuronal populations seen across cortical regions, and the functional maturation of neurons over development and ageing. Samples within each regional cluster were pooled in order to refine our developmental analysis. The results surprising demonstrated that developmental clustering in DFC, MFC, OFC and M1C, S1C, VFC is driven jointly by sample age and sample sex.

### 6.1.3 Clustering bipolar cases and developmental controls

Bipolar and developmental controls were clustered together according to similar patterns of gene expression. Given the etiological framework of accelerated ageing (AA), it was hypothesized that bipolar samples would cluster with older controls. This prediction was not supported; however, bipolar samples did cluster with developmental controls of various ages. To

unpack the differences between the three bipolar subclusters, differential expression analysis was used to identify genes uniquely over- and under-expressed in each, and Gene Set Enrichment Analysis (GSEA) was used to identify biological processes associated with these genes. The youngest cluster containing BD samples was enriched for immune-related functions relative to the older clusters. It was hypothesized that this reflects the earlier production of inflammatory molecules in BD, followed by maladaptive effects that eventually impair the cells that produce them. The other two clusters were enriched for processes that may indicate oxidative stress, mitochondrial dysfunction, and DNA damage.

## 6.2 Significance

### 6.2.1 Unsupervised clustering identifies stable patterns of gene expression

The transcriptome describes the full range of mRNA molecules that are expressed in a tissue at a given point in time. Large-scale studies have used high-throughput sequencing to survey gene expression across multiple brain areas and developmental periods (Kang et al., 2011; Colantuoni et al., 2011). These datasets provide some of the most comprehensive accounts we have of the biological events underlying human brain development and ageing. Numerous groups of genes have been recognized for their involvement in neurogenesis, circuit refinement, and myelination, as well as neurodegeneration and disease. Nevertheless, most analysis methods and descriptive models assume that developmental change in gene expression is continuous and uninterrupted. Perhaps this bias is often overlooked because the emphasis is on *what* is changing during development rather than *how* development itself is changing. Indeed, despite the richness of transcriptomic data and its capacity to recapitulate higher-order functions, few have used it to understand the dynamics of brain development.

Gene expression is determined by the complex, high-dimensional interactions of the gene regulatory network (GRN), which encompasses virtually the entire genome (Huang, 2009). Dynamic systems theory states that the interactions of components in any complex systems will converge on certain stable patterns, also known as attractor states (Prigogine & Nicolis, 1971). To approximate these stable states, the current study leveraged robust and sparse k-means clustering to identify tissue samples with similar patterns of gene expression across the transcriptome. Sample ages were then used to visualize when in developmental time these stable

patterns are present. The resulting model describes the developmental dynamics of the brain transcriptome as a series of overlapping states that progresses across the lifespan.

To our knowledge, this is the first study to use unsupervised clustering expressly to identify subgroups within developmental controls. In transcriptomics, clustering is mostly used to identify genes with similar expression across samples. The inverse is only performed in studies of neuropsychiatric disorders (Gandal et al., 2018), where significant differences between individual cases are expected, or single-cell datasets (Tasic et al., 2016), where the goal is to identify populations of different cell-types. Both of these examples represent the use of clustering by gene expression to identify stable states, whether it is a clinical phenotype or cell type. Why has a similar approach not been applied to developmental datasets? Perhaps it is for the same reasons why the selection criteria for brain donors is so stringent; namely, each sample is considered an error-laden instance of a common developmental program to be characterized. Clustering algorithms like RSKC group together samples that are "close" in a multidimensional feature space defined by gene expression. Unlike regression-based models, differences are not fitted to explanatory variables like age. This allows multiple heterogeneous patterns of gene expression to be identified in samples that overlap in developmental time.

### 6.2.2 Overlapping patterns of gene expression

The overlapping nature of developmental clusters in the current study has several implications. First, it suggests that traditional development periods, defined by anatomical and behavioural milestones, are not described by a single transcriptomic "phenotype." Instead, developmental periods may be defined by two or more patterns of gene expression across the transcriptome. Such patterns are not limited to a particular section of the lifespan, and may be recurring. For example, the transcriptional profile matching samples from cluster 2B in V1C described early development at 120 days as well as ageing at 82 years (Figure 7). Colantuoni et al. (2011) also found that numerous changes in infancy were mirrored by changes in ageing.

While it is impossible to determine with the current experimental design (see Limitations), an intriguing possibility is that cortical tissue transitions between transcriptomic states within an individual. This idea is not unlike McIntosh (2011)'s account of brain activity visiting several metastable network configurations spontaneously and converging on a configuration that meets the demands of the current task. If distinct, overlapping patterns of gene

expression represent metastable transcriptomic states, it is possible that the brain transitions to states that suit current biological needs. This notion is somewhat supported by current study, which demonstrated that only a subset of genes is responsible for the differences between clusters, and the biological principle that minor changes in a few proteins can have cascading effects that ultimately impact the whole organism. Neurodevelopmental disorders may be conceptualized as the failure to transition into suitable transcriptomic states, or remaining "stuck" in an unsuitable state, perhaps in part due to genetic mutations that affect gene expression. For example, normal brain development involves a period of synaptic pruning early in life that is likely orchestrated by a particular pattern of gene expression. Individuals with autism spectrum disorder (ASD) show hypomethylation and consequent over-expression of genes implicated in synaptic pruning (Nardone et al., 2014), suggesting that they may inhabit a transcriptomic state of pruning for a maladaptive period of time.

Lastly, overlapping patterns of gene expression have implications for studying disorders. In order to identify genes that are differentially expressed in disorders, a comparison is often made to a group of age-matched controls. As the current work suggests, however, healthy brain samples of similar age may be represented by more than one distinct transcriptomic pattern. Therefore, it is prudent to compare gene expression first and assess the similarity of samples second. This was one reason why bipolar cases were clustered with all available developmental controls in Chapter 4.

### 6.2.3 Biological significance of clusters and features

RSKC provides a new way to identify developmentally regulated genes of interest. Other approaches, such as co-expression analysis, only detect groups of genes that change together. No group is privileged over another, and they must be visualized and manually selected for further analysis. Feature selection is built into the RSKC algorithm, and presents an unsupervised method of identifying genes with the greatest developmental variance without the linear assumption of other dimensionality reduction approaches (e.g. principal component analysis). The current study identified several feature genes that were directly implicated in neurobiological development. Furthermore, RSKC identified feature genes not previously characterized in development, but that may be implicated developmentally regulated processes (see section 2.3.3). Differences in feature weights were enough to separate functionally and

anatomically distinct brain regions. Additionally, features driving the segregation of postnatal clusters and cortical brain regions were highly enriched for biological processes related to neuronal communication, agreeing with the diversity in neuronal populations seen across cortical regions, and the functional maturation of neurons over development and ageing.

## 6.3 Limitations

While the current study was informed by dynamic systems principles, no formal DST mathematical approaches were utilized (e.g. catastrophe theory). Furthermore, DST is most often used to analyze behavioural or imaging data collected from individuals over time. Time-series data captures the variability and fluctuation of intra-individual measures, which can be analyzed for the emergence of a stable patterns that are then compared between individuals. This experimental design also allows for the detection of transitions between stable states. Unfortunately, due to the cross-sectional design of nearly all human developmental transcriptomic studies, one cannot determine whether differences in gene expression reflect stable inter-individual differences or intrinsic fluctuations over some time scale (hour, days, etc.). This is a general limitation of studies that make use of post-mortem tissue. While the ideal approach to understanding the source and function of biological variability would involve measuring gene expression at a cellular resolution across multiple developmental period within the same individual, this is not yet possible with current technology. Instead, we employed unsupervised clustering, which treats *all* variance as an essential element of the higher-order patterns that constitute the system. Unfortunately, this still does not allow us to speak meaningfully about transitions in the data.

One objective of the current study was to address whether human brain development and ageing can be described as a series of stable, biological states. Our approach involved using clustering to identify developmental samples with similar patterns of gene expression. The distribution of sample clusters across the lifespan served as a proxy for when distinct transcriptomic states are present. However, clustering can be variable and highly contingent on several factors. Some challenges include the selection of an appropriate clustering algorithm and the desired number of clusters. Both of these steps are subjective and have direct consequences on the outcomes of our descriptive model. Another factor is the number and distribution of samples in the original dataset, as demonstrated in Section 4.3.1. Even clustering the same data

can produce different results with each iteration. For these reasons and more, clusters cannot represent *de facto* biological states in human brain development. Instead, the success of the current study lies in demonstrating that the transcriptome *can* be modelled by a series of non-linear states, which also capture important biological mechanisms that may not have been apparent in standard linear analyses.

## 6.4 Future Directions

The current study used three microarray datasets in total for clustering (Kang et al., 2011; Colantuoni et al., 2011; Ryan et al., 2006). While many large-scale transcriptomic stuides use microarray technology, the development of RNA-seq has shown that microarrays have several limitations, including a narrow dynamic range that can lead to underestimation of differential gene expression (Dillman et al., 2017). To better detect developmental differences, future studies should use data generated by RNA-seq. Furthermore, data across developmental studies should be pooled in order to refine the boundaries of developmental clusters.

