

5-2020

An Exploration of Neuroanatomical Signatures of Pediatric Bipolar Disorder Using Structural Neuroimaging

Jonika Tannous

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations



Part of the [Mental Disorders Commons](#)

Recommended Citation

Tannous, Jonika, "An Exploration of Neuroanatomical Signatures of Pediatric Bipolar Disorder Using Structural Neuroimaging" (2020). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 1013.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/1013

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.

**AN EXPLORATION OF NEUROANATOMICAL SIGNATURES OF PEDIATRIC
BIPOLAR DISORDER USING STRUCTURAL NEUROIMAGING**

by

Jonika Tannous, B.A

APPROVED:



Jair C. Soares, M.D, Ph.D.
Advisory Professor



Benson Mwangi, Ph.D.



Khader Hasan, Ph.D.



Scott Lane, Ph.D.



Zhongming Zhao, Ph.D.

APPROVED:

Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

**AN EXPLORATION OF NEUROANATOMICAL SIGNATURES OF PEDIATRIC
BIPOLAR DISORDER USING STRUCTURAL NEUROIMAGING**

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Jonika Tannous, B.A

Houston, Texas

May 2020

Dedication

To my mother-- who saw potential in me and spent hours every day fostering my inherent curiosity. Thank you for the word games, the times tables, and the dictations. You will always be the best teacher I ever had. I'm still not sorry for hiding the Arabic books though.

To my father-- who came home so tired from trying to make a life for us, he was unable to take off his shoes. Yet, he was never too tired to shower me with love and make me feel like I was the reason the sun rose every morning. I hope you find it as cool as I do, that scientists from all over the world have already written "Tannous et al." in their manuscripts.

To my brothers-- who are turning into the kind of men I always hoped they would be.

To my soon-to-be husband-- who for the past seven years has kept me grounded and sane. You've been the best travel, hiking, food, cycling, and snuggle buddy a girl could ask for.

To my friends-- who have no idea how much the pasta dinners, climbing sessions, house parties, café chats, and Facebook Messenger vent sessions helped me get through.

To my mentors-- who built me up to be the scientist I am today. Thank you for taking the time to listen to my ideas, concerns, hopes, and goals. Thank you for inspiring me every day to question everything and take nothing at face value. Thank you for the lab meetings. Thank you for the lunch chats. Thank you for the edits. Thank you for the committee meetings. Thank you for the white board brainstorming sessions that undoubtedly made you have to stay in lab longer than you had to. Thank you for differing to me in meetings even though you knew the answer, just to give me a moment to shine. Thank you for your support.

To my pets-- Ellie and Gunther who gave moral support and countless distractions.

To the little girl -- who wrote in her diary that one day she was going to get a PhD. You did it. Now go do so much more.

**AN EXPLORATION OF NEUROANATOMICAL SIGNATURES OF PEDIATRIC
BIPOLAR DISORDER USING STRUCTURAL NEUROIMAGING**

Jonika Tannous, BA

Advisory Professor: Jair C. Soares, M.D., Ph.D.

Bipolar Disorder (BD) is diagnosed using the Diagnostic and Statistical Manual (DSM) criteria, which relies heavily on symptomatology. This method, however, lends itself to error due to variance in symptom expression and is further complicated during childhood and adolescence—a period marked by major anatomical and behavioral changes. Therefore, in order to institute early and effective interventions, it is imperative that we develop more objective methods of mood disorder characterization and diagnosis. Two proposed solutions for accomplishing this task include 1) the generation of normative development models to assess BD risk and 2) data-driven clustering of patients based on specific neuroanatomical profiles, i.e. biotypes. While the normative development model was unable to quantify BD risk at an individual level, it nevertheless emphasized the heterogeneous nature of both healthy and BD development. The clustering analysis, however, was successful at parsing the variance in the BD sample which resulted in the identification of two distinct BD biotypes. Whereas the BD clusters mapped onto specific anatomical and neurocognitive patterns, symptom-based DSM subtypes were not associated with any empirical measures of mental health. Hence, these findings are testament to the potential of unsupervised and unbiased computational methods in the future of BD diagnostics.

Table of Contents

Approval Page	I
Cover Page	II
Dedication	III
Abstract	IV
Table of Contents	V
List of Images	VII
List of Tables	IX
Chapter 1: Introduction	1
Chapter 2: Processing and Quality Control of MRI and DTI Images	7
Introduction	8
Methods	9
Subjects	9
Imaging	10
Grey Matter Processing	11
White Matter Processing	12
Statistics	13
Results	14
Grey Matter	14
White Matter	15
Discussion	16
Chapter 3: Neuroanatomical and Cognitive Differences Between HCs and BDs	17
Introduction	18
Methods	19
Results	19
Grey Matter	19
White Matter	21
Cognition	21
Discussion	22
Chapter 4: SVM Classification of BDs and HCs	23
Introduction	24
Methods	25
Results	26
Role of PCA Components in SVM Model	26
	V

Model Performance	28
Discussion	29
Chapter 5: Age Prediction	31
Introduction	32
Methods	33
Results	35
Principal Components of Ridge Regression	35
Model Performance	38
Discussion	39
Chapter 6: Clustering of BD	42
Introduction	43
Methods	44
Results	46
Characterizing Clusters	48
Diffusivity Differences Between GM areas of Clusters	49
Cognitive Differences Between Clusters	50
SVM Classification Performance Differences Between BD Clusters	51
Longitudinal Stability of BD Clusters	52
Discussion	52
Chapter 7: Cortical Thickness and Gyrification of BD Clusters	57
Introduction	58
Methods	59
Results	60
Discussion	67
Chapter 8: Discussion	69
References	73
Vita	91

List of Images

Figure 1. Age distribution of sample	10
Figure 2. Example FreeSurfer segmentations	12
Figure 3. Example FA map overlain with principal vectors	13
Figure 4. Example TRACULA Output for a healthy subject.	13
Figure 5. Total subcortical and cortical volume throughout development	14
Figure 6. TRACULA pathway measures with significant age trends.	15
Figure 7. Subcortical areas with significant differences between BDs and HCs	20
Figure 8. Cortical areas with significant differences between BDs and HCs	20
Figure 9. Significant cognitive differences between BDs and HCs	21
Figure 10. Result of knee/elbow method	25
Figure 11. GridSearch results.	26
Figure 12. Coefficients of the seven principal components in the SVM model	27
Figure 13. Loadings of individual brain areas are visualized for each component	28
Figure 14. Insula adjacent cortices accounting for highest loadings in principal component 5	28
Figure 15. Summary of SVM performance	29
Figure 16. Result of knee/elbow method	34
Figure 17. Comparison of GridSearch to iteration results	34
Figure 18. PCA variance explained and heatmap loadings.	36
Figure 19. Coefficients of the principal components in ridge regression model	37
Figure 20. Summary of age prediction model performance for BDs and HCs	38
Figure 21. Distribution of residuals by group	39
Figure 22. K-means clustering results	46

Figure 23. Example subjects from Cluster 1, Cluster 2, and HC groups	46
Figure 24. Principal components of K-means analysis	47
Figure 25. Three-group comparison of principal components over age	48
Figure 26. Age distribution of BD clusters	48
Figure 27. CDRS and YMRS differences between clusters	48
Figure 28. Distribution plot of principal component 1 loadings.	49
Figure 29. Differences in Amygdala MD between Cluster 1, Cluster 2, and HCs	49
Figure 30. Cognitive differences between Cluster 1 and Cluster 2	50
Figure 31. Cognitive differences between Cluster 1, Cluster 2, and HCs	51
Figure 32. Revisiting SVM classification	52
Figure 33. Cluster classification of BD subjects at two time points	52
Figure 34. ICV and total grey matter differences between Cluster 1, Cluster 2, and HCs	58
Figure 35. Average gyrification and cortical thickness in Cluster 1, Cluster 2, and HCs	60
Figure 36. Comparison of the effects of three Cluster 2 outliers on average gyrification results	61
Figure 37. Qdec analyses of gyrification and cortical thickness	62
Figure 38. Comparison of the effects of three Cluster 2 outliers on Qdec gyrification results.	63
Figure 39. Qdec gyrification areas with significant age X group interactions	65
Figure 40. Qdec cortical thickness areas with significant age X group interactions	66