To begin to address whether transitions may occur between clusters, an algorithm called PHATE (potential of heat diffusion for affinity-based transition embedding) may be useful. PHATE constructs a non-linear embedding of high dimensional data that simultaneously denoises the data and emphasizes the continuous nature of any underlying progression and trajectories (Moon et al., 2017). The branching structures produced by this algorithm identify points where developmental trajectories may diverge or converge. Figure 32 shows the results of PHATE applied to samples pooled across regions in the Kang et al. (2011) dataset. The global structure of this data suggests that prenatal samples lie on a distinct branch which transitions into a much broader postnatal branch. The color-coding of samples by RSKC cluster reveals heterogenous substructures within the second branch. With more samples and well-defined gene expression values from RNA-seq, this method could reveal the transitions between developmental clusters.

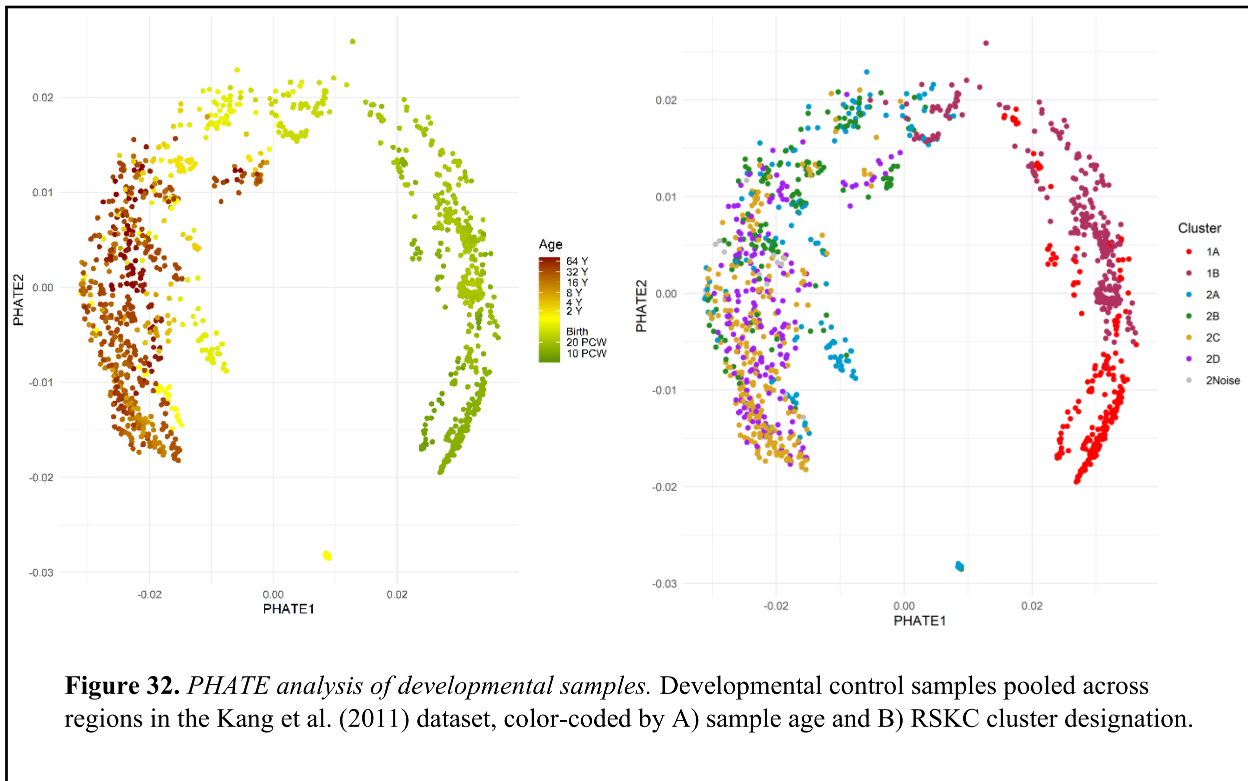


**Figure 32.** *PHATE analysis of developmental samples.* Developmental control samples pooled across regions in the Kang et al. (2011) dataset, color-coded by A) sample age and B) RSKC cluster designation.

# Chapter 7. References

Adams, D. L., & Horton, J. C. (2002). Shadows cast by retinal blood vessels mapped in primary visual cortex. *Science, 298*(5593), 572-576.

Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J., & Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature, 490*(7419), 226-231.

Adolph, K. E., Berger, S. E., Kuhn, D., & Siegler, R. S. (2006). *Handbook of child psychology: Vol. 2*. Cognition, perception, and language.

Agladze, K. I., Krinsky, V. I., & Pertsov, A. M. (1984). Chaos in the non-stirred Belousov–Zhabotinsky reaction is induced by interaction of waves and stationary dissipative structures. *Nature, 308*(5962), 834-835.

Akamatsu, W., Fujihara, H., Mitsuhashi, T., Yano, M., Shibata, S., Hayakawa, Y., ... & Takahashi, T. (2005). The RNA-binding protein HuD regulates neuronal cell identity and maturation. *Proceedings of the National Academy of Sciences*, *102*(12), 4625-4630.

Akbik, F., Cafferty, W. B., & Strittmatter, S. M. (2012). Myelin associated inhibitors: a link between injury-induced and experience-dependent plasticity. *Experimental neurology, 235*(1), 43-52.

Allen, J. S., Bruss, J., Brown, C. K., & Damasio, H. (2005). Normal neuroanatomical variation due to age: the major lobes and a parcellation of the temporal region. *Neurobiology of aging*, *26*(9), 1245-1260.

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., & Hamosh, A. (2015). OMIM. org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic acids research, 43*(D1), D789-D798.

Aminoff, S. R., Hellvin, T., Lagerberg, T. V., Berg, A. O., Andreassen, O. A., & Melle, I. (2013). Neurocognitive features in subgroups of bipolar disorder. *Bipolar disorders*, *15*(3), 272-283.

Anderson, K. D., Merhege, M. A., Morin, M., Bolognani, F., & Perrone-Bizzozero, N. I. (2003). Increased expression and localization of the RNA-binding protein HuD and GAP-43 mRNA to cytoplasmic granules in DRG neurons during nerve regeneration. *Experimental neurology, 183*(1), 100-108.

Angelova, D. M., & Brown, D. R. (2019). Microglia and the aging brain: are senescent microglia the key to neurodegeneration?. *Journal of neurochemistry, 151*(6), 676-688.

Antonini, A., & Stryker, M. P. (1996). Plasticity of geniculocortical afferents following brief or prolonged monocular occlusion in the cat. *Journal of Comparative Neurology, 369*(1), 64-82.

Antonini, A., Fagiolini, M., & Stryker, M. P. (1999). Anatomical correlates of functional plasticity in mouse visual cortex. *Journal of Neuroscience, 19*(11), 4388-4406.

Arambula, S. E., & McCarthy, M. M. (2020). Neuroendocrine-Immune Crosstalk Shapes Sex-Specific Brain Development. *Endocrinology*, *161*(6), bqaa055.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Harris, M. A. (2000). Gene ontology: tool for the unification of biology. *Nature genetics,* 25(1), 25-29.

Bae, B. I., Tietjen, I., Atabay, K. D., Evrony, G. D., Johnson, M. B., Asare, E., ... & Overman, L. (2014). Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. *Science, 343*(6172), 764-768.

Barbosa, I. G., Bauer, M. E., Machado-Vieira, R., & Teixeira, A. L. (2014). Cytokines in bipolar disorder: paving the way for neuroprogression. *Neural plasticity*, *2014*.

Bavelier, D., Levi, D. M., Li, R. W., Dan, Y., & Hensch, T. K. (2010). Removing brakes on adult brain plasticity: from molecular to behavioral interventions. *Journal of Neuroscience, 30*(45), 14964-14971.

Bavelier, D., Levi, D. M., Li, R. W., Dan, Y., & Hensch, T. K. (2010). Removing brakes on adult brain plasticity: from molecular to behavioral interventions. *Journal of Neuroscience, 30*(45), 14964-14971.

Bellivier, F., Golmard, J. L., Rietschel, M., Schulze, T. G., Malafosse, A., Preisig, M., ... & Leboyer, M. (2003). Age at onset in bipolar I affective disorder: further evidence for three subgroups. *American Journal of Psychiatry*, *160*(5), 999-1001.

Belmaker, R. H. (2004). Bipolar disorder. *New England Journal of Medicine*, *351*(5), 476-486.

Berardi, N., Pizzorusso, T., & Maffei, L. (2000). Critical periods during sensory development. *Current opinion in neurobiology, 10*(1), 138-145.

Berk, M., Conus, P., Lucas, N., Hallam, K., Malhi, G. S., Dodd, S., ... & McGorry, P. (2007). Setting the stage: from prodrome to treatment resistance in bipolar disorder. *Bipolar disorders*, *9*(7), 671-678.

Bernal, J. (2000). Thyroid Hormones in Brain Development and Function Endotext. *South Dartmouth, MA*

Berry, W. D. (1993). *Understanding regression assumptions* (Vol. 92). Sage Publications.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine, 34*(4), 537-541.

Blackburn, E. H. (1991). Structure and function of telomeres. *Nature*, *350*(6319), 569-573.

Bolognani, F., Qiu, S., Tanner, D. C., Paik, J., Perrone-Bizzozero, N. I., & Webber, E. J. (2007). Associative and spatial learning and memory deficits in transgenic mice overexpressing the RNA-binding protein HuD. *Neurobiology of learning and memory*, *87*(4), 635-643.

Bolognani, F., Tanner, D. C., Merhege, M., Deschênes-Furry, J., Jasmin, B., & Perrone-Bizzozero, N. I. (2006). In vivo post-transcriptional regulation of GAP-43 mRNA by overexpression of the RNA-binding protein HuD. *Journal of neurochemistry, 96*(3), 790-801.

Bouhours, B., Gjoni, E., Kochubey, O., & Schneggenburger, R. (2017). Synaptotagmin2 (Syt2) drives fast release redundantly with Syt1 at the output synapses of parvalbumin-expressing inhibitory neurons. *Journal of Neuroscience, 37*(17), 4604-4617.

Boulanger, L. M. (2009). Immune proteins in brain development and synaptic plasticity. *Neuron, 64*(1), 93-109.

Bradshaw, K. P., Velez, D. X. F., Habeeb, M., & Gandhi, S. P. (2018). Precocious deposition of perineuronal nets on parvalbumin inhibitory neurons transplanted into adult visual cortex. *Scientific reports, 8*(1), 1-10.

Brainerd, C. J. (1978). The stage question in cognitive-developmental theory. *Behavioral and Brain Sciences, 1*(2), 173-182.

Breen, M. S., Ozcan, S., Ramsey, J. M., Wang, Z., Ma'ayan, A., Rustogi, N., ... & Bahn, S. (2018). Temporal proteomic profiling of postnatal human cortical development. *Translational psychiatry, 8*(1), 267.

Breuss, M. W., Leca, I., Gstrein, T., Hansen, A. H., & Keays, D. A. (2017). Tubulins and brain development–The origins of functional specification. *Molecular and Cellular Neuroscience, 84*, 58-67.

Brietzke, E., Stertz, L., Fernandes, B. S., Kauer-Sant'Anna, M., Mascarenhas, M., Vargas, A. E., ... & Kapczinski, F. (2009). Comparison of cytokine levels in depressed, manic, and euthymic patients with bipolar disorder. *Journal of affective disorders*, *116*(3), 214-217.

Brock, G., Pihur, V., Datta, S., & Datta, S. (2011). clValid, an R package for cluster validation. *Journal of Statistical Software*, *25*(4)

Bronicki, L. M., & Jasmin, B. J. (2013). Emerging complexity of the HuD/ELAVl4 gene; implications for neuronal development, function, and dysfunction. *Rna*, *19*(8), 1019-1037.

Brown, N. C., Andreazza, A. C., & Young, L. T. (2014). An updated meta-analysis of oxidative stress markers in bipolar disorder. *Psychiatry Research*, *218*(1-2), 61-68.\

Burdick, K. E., Russo, M., Frangou, S., Mahon, K., Braga, R. J., Shanahan, M., & Malhotra, A. K. (2014). Empirical evidence for discrete neurocognitive subgroups in bipolar disorder: clinical implications. *Psychological medicine*, *44*(14), 3083.

Burke, E. E., Chenoweth, J. G., Shin, J. H., Collado-Torres, L., Kim, S. K., Micali, N., ... & Chen, H. Y. (2020). Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs. *Nature communications*, *11*(1), 1-14.

Bystron, I., Blakemore, C., & Rakic, P. (2008). Development of the human cerebral cortex: Boulder Committee revisited. *Nature Reviews Neuroscience, 9*(2), 110-122.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods, 3*(1), 1-27.

Carlyle, B. C., Kitchen, R. R., Kanyo, J. E., Voss, E. Z., Pletikos, M., Sousa, A. M., ... & Nairn, A. C. (2017). A multiregional proteomic survey of the postnatal human brain. *Nature neuroscience, 20*(12), 1787-1795.

Cavazzoni, P., Grof, P., Duffy, A., Grof, E., Müller-Oerlinghausen, B., Berghöfer, A., ... & Hajek, T. (2007). Heterogeneity of the risk of suicidal behavior in bipolar-spectrum disorders. *Bipolar disorders*, *9*(4), 377-385.

Cerghet, M., Skoff, R. P., Swamydas, M., & Bessert, D. (2009). Sexual dimorphism in the white matter of rodents. *Journal of the neurological sciences, 286*(1-2), 76-80.

Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E., & Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature, 453*(7194), 544-547.

Charney, A. W., Ruderfer, D. M., Stahl, E. A., Moran, J. L., Chambert, K., Belliveau, R. A., ... & Bromet, E. J. (2017). Evidence for genetic heterogeneity between clinical subtypes of bipolar disorder. *Translational psychiatry*, *7*(1), e993-e993.

Cheng, C., Lau, S. K., & Doering, L. C. (2016). Astrocyte-secreted thrombospondin-1 modulates synapse and spine defects in the fragile X mouse model. *Molecular Brain, 9*(1), 1-15.

Cheng, C., Yu, Z., Zhao, S., Liao, Z., Xing, C., Jiang, Y., ... & Wang, X. (2017). Thrombospondin-1 gene deficiency worsens the neurological outcomes of traumatic brain injury in mice. *International journal of medical sciences, 14*(10), 927.

Cheng, V. Y., Martin, L. J., Elliott, E. M., Kim, J. H., Mount, H. T., Taverna, F. A., ... & Wafford, K. A. (2006). α5GABAA receptors mediate the amnestic but not sedative-hypnotic effects of the general anesthetic etomidate. *Journal of Neuroscience, 26*(14), 3713-3720.

Colantuoni, C., Lipska, B. K., Ye, T., Hyde, T. M., Tao, R., Leek, J. T., ... & Kleinman, J. E. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature, 478*(7370), 519-523.

Collins, W. A., & National Research Council. (1984). Cognitive Development In School-Age Children: Conclusions And New Directions. *In Development During Middle Childhood: The Years from Six to Twelve.* National Academies Press (US).

Colpo, G. D., Leffa, D. D., Koehler, C. A., Kapczinski, F., Quevedo, J., & Carvalho, A. F. (2015). Is bipolar disorder associated with accelerating aging? A meta-analysis of telomere length studies. *Journal of affective disorders*, *186*, 241-248.

Colton, C. A. (2009). Heterogeneity of microglial activation in the innate immune response in the brain. *Journal of neuroimmune pharmacology, 4*(4), 399-418.

Cowan, M., & Petri Jr, W. A. (2018). Microglia: immune regulators of neurodevelopment. *Frontiers in immunology, 9*, 2576.

Crump, C., Sundquist, K., Winkleby, M. A., & Sundquist, J. (2013). Comorbidities and mortality in bipolar disorder: a Swedish national cohort study. *JAMA psychiatry*, *70*(9), 931-939.

Cunningham, C. L., Martínez-Cerdeño, V., & Noctor, S. C. (2013). Microglia regulate the number of neural precursor cells in the developing cerebral cortex. *Journal of Neuroscience, 33*(10), 4216-4233.

D'Mello, C., & Swain, M. G. (2016). Immune-to-brain communication pathways in inflammation-associated sickness and depression. In *Inflammation-Associated Depression: Evidence, Mechanisms and Implications* (pp. 73-94). Springer, Cham.

da Silva, J., Goncalves-Pereira, M., Xavier, M., & Mukaetova-Ladinska, E. B. (2013). Affective disorders and risk of developing dementia: systematic review. *The British Journal of Psychiatry*, *202*(3), 177-186.

Dalmau, I., Finsen, B., Zimmer, J., González, B., & Castellano, B. (1998). Development of microglia in the postnatal rat hippocampus. *Hippocampus, 8*(5), 458-474.

Damani, M. R., Zhao, L., Fontainhas, A. M., Amaral, J., Fariss, R. N., & Wong, W. T. (2011). Age-related alterations in the dynamic behavior of microglia. *Aging cell, 10*(2), 263-276.

Datta, S., & Datta, S. (2006). Evaluation of clustering algorithms for gene expression data. *BMC bioinformatics, 7*(S4), S17.

Daubon, T., Léon, C., Clarke, K., Andrique, L., Salabert, L., Darbo, E., ... & Jeanne, A. (2019). Deciphering the complex role of thrombospondin-1 in glioblastoma development. *Nature communications, 10*(1), 1-15.

Dauth, S., Maoz, B. M., Sheehy, S. P., Hemphill, M. A., Murty, T., Macedonia, M. K., ... & Parker, K. K. (2017). Neurons derived from different brain regions are inherently different in vitro: a novel multiregional brain-on-a-chip. *Journal of neurophysiology*, *117*(3), 1320-1341.

Davis, S., & Meltzer, P. S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics, 23*(14), 1846-1847.

Daw, N. W. (1998). Critical periods and amblyopia. *Archives of ophthalmology, 116*(4), 502-505.

DiFrancesco, D. (1993). Pacemaker mechanisms in cardiac tissue. *Annual review of physiology, 55*(1), 455-472.

DiFrancesco, J. C., & DiFrancesco, D. (2015). Dysfunctional HCN ion channels in neurological diseases. *Frontiers in cellular neuroscience*, *9*, 71.