List of Tables

Table 1. Imaging Parameters	11
Table 2. CANTAB tasks administered	19
Table 3. Coefficients and interpretations of the principal components in the SVM model	27
Table 4. Coefficients and interpretations of the principal components in the ridge regression model.	35
Table 5. Gender of BD clusters	47
Table 6. DSM subtypes of BD clusters	47
Table 7. Cohen's Differences in ICV and total grey matter volumes between groups	57
Table 8. Group X Age interactions for gyrification results	63
Table 9. Group X Age interactions for cortical thickness results	65

Chapter 1: Introduction

Bipolar disorder (BD) is a mental illness characterized by periods of mood irregularity that may shift between mania, hypomania, depression, and euthymia (Grande et al. 2016). About 2.4% of people worldwide have been diagnosed with BD, and the U.S. in particular has a high lifetime prevalence at 4.4% (Kessler et al., 2005; Merikangas et al., 2011). BD is one of the leading causes of disability and is associated with extensive neuroanatomical and cognitive deficits (Alonso et al. 2011). BD subjects are often reported to have smaller brain regions in systems that are responsible for executive functioning and attention, and memory consolidation, as well as emotion and fear response. White matter (WM) pathways responsible for connecting and relaying signals between these systems have also been shown to have decreased integrity largely due to demyelination (F. Lin et al. 2011; Bellani et al. 2009; Heng, Song, and Sim 2010). These anatomical abnormalities are then heavily linked to corresponding neurocognitive deficits that impact psychosocial behavior, mood regulation, and impulsivity (Lima, Peckham, and Johnson 2018).

For many BD patients, the first episode emerges during childhood and adolescence (P.-I. Lin et al. 2006). Pediatric BD is characterized by many of the same deficits seen in adult populations (Leibenluft, Charney, and Pine 2003). Moreover, earlier onset of BD often predicts worse outcomes, which include increased rates of drug and alcohol abuse, comorbidity, and suicidality (Leverich et al. 2007). The poor prognosis of pediatric BD is often due to delays in diagnosis and treatment, thus emphasizing the need for early detection (Post et al. 2010).

Identification of BD, as well as other mood disorders, remains difficult since symptom expression often does not reflect the underlying causes of illness and tends to vary from patient to patient (M. Song et al. 2017). This is further complicated during childhood and adolescence, a critical period in mental health development marked by major anatomical and behavioral changes (Schulenberg, Sameroff, and Cicchetti 2004). Thus, years of persistent reliance on

symptomatology in the clinic, as is standard in the Diagnostic and Statistical Manual (DSM), and research based on false assumptions of population homogeneity has led to symptom-based diagnostic labels that do not align with neurobiology and a lack of standard BD biomarkers that can be used in practice (Marquand, Wolfers, et al. 2016; Wolfers et al. 2018). Therefore, in order to institute early and effective interventions, it is imperative that we develop more objective methods of mood disorder characterization and diagnosis. Two proposed solutions for accomplishing this task include 1) the generation of normative development models and 2) data-driven clustering of patients, i.e. biotypes.

Mood disorders are typically regarded as neurodevelopmental in nature; hence, charting brain changes across development could provide a normative model that would assist in early identification of at-risk patients (Vértes and Bullmore 2015). While many studies have used imaging techniques to track brain changes in gray matter (GM) volumes, GM morphology, and functional connectivity (Wang et al. 2012; Cao et al. 2015; Khundrakpam et al. 2015; Lenroot and Giedd 2006; Yeo et al. 2011), little research has been done to develop an imaging model that can be used to screen for irregularities (Marquand, Rezek, et al. 2016). Ideally, clinicians should have access to a tool that uses neuroanatomical measures to assess mental health risk-- a similar role to the body mass index or the blood pressure chart. Specifically, when charting development, normative models can be used to predict a subject's age. Provided the subject is predicted to be substantially younger than their true age, this discrepancy may imply delayed neurodevelopment, which is a signature of pediatric BD (Roybal et al. 2012). GM volumes could provide the basis of such a model, as GM trajectories are reliably consistent within the literature and have already been shown to be able to predict the ages of healthy subjects (Cao et al. 2015).

Beyond assessing mental illness risk, it is crucial to be able to classify and ultimately treat subjects based not only on their symptoms but also their neurophysiological profile. Recently,

advances in computational modeling have allowed for holistic processing of clinical and neuroimaging data, which has prompted the discovery of biotypes that do not adhere to symptom-based DSM labels (Bak et al. 2017). A groundbreaking study by Clementz et al. examined adult BD, schizoaffective, and schizophrenic patients and used both neurocognitive testing and measures of perceptual processing to develop three distinct biotypes (Clementz et al. 2016). All three biotypes contained patients from each of the illness groups. Further examination showed that each biotype had unique cognitive deficits and differed in social functioning, rates of psychosis, and location of gray matter volume reductions. A patient's DSM label, on the other hand, was not associated with a unique clinical phenotype and was only implicated in symptom severity. This study along with others, which have examined disorders such as adult depression and adolescent ADHD, highlight both the inconsistencies of DSM labels and the need for homogeneous, verifiable biotypes (Drysdale et al. 2017; Fair et al. 2012). Indeed, the DSM specifies three BD subtypes, BD-1, BD-2, and BD-NOS, yet biological distinctions between them have yet to be substantiated and it is unclear whether these subtypes align with data-driven biotypes (Martin Tesli et al. 2015; M. Tesli et al. 2014).

This dissertation aims to address the lack of neurodevelopmental decision support tools in pediatric BD and explore how BD biotypes can make sense of diagnostic heterogeneity. This will be accomplished in three main parts:

Part 1 describes the subject sample and introduces the imaging methodologies that are the foundation of all analyses presented in later chapters. In addition, using traditional assumptions of homogeneity, differences between BDs and HCs are examined with both classical parametric methods and machine learning in order to characterize BD and ascertain whether our sample exhibits expected patterns reported in previous studies. Concordance with the BD literature is thus a testament to the validity of the sampling and

lends credence to the potential reproducibility of the novel analyses conducted in the subsequent chapters. Furthermore, this section allows for the later contrasting of results obtained through assumptions of homogeneity and those resulting from non-diagnostically biased computational methods.

Part 2 aims to generate a normative development model based on the GM volume trajectories of healthy children and adolescents which is then fit to the BD sample. Age estimation errors are interpreted as a sign of altered neurodevelopment. It is hypothesized that BD subjects will exhibit neurodevelopmental delays, and thus the predicted ages of BD subjects will be younger than their true ages. It follows that the larger the difference between a subject's true age and predicted age, the more developmentally delayed a subject is and the higher their neurodevelopmental risk for BD would be.

Part 3 aims to use non-diagnostically biased and unsupervised computational clustering methods to investigate BD heterogeneity. BD subjects are clustered based on GM volume patterns across the cortex and subcortex, which creates groups with distinct GM biotypes. The symptomology, neurocognition, GM integrity, and cortical morphology of each BD cluster is examined. It is hypothesized that GM volume biotypes will not be reflective of DSM subtype labels. Rather, data-driven biotypes will map onto other neurocognitive and morphological abnormalities while DSM subtypes will not.

Ultimately, this dissertation will further strengthen the case for BD diagnostics to extend beyond traditional symptom-centered methods and delve into neurobiological solutions that will progress the field closer to the realm of precision medicine.

Chapter 2: Processing and Quality Control of MRI and DTI Images

Introduction

Structural imaging, specifically T1-Weighted MRI and DTI, allows for *in vivo* characterization of neuroanatomy and can be used to track brain development across childhood and adolescence (Barnea-Goraly et al. 2005; Courchesne et al. 2000).

T1-Weighted sequences produce brain images by using pulse sequences that detect differences in relaxation time to visualize the contrasts between GM and WM tissues as well as CSF (Chen et al. 2015). Images can then be used to assess brain morphometry, including GM volumes, cortical thickness, and gyrification. During childhood, subcortical volume has been shown to increase while cortical volume and thickness decreases due to cortical thinning (Wierenga et al. 2014; Sowell et al. 2002). Gyrification has also been shown to decrease throughout development (Cao et al. 2017).

DTI imaging techniques allow for the visualization of WM fibers by quantifying the diffusive properties of water within tissue. The metrics typically used to describe diffusivity are axial, radial, and mean diffusivity (AD, RD, and MD, respectively), as well as fractional anisotropy (FA) values. AD measures diffusivity running along an axon while RD measures the diffusion perpendicular to an axon (Cascio, Gerig, and Piven 2007). Thus, high AD values can be interpreted as a sign of axonal integrity, while low AD values have been implicated in axonal degeneration (S.-K. Song et al. 2003). Conversely, high RD values suggest possible demyelination (S.-K. Song et al. 2002). MD is a weighted average of AD and RD and provides an inverse indicator of membrane density. FA is the weighted ratio of AD to RD, is sensitive to microstructural changes in WM, and thus serves as a summary metric of WM integrity, with higher FA values associated with healthy WM. During childhood and adolescence, FA is expected to increase with age, while MD and RD decrease (Brouwer et al. 2012). While some studies report increased AD during development, AD has also been found to decrease in certain

WM tracts, but at a slower rate than RD, thus resulting in increased FA (Qiu et al. 2008; Krogsrud et al. 2016).

When investigating the WM and GM signatures in BD, many studies use a case control design, in which BD subjects are compared to HCs. These analyses, however, are credible only if the quality of the data is assured. Thus, in order to ensure validity of results and future reproducibility, it is imperative to perform visual quality control (QC) at multiple stages throughout the image processing pipeline. Furthermore, replicating known age and gender trends within the HC data sample provides confirmation that the comparison group is indeed within the healthy norm. Otherwise, any results found will not aptly capture real-world differences between BDs and HCs, and hence will not have potential to be translated into the clinic.

The following chapter will describe the pediatric dataset that was analyzed throughout this dissertation and will detail the MRI and DTI methodologies that were used to quantify GM and WM features. In addition, this chapter will document the QC measures that were implemented.

Methods

Subjects

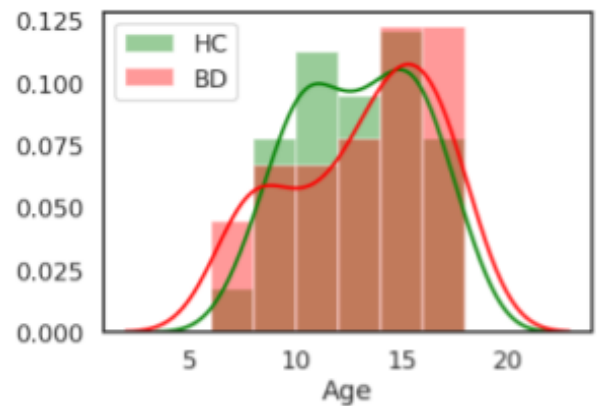
The sample was acquired through the Houston Pediatric Bipolar Consortium, which is a collaboration between McGovern Medical School (PI- Jair C. Soares, M.D., Ph.D.) and the Baylor College of Medicine (PI- Kirti Saxena, M.D. Both sites followed the same data acquisition protocols. Subjects were recruited through flyers, radio, and newspaper advertisements from the local community and psychiatric clinics. To meet the inclusion criteria for the clinical sample, subjects needed to be between six and eighteen years of age and have a

diagnosis for BD-I, BD-II, or BD-NOS according to DSM-IV. All subjects and their parent or legal guardian gave written informed consent and/or assent. This study was approved by the local IRB. Exclusion criteria included head trauma with residual effects, neurological disorders, and uncontrolled major medical conditions. Healthy controls were excluded if they had a history of any Axis I disorders, had a first-degree relative with any Axis I disorder, or used psychoactive medication less than two-weeks before the study. Subjects were evaluated through a socio-demographic history form for age, gender, and years of education. Axis-I diagnoses and clinical characteristics were assessed with the Structured Clinical Interview for DSM-IV (SCID) administered by fully trained research assistants or postdoctoral fellows who were supervised by an experienced research psychiatrist.

The final sample included 43 BD (Male: 17, Female: 26) subjects and 57 HC (Male: 28, Female: 29). There were no differences in age (BD: $M=13.1 \pm 3.5$ years; HC: $M=13.1 \pm 2.9$ years, $t(81)=0.09$, $p=0.93$) or gender ($\chi^2(1)=0.56$, $p=0.45$) between the two groups.

Imaging

McGovern Medical School's Philips Ingenia Medical Systems Scanner was used to acquire T1-weighted and diffusion weighted structural brain images for each subject. Imaging parameters can be found in Table 1. All subject scans were forwarded to a licensed radiologist in order to exclude any subjects with



gross anatomical abnormalities and neurological conditions.

Sequence	TE(ms)	TR (ms)	Voxel size (mm)	Encoding b-factor (smm^{-2})	Matrix (x,y,z)	Scan Time
3D Sag T ₁ -W	3.68	8.1	1x1x1	-	256x256x180	251
3D Sag Flair	292.5	4800	1x1x1	-	256x256x180	317
Axial 2D-Turb-SE	20/90	6800	1x1x2	-	256x256x80	231
DWI	65	9000	1x1x2	lcosa21bpn_ord (42 directions)	256x256x80	528

Table 1. Imaging Parameters.

Grey Matter Processing

After a thorough visual inspection of all the brain images for visual artifacts, MRI scan preprocessing and segmentation was performed using FreeSurfer suite version 6.0 (Dale et al., 1999; Fischl et al., 2002; Jovicich et al., 2006). FreeSurfer is a fully automated software that performs motion correction, intensity normalization, automated topology correction, and atlas based cortical + subcortical segmentation and labeling of MRI images (Bruce Fischl et al. 2004). FreeSurfer is one of the most popular methods of automated segmentation and has been shown to yield comparable results to manual tracing (Grimm et al. 2015; Perlaki et al. 2017). Segmentation results in 101 brain areas, 68 cortical and 33 subcortical. Visual QC was performed on all FreeSurfer segmentation outputs and all subjects were deemed to have been segmented satisfactorily. Example segmentations are shown in Figure 2.

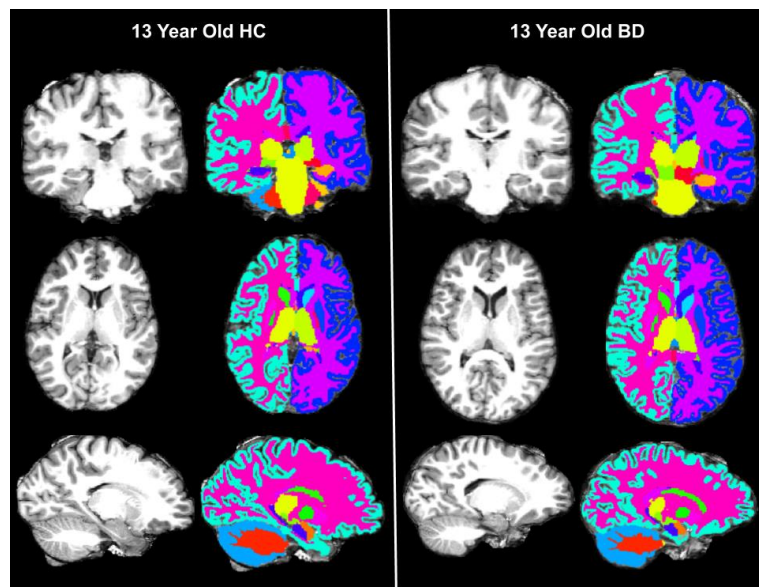


Figure 2. Example Freesurfer subcortical segmentations from two 13 year-old subjects, one healthy and one bipolar.

White Matter Processing

Using FSL's FMRIB Diffusion Toolbox (fsl.fmrib.ox.ac.uk/), FA, MD, RD, and AD maps for each subject were generated after correcting for eddy current distortions, creating a brain mask by removing all non-brain tissue using the brain extraction tool (BET), and fitting a diffusion tensor model to raw diffusion data (DTIFIT). As a visual QC, subjects' FA maps were examined for artifacts and abnormalities. Seven subjects were excluded from analysis. In order to assure that the principal vectors aligned well with neuroanatomy, the remaining subjects had their FA maps overlaid with principal eigenvector, or "V1", maps. An example overlay is shown in Figure 3.