Dillman, A. A., Majounie, E., Ding, J., Gibbs, J. R., Hernandez, D., Arepalli, S., ... & Cookson, M. R. (2017). Transcriptomic profiling of the human brain reveals that altered synaptic gene expression is associated with chronological aging. *Scientific reports, 7*(1), 1-12.

do Prado, C. H., Rizzo, L. B., Wieck, A., Lopes, R. P., Teixeira, A. L., Grassi-Oliveira, R., & Bauer, M. E. (2013). Reduced regulatory T cells are associated with higher levels of Th1/TH17 cytokines and activated MAPK in type 1 bipolar disorder. *Psychoneuroendocrinology*, *38*(5), 667-676.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.

Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics, 4*(1), 95-104.

Fagiolini, M., & Hensch, T. K. (2000). Inhibitory threshold for critical-period activation in primary visual cortex. *Nature, 404*(6774), 183-186.

Fagiolini, M., Fritschy, J. M., Löw, K., Möhler, H., Rudolph, U., & Hensch, T. K. (2004). Specific GABAA circuits for visual cortical plasticity. *Science, 303*(5664), 1681-1683.

Feldman, D. H. (2004). Piaget's stages: the unfinished symphony of cognitive development. *New Ideas in Psychology, 22*(3), 175-231.

Fischer, K. W., & Bullock, D. (1981). Patterns of data: Sequence, synchrony, and constraint in cognitive development. *New Directions for Child and Adolescent Development, 1981*(12), 1-20.

Fischer, K. W., & Silvern, L. (1985). Stages and individual differences in cognitive development. *Annual Review of Psychology, 36*(1), 613-648.

Flavell, J. H. (1971). Stage-related properties of cognitive development. *cognitive Psychology, 2*(4), 421-453.

Foulquier, S., Caolo, V., Swennen, G., Milanova, I., Reinhold, S., Recarti, C., ... & Post, M. J. (2019). The role of receptor MAS in microglia-driven retinal vascular development. *Angiogenesis, 22*(4), 481-489.

Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., ... & Cevenini, E. (2007). Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mechanisms of ageing and development*, *128*(1), 92-105.

Fries, G. R., Pfaffenseller, B., Stertz, L., Paz, A. V. C., Dargél, A. A., Kunz, M., & Kapczinski, F. (2012). Staging and neuroprogression in bipolar disorder. *Current psychiatry reports*, *14*(6), 667-675.

Frisca, F., Colquhoun, D., Goldshmit, Y., Änkö, M. L., Pébay, A., & Kaslin, J. (2016). Role of ectonucleotide pyrophosphatase/phosphodiesterase 2 in the midline axis formation of zebrafish. *Scientific reports, 6*, 37678.

Fritz, G., Botelho, H. M., Morozova-Roche, L. A., & Gomes, C. M. (2010). Natural and amyloid self-assembly of S100 proteins: structural basis of functional diversity. *The FEBS journal, 277*(22), 4578-4590.

Funa, K., & Sasahara, M. (2014). The roles of PDGF in development and during neurogenesis in the normal and diseased nervous system. *Journal of Neuroimmune Pharmacology, 9*(2), 168-181.

Gandal, M. J., Zhang, P., Hadjimichael, E., Walker, R. L., Chen, C., Liu, S., ... & Shieh, A. W. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science, 362*(6420).

Garay, P. A., & McAllister, A. K. (2010). Novel roles for immune molecules in neural development: implications for neurodevelopmental disorders. *Frontiers in synaptic neuroscience, 2*, 136.

Garrity, A. G., Pearlson, G. D., McKiernan, K., Lloyd, D., Kiehl, K. A., & Calhoun, V. D. (2007). Aberrant "default mode" functional connectivity in schizophrenia. *American journal of psychiatry, 164*(3), 450-457.

Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. A. (2004). affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinforma. 20*, 307–315.

Genolini, C., Alacoque, X., Sentenac, M., & Arnaud, C. (2015). kml and kml3d: R packages to cluster longitudinal data. *Journal of Statistical Software, 65*(4), 1-34.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., ... & Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature neuroscience*, *2*(10), 861-863.

Gladyshev, V. N. (2016). Aging: progressive decline in fitness due to the rising deleteriome adjusted by genetic, environmental, and stochastic processes. *Aging cell*, *15*(4), 594-602.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., ... & Rapoport, J. L. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. Proceedings of the National Academy of Sciences, 101(21), 8174-8179.

Goldstein, B. I., Kemp, D. E., Soczynska, J. K., & McIntyre, R. S. (2009). Inflammation and the phenomenology, pathophysiology, comorbidity, and treatment of bipolar disorder: a systematic review of the literature. *Journal of Clinical Psychiatry*, *70*(8), 1078.

Goldstein, J. M., Jerram, M., Poldrack, R., Anagnoson, R., Breiter, H. C., Makris, N., ... & Seidman, L. J. (2005). Sex differences in prefrontal cortical brain activity during fMRI of auditory verbal working memory. *Neuropsychology, 19*(4), 509.

Gordon, J. A., & Stryker, M. P. (1996). Experience-dependent plasticity of binocular responses in the primary visual cortex of the mouse. *Journal of Neuroscience, 16*(10), 3274-3286.

Grande, I., Kapczinski, F., Camino, S., Vázquez, G., & Vieta, E. (2019). Allostasis and neuroprogression in major psychiatric disorders. *Neuroprogression in Psychiatry*, 2.

Griffiths, P. E., & Tabery, J. (2013). Developmental systems theory: What does it explain, and how does it explain it? *Advances in child development and behavior* (Vol. 44, pp. 65-94). JAI.

Grillet, N., Dubreuil, V., Dufour, H. D., & Brunet, J. F. (2003). Dynamic expression of RGS4 in the developing nervous system and regulation by the neural type-specific transcription factor Phox2b. *Journal of Neuroscience, 23*(33), 10613-10621.

Gurung, P., Lukens, J. R., & Kanneganti, T. D. (2015). Mitochondria: diversity in the regulation of the NLRP3 inflammasome. *Trends in molecular medicine*, *21*(3), 193-201.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol, 6*(7), e159.

Hagmeyer, S., Romão, M. A., Cristóvão, J. S., Vilella, A., Zoli, M., Gomes, C. M., & Grabrucker, A. M. (2019). Distribution and relative abundance of S100 proteins in the brain of the APP23 Alzheimer's disease model mice. *Frontiers in neuroscience, 13*, 640.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of intelligent information systems, 17*(2-3), 107-145.

Halliwell, B. (1992). Reactive oxygen species and the central nervous system. *Journal of neurochemistry*, *59*(5), 1609-1623.

Handl, J., Knowles, J., & Kell, D. B. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics, 21*(15), 3201-3212.

Hanisch, U. K., & Kettenmann, H. (2007). Microglia: active sensor and versatile effector cells in the normal and pathologic brain. *Nature neuroscience, 10*(11), 1387-1394.

Harauz, G., & Boggs, J. M. (2013). Myelin management by the 18.5-kDa and 21.5-kDa classic myelin basic protein isoforms. *Journal of neurochemistry, 125*(3), 334-361.

Harauz, G., Ladizhansky, V., & Boggs, J. M. (2009). Structural polymorphism and multifunctionality of myelin basic protein. *Biochemistry, 48*(34), 8094-8104.

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., ... & Abajian, C. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature, 489*(7416), 391-399.

He, H. Y., Hodos, W., & Quinlan, E. M. (2006). Visual deprivation reactivates rapid ocular dominance plasticity in adult visual cortex. *Journal of Neuroscience, 26*(11), 2951-2955.\

Hennig, C. (2010). fpc: Flexible procedures for clustering. *R package version, 2*(2), 0-3.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience, 6*(11), 877-888.

Hensch, T. K., Fagiolini, M., Mataga, N., Stryker, M. P., Baekkeskov, S., & Kash, S. F. (1998). Local GABA circuit control of experience-dependent plasticity in developing visual cortex. *Science, 282*(5393), 1504-1508.

Hickman, S. E., Kingery, N. D., Ohsumi, T. K., Borowsky, M. L., Wang, L. C., Means, T. K., & El Khoury, J. (2013). The microglial sensome revealed by direct RNA sequencing. *Nature neuroscience, 16*(12), 1896-1905.

Holguera, I., & Desplan, C. (2018). Neuronal specification in space and time. *Science, 362*(6411), 176-180.

Holmes, J. M., Lazar, E. L., Melia, B. M., Astle, W. F., Dagi, L. R., Donahue, S. P., ... & Weise, K. K. (2011). Effect of age on response to amblyopia treatment in children. *Archives of ophthalmology, 129*(11), 1451-1457.

Hope, S., Melle, I., Aukrust, P., Steen, N. E., Birkenaes, A. B., Lorentzen, S., ... & Andreassen, O. A. (2009). Similar immune profile in bipolar disorder and schizophrenia: selective

increase in soluble tumor necrosis factor receptor I and von Willebrand factor. *Bipolar disorders, 11*(7), 726-734.

Hu, H., & Jonas, P. (2014). A supercritical density of Na+ channels ensures fast signaling in GABAergic interneuron axons. *Nature neuroscience, 17*(5), 686-693.

Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development, 136*(23), 3853-3862.

Huang, S., & Kauffman, S. A. (2009). Complex gene regulatory networks-from structure to biological observables: cell fate determination.