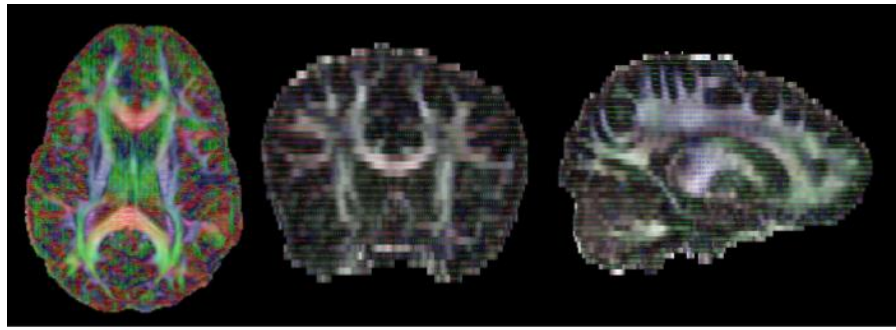


Figure 3. Example FA map overlay with principal vectors (V1) map. V1 color key: Red, left ↔ right; Blue, dorsal ↔ ventral; Green, anterior ↔ posterior.

TRACULA (TRActs Constrained by UnderLying Anatomy), an automated tractography software, was used to generate 18 major WM pathways. Specifically, diffusion maps fitted by FSL are used along with a subject's FreeSurfer segmentations and a WM atlas to conduct probabilistic tractography. An example TRACULA output is shown in Figure 4.

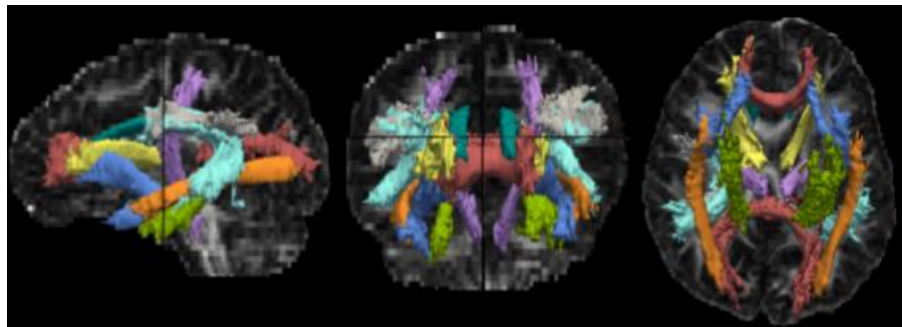


Figure 4. Example TRACULA Output for a healthy subject.

Statistics

After visual QC, the relationships between age, gender, and grey matter (GM) volumes were examined in healthy subjects to establish the validity of patterns observed in the data. General linear models examined the relationships between age and total cortical volume, subcortical volume, and FA, AD, RD, and MD for the 18 TRACULA tracts in HCs. All GLMs accounted for interactions between age and gender. General linear models were also used to look for gender differences in GM and WM. All volumetric analyses controlled for Total

Estimated Intracranial Volume (ICV). GLM analyses were conducted in R-Studio version 1.2.1335 and significance was set at $p < 0.05$.

Results

Grey Matter

Controlling for ICV, cortical volume was found to decrease with age ($p = 0.02$) and subcortical volume was found to increase with age ($p = 0.003$) (Figure 5). There were no significant interactions between age and gender. Females were found to have smaller total subcortical and cortical brain volumes than males ($p = 8.52 \times 10^{-10}$, $p = 9.8 \times 10^{-7}$).

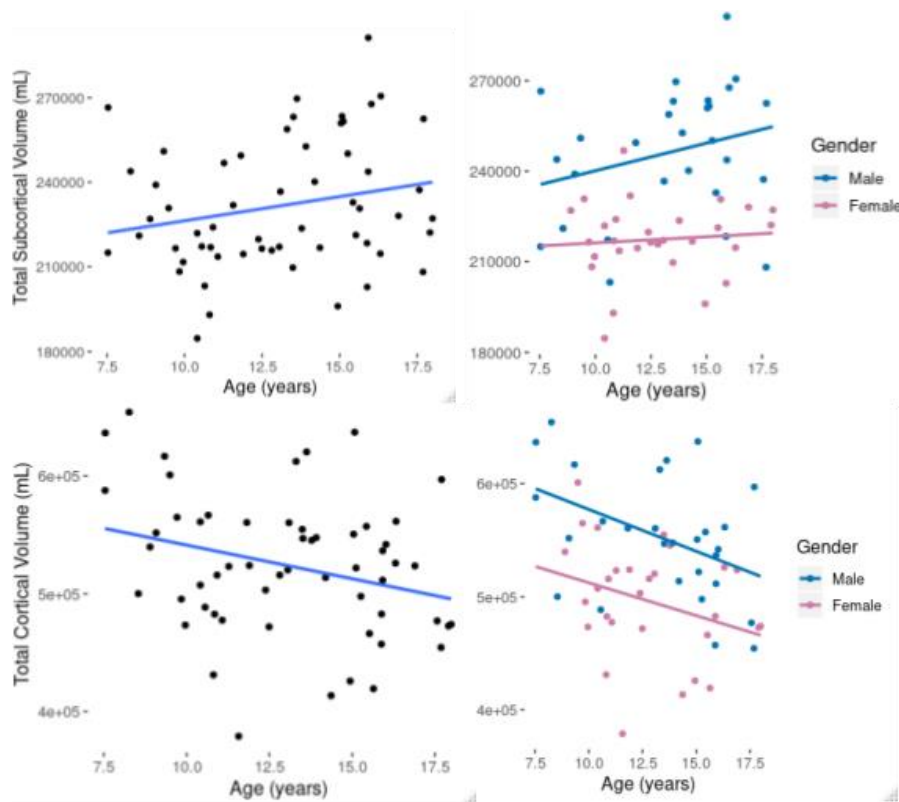


Figure 5. Total subcortical and cortical volume throughout development. Subcortical volume increases with age while cortical volume decreases. Males had significantly larger volumes in subcortical areas and cortical areas.

White Matter

No significant age effects were found in the TRACULA generated pathways when accounting for interactions between age and gender, despite only one measure having a significant interaction (FA of right inferior longitudinal fasciculus: $p_{\text{interaction}}=0.05$). This was likely due to two interaction terms decreasing the power of the analysis. When the GLM was changed to control for gender and only tested for age X group interactions, eleven measures had significant relationships with age (Figure 6). With regards to gender differences, only the FA and AD values in the cerebrospinal tract were significant, with females having lower values for both measures ($p=0.008$; $p=0.01$).

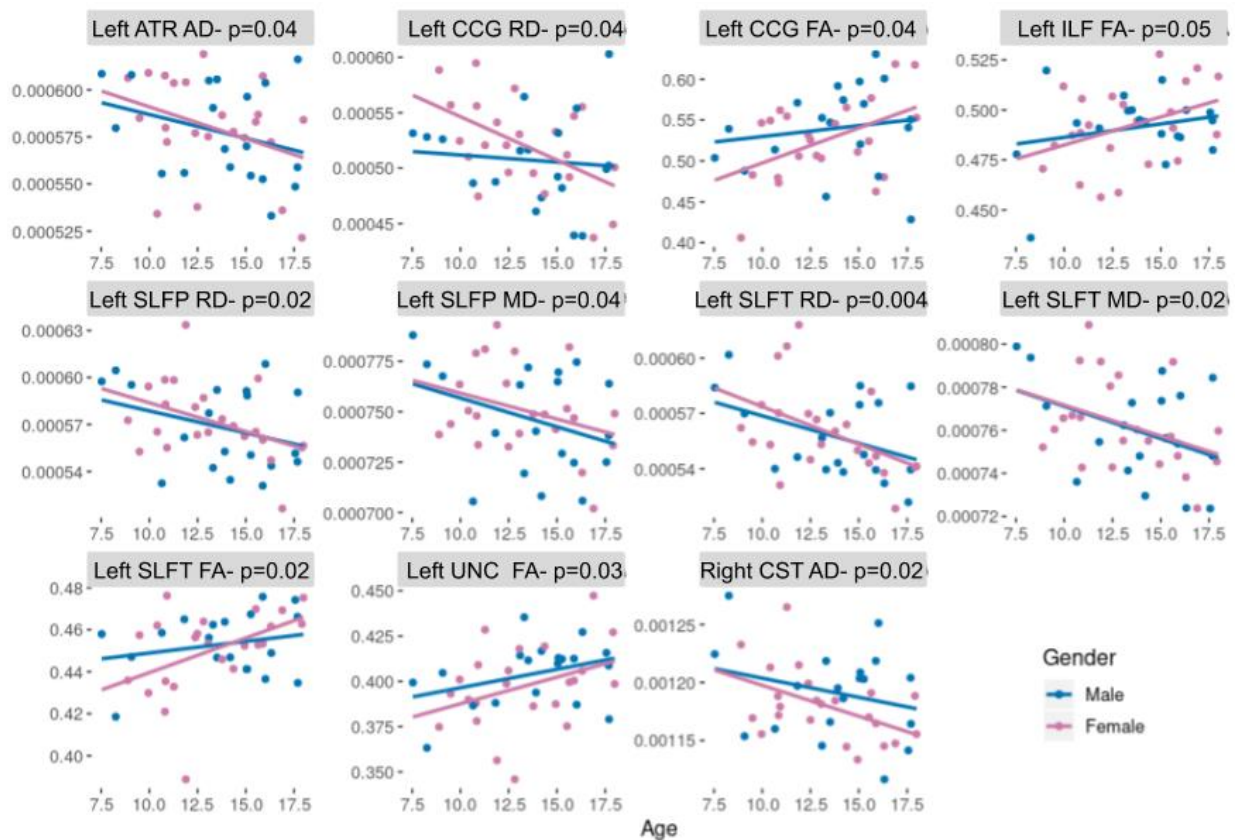


Figure 6. TRACULA pathway measures with significant age trends. Abbreviations: ATR, anterior thalamic radiations; CCG, cingulum - cingulate gyrus endings; ILF, inferior longitudinal fasciculus; SLFP, superior longitudinal fasciculus - parietal endings; SLFT, superior longitudinal fasciculus - temporal endings; UNC, uncinate fasciculus; CST, corticospinal tract.

Discussion

In this sample of 100 subjects, visual QC retained all GM data, while 7 subjects were excluded from DTI analysis. HCs exhibited trends in accordance with the literature, specifically increases in ICV and subcortical GM volumes and decreases in cortical GM volume (Sgouros et al. 1999; Sowell et al. 2002). Expected gender effects stemming from males typically having larger brains were also observed (Groeschel et al. 2010). WM showed increases in FA, which corresponded with greater WM integrity, as HC subjects develop, which is also in accordance with the literature (Lebel and Beaulieu 2011).

The confirmed age trends as well as the imaging methods and QC measures described in this chapter lay the foundation for the subsequent analyses within this dissertation. Ultimately, results are only as valid as the data that is used to generate them. Hence, assuring the quality of data and testing for reproducible patterns lends credence to the more computationally complex analyses to come.

Chapter 3: Neuroanatomical and Cognitive Differences

Between HCs and BDs

Introduction

BD in both youth and adulthood is associated with GM, WM, and cognitive abnormalities which have been shown to be linked to increased illness severity and poorer psychosocial functioning (Goswami et al. 2006; Heng, Song, and Sim 2010; Bearden, Woogen, and Glahn 2010). BD patients are often found to have smaller volumes in many brain regions throughout the cortex and subcortex. Some of these regions include structures within the limbic system, such as the hippocampus, amygdala, thalamus, and hypothalamus, which are integral to memory, sensory, cognitive, and emotional processing (Catani, Dell'acqua, and Thiebaut de Schotten 2013). Areas within the frontal cortex which are responsible for executive functioning and inhibition control have also been reported to be smaller in subjects with BD (Arnone et al. 2009) .

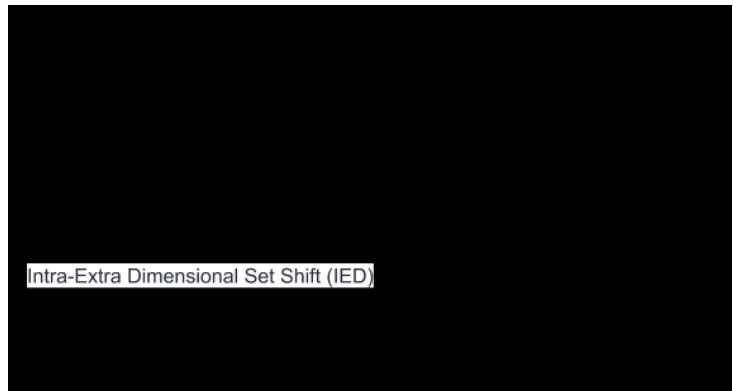
Beyond volumetric deficits, studies have suggested impaired WM connectivity in fronto-limbic areas due to decreased FA (Barnea-Goraly et al. 2009). The corpus callosum, which is the main pathway for inter-hemispheric signaling, is frequently cited as being smaller and less myelinated in BD (Brambilla et al. 2003; Caetano et al. 2008; Atmaca, Ozdemir, and Yildirim 2007). This impairment in inter-hemispheric networks is reflected in decreased efficiency and abnormal network organization in BDs (Leow et al. 2013).

Based on the literature, it appears that BD subjects suffer from widespread neuroanatomical abnormalities that are linked to impaired performance in numerous cognitive domains (Haldane et al. 2008; Poletti et al. 2015). Thus, to capture the full extent of neuroanatomical and cognitive abnormality in pediatric BD, the following chapter will explore GM from areas across the whole brain, WM in 18 major pathways, and cognitive differences in a diverse battery of tasks. It is hypothesized that fronto-limbic volumes will be smaller, WM

pathways connecting the frontolimbic system will have less structural integrity, and that cognitive performance will in turn be worse in BD youth.

Methods

GLMs examined group differences between HCs and BDs in GM, WM, and Cognition. GM and WM data were acquired using the FreeSurfer and TRACULA methodologies described in Chapter 1. Cognitive data was acquired using the Cambridge Neuropsychological Test Automated Battery (CANTAB). The list of tasks administered and the cognitive domains they measure are listed in Table 2. All analyses controlled for age and gender. GM analyses also controlled for ICV. Results were FDR-corrected for multiple comparisons. All analyses were conducted in RStudio version 1.2.1335 and significance was set at $p < 0.05$.



Results

Grey Matter

Volumes in 73 out of the 101 brain areas were significantly different in BD subjects than HCs. 71 of these regions survived FDR correction. Due to a large number of significant areas, the Bonferroni method of correction was implemented as it is more conservative than FDR (Narum 2006). After Bonferroni correction, 26 areas remained significant: 9 subcortical and 17 cortical. BDs had smaller volumes in all significant subcortical areas, which included the

bilateral cerebellum and hippocampus, the right ventral pallidum and ventral diencephalon, and left thalamus and caudate (Figure 7). The 17 significant cortical areas, which are visualized along with their Bonferroni corrected significance in Figure 8, were also smaller in BDs.