Huang, Z. J., Kirkwood, A., Pizzorusso, T., Porciatti, V., Morales, B., Bear, M. F., ... & Tonegawa, S. (1999). BDNF regulates the maturation of inhibition and the critical period of plasticity in mouse visual cortex. *Cell, 98*(6), 739-755.

Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *The Journal of physiology, 206*(2), 419-436.

Hubel, D. H., Wiesel, T. N., & LeVay, S. (1976, January). Functional architecture of area 17 in normal and monocularly deprived macaque monkeys. *Cold Spring Harbor Symposia on Quantitative Biolog*y (Vol. 40, pp. 581-589). Cold Spring Harbor Laboratory Press.

Hussain, S., & Davanger, S. (2015). Postsynaptic VAMP/synaptobrevin facilitates differential vesicle trafficking of GluA1 and GluA2 AMPA receptor subunits. *PloS one, 10*(10), e0140868.

Huttenlocher, P. R., de Courten, C., Garey, L. J., & Van der Loos, H. (1982). Synaptogenesis in human visual cortex—evidence for synapse elimination during normal development. *Neuroscience letters*, *33*(3), 247-252.

Ikegami, K., Saigoh, K., Fujioka, A., Nagano, M., Kitajima, K., Sato, C., ... & Shigeyoshi, Y. (2019). Effect of expression alteration in flanking genes on phenotypes of St8sia2-deficient mice. *Scientific reports, 9*(1), 1-11.

Ingi, T., & Aoki, Y. (2002). Expression of RGS2, RGS4 and RGS7 in the developing postnatal brain. *European Journal of Neuroscience, 15*(5), 929-936.

Jakovcevski, I., Filipovic, R., Mo, Z., Rakic, S., & Zecevic, N. (2009). Oligodendrocyte development and the onset of myelination in the human fetal brain. *Frontiers in neuroanatomy, 3,* 5.

Jernigan, T. L., & Tallal, P. (1990). Late childhood changes in brain morphology observable with MRI. *Developmental Medicine & Child Neurology*, *32*(5), 379-385.

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118-127.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), 20150202.

Kaern, Mads, Timothy C. Elston, William J. Blake, and James J. Collins. "Stochasticity in gene expression: from theories to phenotypes." *Nature Reviews Genetics 6, no. 6* (2005): 451-464.

Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., ... & Guennel, T. (2011). Spatio-temporal transcriptome of the human brain. *Nature, 478*(7370), 483-489.

Kapczinski, F., Dias, V. V., Kauer-Sant'Anna, M., Frey, B. N., Grassi-Oliveira, R., Colom, F., & Berk, M. (2009). Clinical implications of a staging model for bipolar disorders. *Expert review of neurotherapeutics*, *9*(7), 957-966.

Kapczinski, F., Magalhães, P. V. S., Balanzá-Martinez, V., Dias, V. V., Frangou, S., Gama, C. S., ... & Kunz, M. (2014). Staging systems in bipolar disorder: an International Society for Bipolar Disorders T ask Force Report. *Acta psychiatrica scandinavica*, *130*(5), 354-363.

Kapczinski, F., Vieta, E., Andreazza, A. C., Frey, B. N., Gomes, F. A., Tramontina, J., ... & Post, R. M. (2008). Allostatic load in bipolar disorder: implications for pathophysiology and treatment. *Neuroscience & Biobehavioral Reviews*, *32*(4), 675-692.

Kapczinski, F., Dal-Pizzol, F., Teixeira, A. L., Magalhaes, P. V., Kauer-Sant'Anna, M., Klamt, F., ... & Gama, C. S. (2011). Peripheral biomarkers and illness activity in bipolar disorder. *Journal of psychiatric research*, *45*(2), 156-161.

Katz, L. C., & Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. *Science, 274*(5290), 1133-1138.

Kauer-Sant'Anna, M., Kapczinski, F., Andreazza, A. C., Bond, D. J., Lam, R. W., Young, L. T., & Yatham, L. N. (2009). Brain-derived neurotrophic factor and inflammatory markers in patients with early-vs. late-stage bipolar disorder. *International Journal of Neuropsychopharmacology*, *12*(4), 447-458.

Kauffman, S. A. (1993). The origins of order: Self-organization and selection in evolution. Oxford University Press, USA.

Kaufmann, B. B., & van Oudenaarden, A. (2007). Stochastic gene expression: from single molecules to the proteome. *Current opinion in genetics & development, 17*(2), 107-112.

Keil, J. M., Qalieh, A., & Kwan, K. Y. (2018). Brain transcriptome databases: a user's guide. *Journal of Neuroscience, 38*(10), 2399-2412.

Kelley, K. W., Nakao-Inoue, H., Molofsky, A. V., & Oldham, M. C. (2018). Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes. *Nature neuroscience, 21*(9), 1171-1184.

Kim, C. S., & Johnston, D. (2018). A possible link between HCN channels and depression. *Chronic Stress*, *2*, 2470547018787781.

Kishimoto, J., Tsuchiya, T., Cox, H., Emson, P. C., & Nakayama, Y. (1998). Age-related changes of calbindin-D28k, calretinin, and parvalbumin mRNAs in the hamster brain. *Neurobiology of aging, 19*(1), 77-82.

Knijff, E. M., Nadine Breunis, M., Kupka, R. W., De Wit, H. J., Ruwhof, C., Akkerhuis, G. W., ... & Drexhage, H. A. (2007). An imbalance in the production of IL-1β and IL-6 by monocytes of bipolar patients: restoration by lithium treatment. *Bipolar disorders, 9*(7), 743-753.

Koellhoffer, E. C., McCullough, L. D., & Ritzel, R. M. (2017). Old maids: aging and its impact on microglia function. International journal of molecular sciences, 18(4), 769.

Kondo, Y., Salibian-Barrera, M., & Zamar, R. (2016). RSKC: An R package for a robust and sparse k-means clustering algorithm. *Journal of Statistical Software, 72*(5), 1-26.

Kooijmans, R. N., Self, M. W., Wouterlood, F. G., Beliën, J. A., & Roelfsema, P. R. (2014). Inhibitory interneuron classes express complementary AMPA-receptor patterns in macaque primary visual cortex. *Journal of Neuroscience, 34(*18), 6303-6315.

Kook, S. Y., Jeong, H., Kang, M. J., Park, R., Shin, H. J., Han, S. H., ... & Yi, E. C. (2014). Crucial role of calbindin-D 28k in the pathogenesis of Alzheimer's disease mouse model. *Cell Death & Differentiation, 21*(10), 1575-1587.

Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M. P., ... & Hussain, N. K. (2019). SynGO: an evidence-based, expert-curated knowledge base for the synapse. *Neuron*, *103*(2), 217-234.

Kotz, C. M., Weldon, D., Billington, C. J., & Levine, A. S. (2004). Age-related changes in brain proDynorphin gene expression in the rat. *Neurobiology of aging, 25*(10), 1343-1347.

Kraepelin, E. (1921). Manic depressive insanity and paranoia. *The Journal of Nervous and Mental Disease*, *53*(4), 350.

Kroon, T., Sierksma, M., & Meredith, R. M. (2013). Investigating mechanisms underlying neurodevelopmental phenotypes of autistic and intellectual disability disorders: a perspective. *Frontiers in systems neuroscience, 7*, 75.

Kwan, K. Y., Lam, M. M., Johnson, M. B., Dube, U., Shim, S., Rašin, M. R., ... & Chan, D. W. (2012). Species-dependent posttranscriptional regulation of NOS1 by FMRP in the developing cerebral cortex. *Cell, 149*(4), 899-911.

Lafon-Cazal, M., Culcasi, M., Gaven, F., Pietri, S., & Bockaert, J. (1993). Nitric oxide, superoxide and peroxynitrite: putative mediators of NMDA-induced cell death in cerebellar granule cells. *Neuropharmacology*, *32*(11), 1259-1266.

Lazar, C., Meganck, S., Taminau, J., Steenhoff, D., Coletta, A., Molter, C., ... & Nowé, A. (2013). Batch effect removal methods for microarray gene expression data integration: a survey. *Briefings in bioinformatics*, *14*(4), 469-490.

Leek, J. T., Johnson, W. E., Parker, H. S., Fertig, E. J., Jaffe, A. E., Storey, J. D., ... & Torres, L. C. (2014). Package 'sva'.

Lennarz, W. J., & Lane, M. D. (2013). *Encyclopedia of biological chemistry.* Academic Press.

Levin, S. G., & Godukhin, O. V. (2017). Modulating effect of cytokines on mechanisms of synaptic plasticity in the brain. *Biochemistry (Moscow), 82*(3), 264-274.

Lewis, M. D., Lamey, A. V., & Douglas, L. (1999). A new dynamic systems method for the analysis of early socioemotional development. *Developmental Science, 2*(4), 457-475.

Liu, H. C., Yang, Y. Y., Chou, Y. M., Chen, K. P., Shen, W. W., & Leu, S. J. (2004). Immunologic variables in acute mania of bipolar disorder. *Journal of neuroimmunology, 150*(1-2), 116-122.

Lloyd, A. F., & Miron, V. E. (2019). The pro-remyelination properties of microglia in the central nervous system. *Nature Reviews Neurology, 15*(8), 447-458.