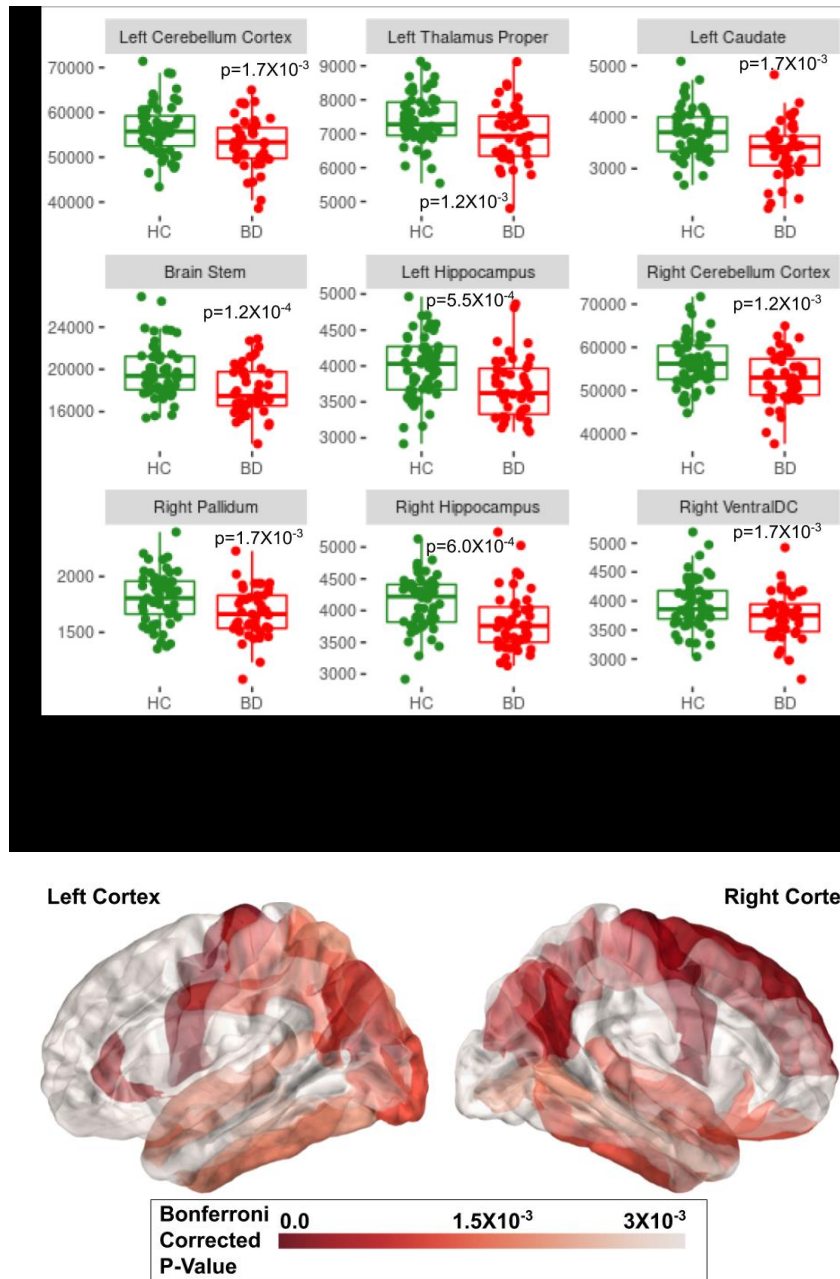


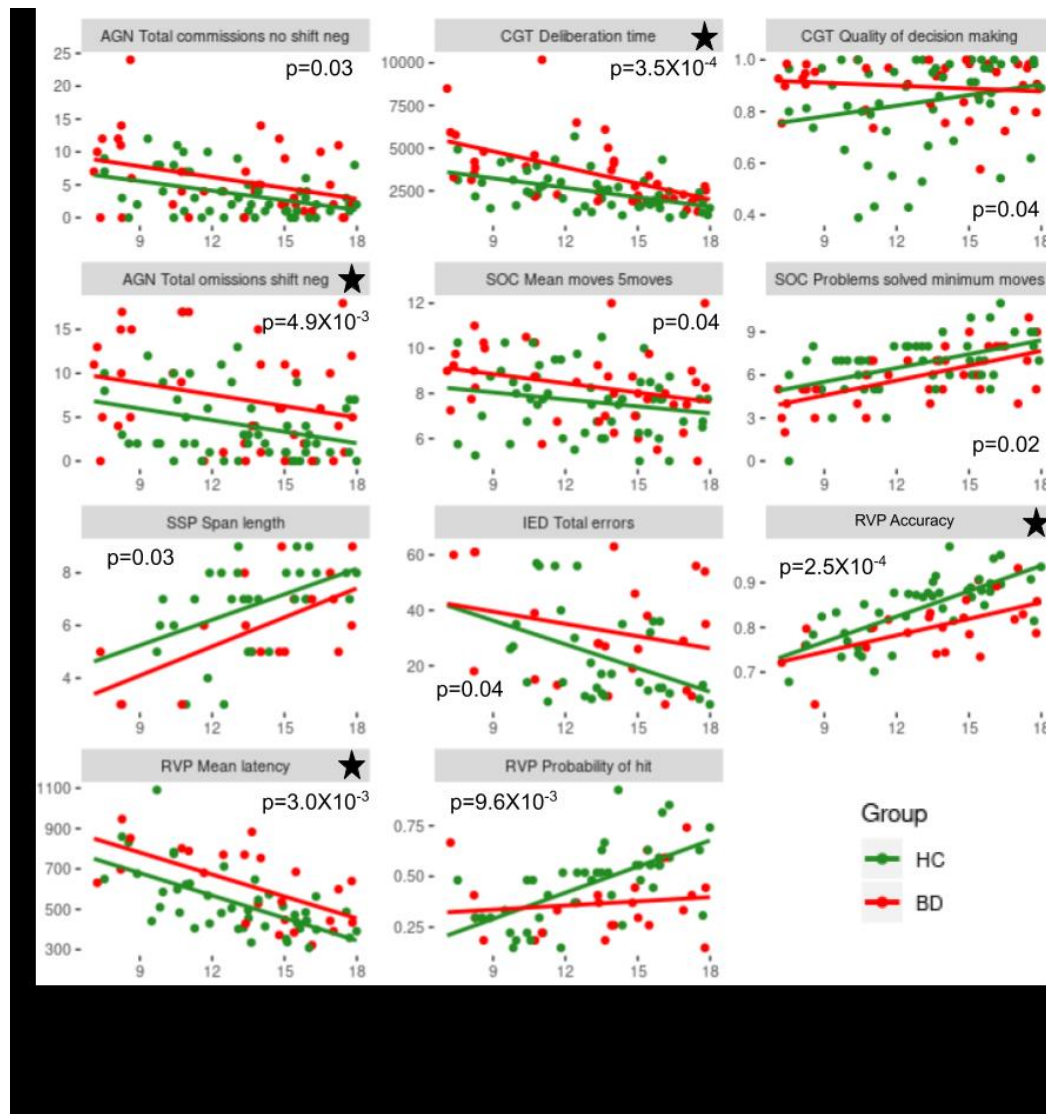
Figure 8. Cortical areas with significant differences between BDs and HCs. Seventeen cortical areas survived Bonferroni correction. In all areas, BDs had smaller cortical volumes than HCs. The figure visualizes all significant areas in red. The redder the shade, the smaller the p-value and the more significant the cortical area. All significant Bonferroni corrected p-values were less than 3.0×10^{-3} .

White Matter

There were no WM differences between HCs and BDs.

Cognition

There were differences in BD and HC cognitive performance in the AGN, CGT, SOC, SSP, IED, and RVP tests (Figure 9). There were eleven significant measures and in all except for the CGT quality of decision making, BDs performed worse than HCs. Four of the eleven significant measures survived FDR correction: AGN total omissions shift negative, CGT deliberation time, RVP accuracy, and RVP mean latency.



Discussion

Diffuse volumetric deficits were found across the subcortex and much of the cortex of BD subjects. Although there were no differences in WM integrity, compared to HCs, BD subjects showed deficits in affective processing, spatial reasoning and working memory, executive functioning and planning, and attention. These results are largely in accordance with the literature, and provide evidence for pervasive GM abnormalities and cognitive deficits in BD (Sarnicola et al. 2009).

The widespread nature of the reported BD volumetric abnormalities alludes to system-wide impairments. Thus, the lack of significant WM differences is unexpected as WM is responsible for connecting said systems. Although WM abnormalities are frequently cited in the BD literature, the absence of WM differences in BD has also been reported (López-Larson et al. 2002). It is possible that GM signatures are more robust and thus are better detected at earlier stages of BD. More studies examining both GM and WM differences in BDs and HCs are necessary to determine which is the more consistent indicator of BD pathology.

The overwhelming presentation of GM abnormalities within the BD group provides support for the possible utility of using GM metrics to build future risk assessment and decision support tools, which will be explored in later chapters.

Chapter 4: SVM Classification of BDs and HCs

Introduction

While group-wise analyses, such as those conducted in the previous chapter, give crucial insights regarding BD neuroanatomy, these group differences cannot be used to inform diagnostic decisions at the individual level (Mwangi, Tian, and Soares 2014). In addition, the widespread nature of GM abnormalities in BD reported demonstrates the necessity for models that can synthesize data from across the whole-brain (Orrù et al. 2012). Such analyses may uncover underlying links between GM areas and identify how these systems may then contribute to the neurological underpinnings of BD.

Machine learning methodologies have gained popularity within psychiatric neuroimaging due to computational capabilities that allow for the synthesizing of information from across the whole brain. Models can then use these patterns to predict diagnosis, treatment response, and other clinically valuable outcomes (Bzdok and Meyer-Lindenberg 2018; Chekroud et al. 2016; Mwangi et al. 2015). While none of these studies have yet to be successfully translated into the clinic, they serve to highlight the main predictors of mental illnesses and lay the foundation for future targeted interventions and larger scale predictive models (Gillan and Whelan 2017).

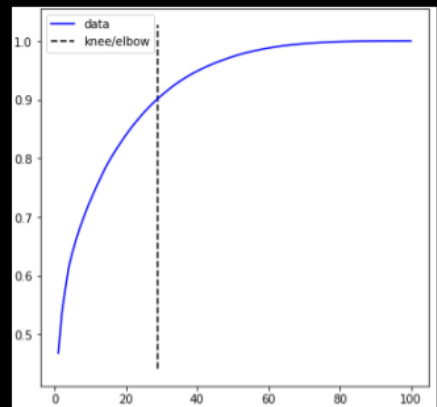
The following chapter will use GM volumes from across the cortex and subcortex to predict BD vs HC diagnosis using a principal component analysis (PCA) + support vector machine (SVM) method in children and adolescents. Based on the volumetric differences reported in the literature and in Chapter 2, it is hypothesized that GM volumes will be able to reliably separate BDs from HCs and highlight neuroanatomical signatures of pediatric BD.

Methods

The classification model was generated using ICV-corrected and scaled data in order to mitigate the effects of brain size and ensure that large brain regions do not have an outsized influence on the classification model. The 101 corrected and scaled brain regions were then fit by a Principal Component Analysis (PCA) in order to reduce the dimensionality of the dataset. PCAs collapse correlated data into orthogonal vectors, i.e. components, that preserve much of the variation in a dataset (Monfreda 2012). As many brain region volumes, especially those within the same systems, are correlated, with larger volumes in one area often being linked to larger volumes in others, GM volumetric data can be reduced to a much smaller dataset, thus increasing the power of the analysis and decreasing the risk for overfitting. The resulting principal components were then fit by a 2-class linear SVM intended to classify subjects as HCs or BDs. Briefly, SVMs use training data to fit a hyperplane that will most optimally separate subject groups in a multi-dimensional data space (Vapnik 2000; Mwangi et al. 2015).

Before running a PCA, the number of PCA components (n) must be pre-selected. This decision was made based on which n would produce the most accurate classifications. The knee/elbow method was used to select that max n that would be tested for model optimization to avoid including too many components into the final model, which could possibly render it uninterpretable (Figure 10) (Satopaa et al. 2011). The

Scikit-learn GridSearch function with leave-one-out cross-validation, was used to select the



optimal n of components in the PCA. The GridSearch tested all n components from one to the max n . The PCA result was then Z-scored using Scikit-learn's Standard Scaler before being fit by a linear SVM with the following parameters: class weight was set to the proportion of BDs and HCs in the sample (BD: 43, HC: 57), maximum iterations to 1×10^6 , tolerance to 1×10^{-6} , and the rest of the parameters remained at the Scikit-learn model defaults. The n that yielded the highest average SVM prediction accuracy, $\frac{\# \text{correctly classified subjects}}{\text{total \# subjects}}$, across all test sets, was then selected as the final n . In order to interpret the meaning of the components included in the final SVM model, for each component, the top loadings from individual brain areas were examined.

Results

Role of PCA Components in SVM Model

The knee method indicated the max n to be 29 PCAs (Figure 10) and the Gridsearch indicated that seven principal components, which capture 68.0% of the variance within the data, yielded the best SVM accuracy (Figure 11). The coefficients of the principal components making up the SVM are shown in Figure 12. Principal components with negative coefficients are characteristic of HCs, while positive

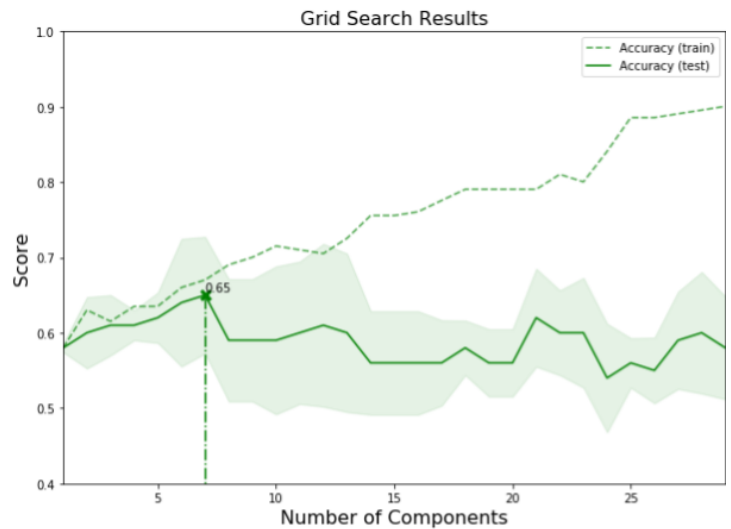


Figure 11. Gridsearch results. The n of principal components that resulted in the most accurate SVM classification model was 7. The Gridsearch shows that as the number of components is increased, training accuracy went up. However, after the 7th component, this does not improve accuracy of prediction. This is due to the increased number of PCAs causing the model to overfit the training data and thus underperform on test data.

are characteristic of BDs. Using the loadings of brain areas (Figure 13), the anatomical meaning of each component was summarized as is listed in Table 3.

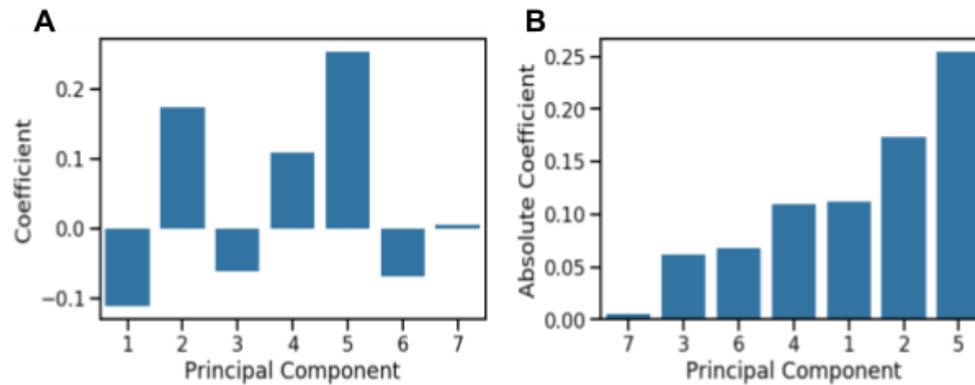


Figure 12. Coefficients of the seven principal components in the SVM models. A) Components with negative coefficients are characteristic of HCs, while positive coefficients are characteristic of BDs. B) The absolute value of coefficients is graphed from the largest contribution to the SVM model to the lowest.