Lombaert, N., Hennes, M., Gilissen, S., Schevenels, G., Aerts, L., Vanlaer, R., ... & Arckens, L. (2018). 5-HTR 2A and 5-HTR 3A but not 5-HTR 1A antagonism impairs the cross-modal reactivation of deprived visual cortex in adulthood. *Molecular brain, 11*(1), 1-19.

Lyckman, A. W., Horng, S., Leamey, C. A., Tropea, D., Watakabe, A., Van Wart, A., ... & Sur, M. (2008). Gene expression patterns in visual cortex during the critical period: synaptic stabilization and reversal by visual deprivation. *Proceedings of the National Academy of Sciences*, *105*(27), 9409-9414.

Machado-Vieira, R., Andreazza, A. C., Viale, C. I., Zanatto, V., Cereser Jr, V., da Silva Vargas, R., ... & Gentil, V. (2007). Oxidative stress parameters in unmedicated and treated bipolar subjects during initial manic episode: a possible role for lithium antioxidant effects. *Neuroscience letters*, *421*(1), 33-36.

MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. *In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).

Magnin, E., Francavilla, R., Amalyan, S., Gervais, E., David, L. S., Luo, X., & Topolnik, L. (2019). Input-specific synaptic location and function of the α5 GABAA receptor subunit in the mouse CA1 hippocampal neurons. *Journal of Neuroscience, 39*(5), 788-801.

Mahfouz, A., Huisman, S. M., Lelieveldt, B. P., & Reinders, M. J. (2017). Brain transcriptome atlases: a computational perspective. *Brain Structure and Function, 222*(4), 1557-1580.

Malenka, R. C., Kauer, J. A., Perkel, D. J., & Nicoll, R. A. (1989). The impact of postsynaptic calcium on synaptic transmission—its role in long-term potentiation. *Trends in neurosciences, 12*(11), 444-450.

Martin, L. J., Oh, G. H., & Orser, B. A. (2009). Etomidate targets α5γ-aminobutyric acid subtype A receptors to regulate synaptic plasticity and memory blockade. *Anesthesiology: The Journal of the American Society of Anesthesiologists, 111*(5), 1025-1035.

Martin, L. J., Zurek, A. A., MacDonald, J. F., Roder, J. C., Jackson, M. F., & Orser, B. A. (2010). α5GABAA receptor activity sets the threshold for long-term potentiation and constrains hippocampus-dependent memory. *Journal of Neuroscience, 30*(15), 5269-5282.

Marton, N., Baricza, E., Érsek, B., Buzás, E. I., & Nagy, G. (2015). The emerging and diverse roles of src-like adaptor proteins in health and disease. *Mediators of inflammation*.

Maślińska, D., Laure-Kamionowska, M., & Kaliszek, A. (1998). Morphological forms and localization of microglial cells in the developing human cerebellum. *Folia neuropathologica, 36*(3), 145.

Mataga, N., Mizuguchi, Y., & Hensch, T. K. (2004). Experience-dependent pruning of dendritic spines in visual cortex by tissue plasminogen activator. *Neuron, 44*(6), 1031-1041.

Maze, I., Wenderski, W., Noh, K. M., Bagot, R. C., Tzavaras, N., Purushothaman, I., ... & Tamminga, C. A. (2015). Critical role of histone turnover in neuronal transcription and plasticity. *Neuron*, *87*(1), 77-94.

McGee, A. W., Yang, Y., Fischer, Q. S., Daw, N. W., & Strittmatter, S. M. (2005). Experience-driven plasticity of visual cortex limited by myelin and Nogo receptor. *Science, 309*(5744), 2222-2226.

Meilă, M. (2007). Comparing clusterings—an information-based distance. *Journal of multivariate analysis, 98*(5), 873-895.

Ménard, C., Tse, Y. C., Cavanagh, C., Chabot, J. G., Herzog, H., Schwarzer, C., ... & Quirion, R. (2013). Knockdown of prodynorphin gene prevents cognitive decline, reduces anxiety, and rescues loss of group 1 metabotropic glutamate receptor function in aging. *Journal of Neuroscience*, *33*(31), 12792-12804.

Meredith, R. M., Dawitz, J., & Kramvis, I. (2012). Sensitive time-windows for susceptibility in neurodevelopmental disorders. *Trends in neurosciences, 35*(6), 335-344.

Miller, D. J., Duka, T., Stimpson, C. D., Schapiro, S. J., Baze, W. B., McArthur, M. J., ... & Lipovich, L. (2012). Prolonged myelination in human neocortical evolution. *Proceedings of the National Academy of Sciences, 109*(41), 16480-16485.

Modabbernia, A., Taslimi, S., Brietzke, E., & Ashrafi, M. (2013). Cytokine alterations in bipolar disorder: a meta-analysis of 30 studies. *Biological psychiatry*, *74*(1), 15-25.

Morales, B., Choi, S. Y., & Kirkwood, A. (2002). Dark rearing alters the development of GABAergic transmission in visual cortex. *Journal of Neuroscience, 22*(18), 8084-8090.

Mota, R., Gazal, M., Acosta, B. A., de Leon, P. B., Jansen, K., Pinheiro, R. T., ... & Lara, D. R. (2013). Interleukin-1β is associated with depressive episode in major depression but not in bipolar disorder. *Journal of psychiatric research, 47(*12), 2011-2014.

Müller, N., & Schwarz, M. J. (2007). The immune-mediated alteration of serotonin and glutamate: towards an integrated view of depression. *Molecular psychiatry*, *12*(11), 988-1000.

Muneer, A. (2016). The neurobiology of bipolar disorder: an integrated approach. *Chonnam medical journal*, *52*(1), 18-37.

Murphy, E. V., Zhang, Y., Zhu, W., & Biggs, J. (1995). The human glioma pathogenesis-related protein is structurally related to plant pathogenesis-related proteins and its gene is expressed specifically in brain tumors. *Gene, 159*(1), 131-135.

Murphy, K. M., Tcharnaia, L., Beshara, S., & Jones, D. G. (2012). Cortical development of AMPA receptor trafficking proteins. *Frontiers in Molecular Neuroscience, 5*, 65.

Nakagami, Y., Watakabe, A., & Yamamori, T. (2013). Monocular inhibition reveals temporal and spatial changes in gene expression in the primary visual cortex of marmoset. *Frontiers in neural circuits, 7*, 43.

Nardone, S., Sams, D. S., Reuveni, E., Getselter, D., Oron, O., Karpuj, M., & Elliott, E. (2014). DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Translational psychiatry*, *4*(9), e433-e433.

Newell, K. M., Liu, Y. T., & Mayer-Kress, G. (2003). A dynamical systems interpretation of epigenetic landscapes for infant motor development. *Infant Behavior and Development, 26*(4), 449-472.

Ni, L., Xu, Y., Dong, S., Kong, Y., Wang, H., Lu, G., ... & Sun, H. (2020). The potential role of the HCN1 ion channel and BDNF-mTOR signaling pathways and synaptic transmission in the alleviation of PTSD. *Translational Psychiatry*, *10*(1), 1-15.

Oliveira-Pinto, A. V., Santos, R. M., Coutinho, R. A., Oliveira, L. M., Santos, G. B., Alho, A. T., ... & Pasqualucci, C. A. (2014). Sexual dimorphism in the human olfactory bulb: females have more neurons and glial cells than males. *PloS one*, *9*(11), e111733.

Oray, S., Majewska, A., & Sur, M. (2004). Dendritic spine dynamics are regulated by monocular deprivation and extracellular matrix degradation. *Neuron, 44*(6), 1021-1030.

Ortiz-Domínguez, A., Hernández, M. E., Berlanga, C., Gutiérrez-Mora, D., Moreno, J., Heinze, G., & Pavón, L. (2007). Immune variations in bipolar disorder: phasic differences. *Bipolar disorders*, *9*(6), 596-602.

Papadimitriou, G. N., Calabrese, J. R., Dikeos, D. G., & Christodoulou, G. N. (2005). Rapid cycling bipolar disorder: biology and pathogenesis. *International Journal of Neuropsychopharmacology*, *8*(2), 281-292.

Pape, H. C. (1996). Queer current and pacemaker: the hyperpolarization-activated cation current in neurons. *Annual review of physiology, 58*(1), 299-327.

Park, H. J., Kong, J. H., Kang, Y. S., Park, W. M., Jeong, S. A., Park, S. M., ... & Jeon, C. J. (2002). The distribution and morphology of calbindin D28K-and calretinin-immunoreactive neurons in the visual cortex of mouse. *Molecules and cells, 14*(1), 143-149.

Petralia, R. S., Mattson, M. P., & Yao, P. J. (2014). Communication breakdown: the impact of ageing on synapse structure. Ageing research reviews, 14, 31-42.

Pfaffenseller, B., da Silva Magalhães, P. V., De Bastiani, M. A., Castro, M. A. A., Gallitano, A. L., Kapczinski, F., & Klamt, F. (2016). Differential expression of transcriptional regulatory units in the prefrontal cortex of patients with bipolar disorder: potential role of early growth response gene 3. *Translational psychiatry*, *6*(5), e805-e805.

Pfeffer, C. K., Xue, M., He, M., Huang, Z. J., & Scanziani, M. (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience, 16*(8), 1068-1076.

Piaget, J., & Cook, M. (1952). The origins of intelligence in children (Vol. 8, No. 5, p. 18). New York: International Universities Press.

Piaget, J., & Inhelder, B. (1969). The psychology of the child. New York, NY: Basic Book. Inc., Harper.

Pinto, J. G., Hornby, K. R., Jones, D. G., & Murphy, K. M. (2010). Developmental changes in GABAergic mechanisms in human visual cortex across the lifespan. *Frontiers in cellular neuroscience, 4*, 16.


Pinto, J. G., Hornby, K. R., Jones, D. G., & Murphy, K. M. (2010). Developmental changes in GABAergic mechanisms in human visual cortex across the lifespan. *Frontiers in cellular neuroscience, 4*, 16.

Pizzorusso, T., Medini, P., Berardi, N., Chierzi, S., Fawcett, J. W., & Maffei, L. (2002). Reactivation of ocular dominance plasticity in the adult visual cortex. *Science, 298*(5596), 1248-1251.

Pribiag, H., & Stellwagen, D. (2014). Neuroimmune regulation of homeostatic synaptic plasticity. *Neuropharmacology, 78*, 13-22.

Prigogine, I., & Nicolis, G. (1971). Biological order, structure and instabilities.

Qiu, W., & Joe, H. (2006). Generation of random clusters with specified degree of separation. *Journal of Classification, 23*(2), 315-334.

Raz, N., & Rodrigue, K. M. (2006). Differential aging of the brain: patterns, cognitive correlates and modifiers. *Neuroscience & Biobehavioral Reviews*, *30*(6), 730-748.

Reiner, O. (2013). LIS1 and DCX: implications for brain development and human disease in relation to microtubules. *Scientifica*, *2013*.

Rezaie, P., & Male, D. (1999). Colonisation of the developing human brain and spinal cord by microglia: a review. Microscopy research and technique, 45(6), 359-382.

Richter, S., Gorny, X., Marco-Pallares, J., Krämer, U. M., Machts, J., Barman, A., ... & Reissner, C. (2011). A potential role for a genetic variation of AKAP5 in human aggression and anger control. *Frontiers in human neuroscience*, *5*, 175.

Ridout, K. K., Ridout, S. J., Price, L. H., Sen, S., & Tyrka, A. R. (2016). Depression and telomere length: A meta-analysis. *Journal of affective disorders*, *191*, 237-247.

Rilett, K. C., Friedel, M., Ellegood, J., MacKenzie, R. N., Lerch, J. P., & Foster, J. A. (2015). Loss of T cells influences sex differences in behavior and brain structure. *Brain, behavior, and immunity*, *46*, 249-260.

Risher, W. C., Patel, S., Kim, I. H., Uezu, A., Bhagat, S., Wilton, D. K., ... & Stevens, B. (2014). Astrocytes refine cortical connectivity at dendritic spines. *Elife, 3,* e04047.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research, 43*(7), e47.

Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., Reus, L. M., Alloza, C., ... & Liewald, D. C. (2018). Sex differences in the adult human brain: evidence from 5216 UK Biobank participants. *Cerebral Cortex*, *28*(8), 2959-2975.

Robertson, H. R., Gibson, E. S., Benke, T. A., & Dell'Acqua, M. L. (2009). Regulation of postsynaptic structure and function by an A-kinase anchoring protein–membrane-

associated guanylate kinase scaffolding complex. *Journal of Neuroscience*, *29*(24), 7929-7943.

Rosenblat, J. D., & McIntyre, R. S. (2015). Are medical comorbid conditions of bipolar disorder due to immune dysfunction?. *Acta Psychiatrica Scandinavica*, *132*(3), 180-191.

Rothermundt, M., Peters, M., Prehn, J. H., & Arolt, V. (2003). S100B in brain damage and neurodegeneration. *Microscopy research and technique, 60*(6), 614-632.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics, 20*, 53-65.

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC bioinformatics, 18(*1), 529.

Ryan, M. M., Lockstone, H. E., Huffaker, S. J., Wayland, M. T., Webster, M. J., & Bahn, S. (2006). Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Molecular psychiatry*, *11*(10), 965-978.

Santoro, B., & Baram, T. Z. (2003). The multiple personalities of h-channels. *Trends in neurosciences, 26*(10), 550-554.

Sassi, R. B., Brambilla, P., Nicoletti, M., Mallinger, A. G., Frank, E., Kupfer, D. J., ... & Soares, J. C. (2003). White matter hyperintensities in bipolar and unipolar patients with relatively mild-to-moderate illness severity. *Journal of affective disorders*, *77*(3), 237-245.

Sawtell, N. B., Frenkel, M. Y., Philpot, B. D., Nakazawa, K., Tonegawa, S., & Bear, M. F. (2003). NMDA receptor-dependent ocular dominance plasticity in adult visual cortex. *Neuron, 38*(6), 977-985.

Scala, F., Kobak, D., Shan, S., Bernaerts, Y., Laturnus, S., Cadwell, C. R., ... & Papadopoulos, S. (2019). Layer 4 of mouse neocortex differs in cell types and circuit organization between sensory areas. *Nature communications, 10*(1), 1-12.

Schafer, D. P., & Stevens, B. (2015). Microglia function in central nervous system development and plasticity. *Cold Spring Harbor perspectives in biology, 7*(10), a020545.

Schafer, D. P., Lehrman, E. K., Kautzman, A. G., Koyama, R., Mardinly, A. R., Yamasaki, R., ... & Stevens, B. (2012). Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron, 74*(4), 691-705.

Schuepbach, D., Novick, D., Haro, J. M., Reed, C., Boeker, H., Noda, S., ... & Hell, D. (2008). Determinants of voluntary vs. involuntary admission in bipolar disorder and the impact of adherence. *Pharmacopsychiatry*, *41*(01), 29-36.

Seo, H., Seol, M. J., & Lee, K. (2015). Differential expression of hyperpolarization-activated cyclic nucleotide-gated channel subunits during hippocampal development in the mouse. *Molecular brain, 8*(1), 13.

Shatz, C. J., & Stryker, M. P. (1978). Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *The Journal of physiology, 281*(1), 267-283.

Shi, R. M., Kobayashi, T., Kikuchi, A., Sato, R., Uematsu, M., An, K., & Kure, S. (2017). Phenytoin-responsive epileptic encephalopathy with a tandem duplication involving FGF12. *Neurology Genetics, 3*(1).

Shim, S., Kwan, K. Y., Li, M., Lefebvre, V., & Šestan, N. (2012). Cis-regulatory control of corticospinal system development and evolution. *Nature, 486*(7401), 74-79.

Shioda, N., Moriguchi, S., Oya, T., Ishii, Y., Shen, J., Matsushima, T., ... & Fukunaga, K. (2012). Aberrant hippocampal spine morphology and impaired memory formation in neuronal platelet-derived growth factor beta-receptor lacking mice. *Hippocampus, 22*(6), 1371-1378.

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., ... & Alon, U. (2006). Variability and memory of protein levels in human cells. *Nature, 444*(7119), 643-646.

Şimşek, Ş., Gençoğlan, S., Yüksel, T., Kaplan, İ., Alaca, R., & Aktaş, H. (2016). Oxidative stress and DNA damage in untreated first-episode psychosis in adolescents. *Neuropsychobiology*, *73*(2), 92-97.

Siu, C. R., & Murphy, K. M. (2018). The development of human visual cortex and clinical implications. *Eye and brain, 10*, 25.

Siu, C. R., Balsor, J. L., Jones, D. G., & Murphy, K. M. (2015). Classic and Golli Myelin Basic Protein have distinct developmental trajectories in human visual cortex. *Frontiers in neuroscience, 9,* 138.

Siu, C. R., Beshara, S. P., Jones, D. G., & Murphy, K. M. (2017). Development of glutamatergic proteins in human visual cortex across the lifespan. Journal of *Neuroscience, 37*(25), 6031-6042.

Slack, J. M. (2002). Conrad Hal Waddington: the last renaissance biologist?. *Nature Reviews Genetics, 3*(11), 889-895.

Söderlund, J., Olsson, S. K., Samuelsson, M., Walther-Jallow, L., Johansson, C., Erhardt, S., ... & Engberg, G. (2011). Elevation of cerebrospinal fluid interleukin-1β in bipolar disorder. *Journal of psychiatry & neuroscience: JPN*, *36*(2), 114.

Sommeijer, J. P., & Levelt, C. N. (2012). Synaptotagmin-2 is a reliable marker for parvalbumin positive inhibitory boutons in the mouse visual cortex. *PloS one, 7*(4), e35323.

Song, Y. H., Hwang, Y. S., Kim, K., Lee, H. R., Kim, J. H., Maclachlan, C., ... & Lee, S. H. (2020). Somatostatin enhances visual processing and perception by suppressing excitatory inputs to parvalbumin-positive interneurons in V1. *Science Advances, 6*(17), eaaz0517.

Sørensen, J. B., Wiederhold, K., Müller, E. M., Milosevic, I., Nagy, G., de Groot, B. L., ... & Fasshauer, D. (2006). Sequential N-to C-terminal SNARE complex assembly drives priming and fusion of secretory vesicles. *The EMBO journal, 25*(5), 955-966.

Sorokowski, P., Karwowski, M., Misiak, M., Marczak, M. K., Dziekan, M., Hummel, T., & Sorokowska, A. (2019). Sex differences in human olfaction: A meta-analysis. *Frontiers in psychology,* 10, 242.

Sowell, E. R., Thompson, P. M., Tessner, K. D., & Toga, A. W. (2001). Mapping continued brain growth and gray matter density reduction in dorsal frontal cortex: Inverse relationships during post adolescent brain maturation. *Journal of Neuroscience*, *21*(22), 8819-8829.

Sparkman, N. L., & Johnson, R. W. (2008). Neuroinflammation associated with aging sensitizes the brain to the effects of infection or stress. *Neuroimmunomodulation, 15*(4-6), 323-330.

Spencer, J. P., Austin, A., & Schutte, A. R. (2012). Contributions of dynamic systems theory to cognitive development. *Cognitive Development, 27*(4), 401-418.

Spencer, J. P., Perone, S., & Buss, A. T. (2011). Twenty years and going strong: A dynamic systems revolution in motor and cognitive development. *Child Development Perspectives, 5*(4), 260-266.

Spencer, J. P., Simmering, V. R., Schutte, A. R., & Schöner, G. (2007). What does theoretical neuroscience have to offer the study of behavioral development? Insights from a dynamic field theory of spatial cognition.

Stevens, J. S., & Hamann, S. (2012). Sex differences in brain activation to emotional stimuli: a meta-analysis of neuroimaging studies. *Neuropsychologia*, *50*(7), 1578-1593.

Streit, W. J. (2006). Microglial senescence: does the brain's immune system have an expiration date?. *Trends in neurosciences*, *29*(9), 506-510.

Streit, W. J., Sammons, N. W., Kuhns, A. J., & Sparks, D. L. (2004). Dystrophic microglia in the aging human brain. Glia, 45(2), 208-212.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545-15550.

Svitkina, T., Lin, W. H., Webb, D. J., Yasuda, R., Wayman, G. A., Van Aelst, L., & Soderling, S. H. (2010). Regulation of the postsynaptic cytoskeleton: roles in development, plasticity, and disorders. *Journal of Neuroscience, 30*(45), 14937-14942.

Taminau, J., Meganck, S., Lazar, C., Steenhoff, D., Coletta, A., Molter, C., ... & Nowé, A. (2012). Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. *BMC bioinformatics*, *13*(1), 1-9.

Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., ... & Bertagnolli, D. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature neuroscience, 19*(2), 335-346.

Thelen, E., & Smith, L. B. (1994). A dynamic systems approach to the development of perception and action.

Thelen, E., & Smith, L. B. (2007). Dynamic systems theories. Handbook of child psychology, 1.

Thorndike, R. L. (1953). Who belongs in the family?. *Psychometrika, 18*(4), 267-276.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63*(2), 411-423.

Tremblay, M. È., & Majewska, A. K. (2011). A role for microglia in synaptic plasticity?. *Communicative & integrative biology, 4*(2), 220-222.

Tremblay, M. È., Lowery, R. L., & Majewska, A. K. (2010). Microglial interactions with synapses are modulated by visual experience. *PLoS Biol, 8*(11), e1000527.

Trimble, William S., Thackery S. Gray, Lisa A. Elferink, Michael C. Wilson, and Richard H. Scheller. "Distinct patterns of expression of two VAMP genes within the rat brain." *Journal of Neuroscience 10, no. 4* (1990): 1380-1387.

Van Dijk, M., & Van Geert, P. (2007). Wobbles, humps, and sudden jumps: A case study of continuity, discontinuity, and variability in early language development. *Infant and Child Development: An International Journal of Research and Practice, 16*(1), 7-33.

Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological review, 98*(1), 3.

Van Geert, P. (2003). Dynamic systems approaches and modeling of developmental processes. *Handbook of developmental psychology*, 640672.

Van Geert, P. (Ed.). (1986). Theory building in developmental psychology. Newnes.

Van Geert, P. L. (2019). Dynamic systems, process, and development. *Human Development, 63*(3-4), 153-179.

Vincent, A. J., Lau, P. W., & Roskams, A. J. (2008). SPARC is expressed by macroglia and microglia in the developing and mature nervous system. *Developmental dynamics: an official publication of the American Association of Anatomists, 237*(5), 1449-1462.

Votinov, M., Pripfl, J., Windischberger, C., Moser, E., Sailer, U., & Lamm, C. (2015). A functional polymorphism in the prodynorphin gene affects cognitive flexibility and brain activation during reversal learning. *Frontiers in behavioral neuroscience*, *9*, 172.

Waddington, C. H. (1957). A discussion of some aspects of theoretical biology. *The Strategy of the Genes*. *London George Allen and Unwin*., London.

Wakselman, S., Béchade, C., Roumier, A., Bernard, D., Triller, A., & Bessis, A. (2008). Developmental neuronal death in hippocampus requires the microglial CD11b integrin and DAP12 immunoreceptor. *Journal of Neuroscience, 28*(32), 8138-8143.

Wang, K. C., Koprivica, V., Kim, J. A., Sivasankaran, R., Guo, Y., Neve, R. L., & He, Z. (2002). Oligodendrocyte-myelin glycoprotein is a Nogo receptor ligand that inhibits neurite outgrowth. *Nature, 417*(6892), 941-944.

Wang, A. K., & Miller, B. J. (2018). Meta-analysis of cerebrospinal fluid cytokine and tryptophan catabolite alterations in psychiatric patients: comparisons between schizophrenia, bipolar disorder, and depression. *Schizophrenia Bulletin*, *44*(1), 75-83.

Watakabe, A., Komatsu, Y., Sadakane, O., Shimegi, S., Takahata, T., Higo, N., ... & Sakamoto, H. (2009). Enriched expression of serotonin 1B and 2A receptor genes in macaque visual cortex and their bidirectional modulatory effects on neuronal responses. *Cerebral cortex, 19*(8), 1915-1928.

Watkins, C. C., Sawa, A., & Pomper, M. G. (2014). Glia and immune cell signaling in bipolar disorder: insights from neuropharmacology and molecular imaging to clinical application. *Translational psychiatry*, *4*(1), e350-e350.

Weinberg, E. D. (2008). Iron out-of-balance: a risk factor for acute and chronic diseases. *Hemoglobin, 32*(1-2), 117-122.

Weisenhaus, M., Allen, M. L., Yang, L., Lu, Y., Nichols, C. B., Su, T., ... & McKnight, G. S. (2010). Mutations in AKAP5 disrupt dendritic signaling complexes and lead to electrophysiological and behavioral phenotypes in mice. *PLoS One*, *5*(4), e10325.

Wieck, A., Grassi-Oliveira, R., do Prado, C. H., Rizzo, L. B., de Oliveira, A. S., Kommers-Molina, J., ... & Bauer, M. E. (2013). Differential neuroendocrine and immune responses to acute psychosocial stress in women with type 1 bipolar disorder. *Brain, behavior, and immunity*, *34*, 47-55.

Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of neurophysiology, 26*(6), 1003-1017.

Wildburger, N. C., Ali, S. R., Hsu, W. C. J., Shavkunov, A. S., Nenov, M. N., Lichti, C. F., ... & Nilsson, C. L. (2015). Quantitative proteomics reveals protein–protein interactions with fibroblast growth factor 12 as a component of the voltage-gated sodium channel 1.2 (Nav1. 2) macromolecular complex in mammalian brain. *Molecular & Cellular Proteomics, 14(*5), 1288-1300.

Williamson, J. M., & Lyons, D. A. (2018). Myelin dynamics throughout life: an ever-changing landscape?. *Frontiers in cellular neuroscience*, *12*, 424.

Witten, D. M., & Tibshirani, R. (2010). A framework for feature selection in clustering. *Journal of the American Statistical Association, 105*(490), 713-726.

Yakovleva, T., Marinova, Z., Kuzmin, A., Seidah, N. G., Haroutunian, V., Terenius, L., & Bakalkin, G. (2007). Dysregulation of dynorphins in Alzheimer disease. *Neurobiology of aging, 28*(11), 1700-1708.

Yamaguchi, S., Aoki, N., Kitajima, T., Iikubo, E., Katagiri, S., Matsushima, T., & Homma, K. J. (2012). Thyroid hormone determines the start of the sensitive period of imprinting and primes later learning. *Nature communications, 3*(1), 1-13.

York, E. M., Bernier, L. P., & MacVicar, B. A. (2018). Microglial modulation of neuronal activity in the healthy brain. *Developmental neurobiology, 78*(6), 593-603.

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*, *16*(5), 284-287.

Yuan, A., Rao, M. V., Sasaki, T., Chen, Y., Kumar, A., Liem, R. K., ... & Nixon, R. A. (2006). α-Internexin is structurally and functionally associated with the neurofilament triplet proteins in the mature CNS. *Journal of Neuroscience, 26*(39), 10006-10019.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology, 4*(1).

Zhou, J., Pfaff, D. W., & Chen, G. (2005). Sex differences in estrogenic regulation of neuronal activity in neonatal cultures of ventromedial nucleus of the hypothalamus. *Proceedings of the National Academy of Sciences*, 102(41), 14907-14912.

Zimmermann, J., Trimbuch, T., & Rosenmund, C. (2014). Synaptobrevin 1 mediates vesicle priming and evoked release in a subpopulation of hippocampal neurons. *Journal of neurophysiology, 112*(6), 1559-1565.