Principal Component	Coefficient in Model	Summary	Characteristic of BDs or HCs
5	0.25	Measure of larger bilateral insula and insula adjacent cortices, specifically the bilateral pars triangularis and transverse-temporal volumes (figure 14)	BDs
2	0.17	Measure of a both a large brain stem, cerebellum, diencephalon and CSF system	BDs
1	-0.11	General measure of how small areas are	HCs
4	0.11	Measure of large bilateral cuneus and precuneus volumes	BDs
6	-0.07	Measure of large entorhinal cortices.	Small Effect
3	-0.06	Measure of a large CSF system	Small Effect
7	0.005	Measure of large rostral-middle-frontal cortex	Small Effect

Table 3. Coefficients and interpretations of the principal components in the SVM model. Principal components are sorted based on their absolute coefficients. Larger absolute coefficients indicate greater contributions to the classification model. Positive coefficients are characteristics of BDs while negative are characteristic of HCs.

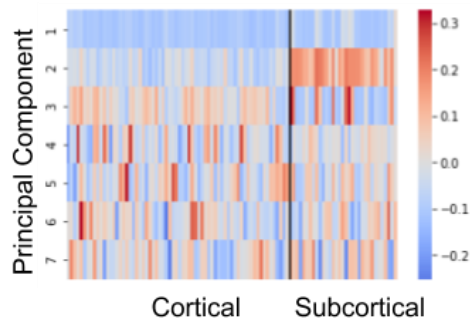


Figure 13. Loadings of individual brain areas are visualized for each component. Each column is a specific cortical area. Red bars indicate large areas contribute to a higher component value, blue bars indicate small areas contribute to a higher component value.

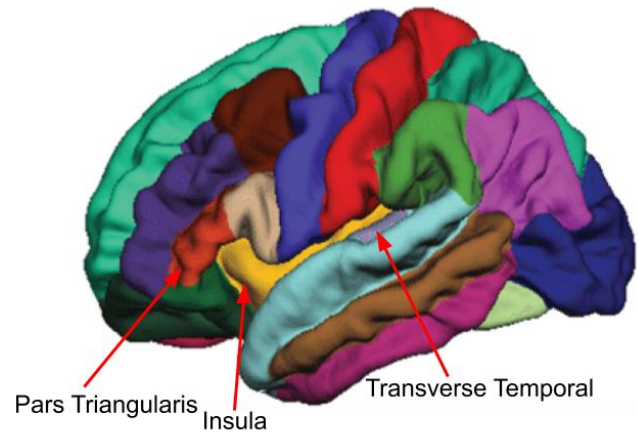


Figure 14. Insula adjacent cortices accounting for the highest loadings in principal component 5. BDs had larger volumes in these areas.

Model Performance

The final model had 63% accuracy. Figure 15 details the model's performance for HC's and BD's and indicates that the model performed better for HCs than BDs (84% VS 35% prediction accuracy). The model performed equally well for both genders, and the age distributions of correct and incorrect responses were not significantly different.

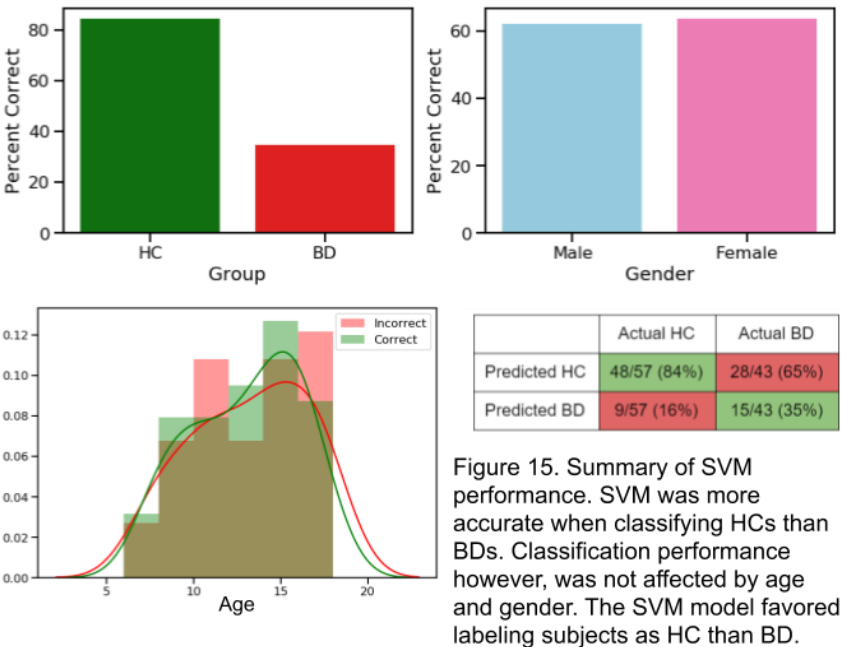
Discussion

The HC vs BD classification model had 63% accuracy.

Interestingly, the model classified 83% of HCs correctly while accuracy for BDs was much lower at 35%. Furthermore, despite accounting for an unequal

number of BDs and HCs in the training set, the model still classified 76% of subjects as HC versus the sample rate of 57%. Examining the predictors of the SVM model indicated predictive patterns contrary to the BD literature as well the results reported in Chapter 2 using classical statistical methods. ICV scaling and heterogeneity of BD are possible reasons for the discrepancies between the patterns identified by the SVM model and the BD literature as well as the poor BD classification.

Accounting for the effects of ICV in neuroimaging has become standard. This is done in order to minimize the confounding effects of cranium size on GM patterns observed (Voevodskaya et al. 2014). In classical statistical methods such as GLMS, this is often done by including ICV as a covariate (Sanfilipo et al. 2004). When conducting a PCA, however, data must be scaled, i.e. divided by ICV, as there is no method to account for covariates otherwise. These differing methods of ICV adjustment may have vast consequences on results. Indeed, the classical statistical analysis in Chapter 2 was repeated with ICV scaled data instead of the original ICV controlled analysis and all previously significant regions became null. Other



studies have also reported on the pronounced effects ICV scaling has on results (Pintzka et al. 2015; Voevodskaya et al. 2014). It is possible that scaling for ICV makes analyses less sensitive to group differences as BDs in our sample have significantly smaller ICVs than HCs ($p=0.003$) and ICV is highly correlated to GM volume. Chapter 5 will expound upon the role ICV has to play in the GM signatures of BD.

Heterogeneity of mental disorders such as BD complicate comparisons and classifications between BDs and HCs due to the fact that many analyses, including the present SVM, regard BD as being one uniform group (Van Rheenen et al. 2017; Dacquino, De Rossi, and Spalletta 2015). Research on mental disorders as well as discoveries in other fields, such as oncology and cardiology, have highlighted the fact that having the same diagnosis and self-reported symptoms does not guarantee homogeneousness when it comes to the mechanism, pathophysiology, or treatment of any given illness (Senni et al. 2014; von Minckwitz et al. 2011; Ivleva et al. 2017). Thus, Chapter 6 will explore the heterogeneity of the BD sample, how it may have affected SVM performance, and how accounting for it provides a clearer picture of neuroanatomical abnormalities of the BD sample.

Chapter 5: Normative Age Prediction Model

Introduction

BD is often regarded as a neurodevelopmental disorder and early onset BD has been associated with poorer prognosis and increased disease severity due to delays in treatment (Leverich et al. 2007). It has been postulated that characterizing healthy GM development and tracking how BD youth deviate from the typical trajectory will aid scientists in early identification and treatment of pediatric BD (Vértes and Bullmore 2015). It is important, however, that these normative models can account for the fact that the whole brain is part of the neurodevelopmental process and different regions have differing volumetric trends over time (Paus 2005). In addition, using computational resources to integrate signals from across the cortex and subcortex is likely to provide a larger effect size than relying on specific regions alone (Reddan, Lindquist, and Wager 2017).

One way to estimate normative development is through age prediction models. Specifically, pediatric subjects predicted to be younger than their age are considered to be developmentally delayed while those predicted to be older are developmentally accelerated.

Building on GM as a marker of BD abnormality, GM volumes from cortical and subcortical areas of HCs will be used to build a normative development model that would be used to assess the BD subjects. Due to BDs often having smaller GM volumes than their HC cohorts, along with other deficits in cognition and emotional regulation, it is hypothesized that when fitted by the normative model, BDs would be predicted to be younger than their true ages (Robinson et al. 2006; Dickstein and Leibenluft 2006). This would then provide further evidence of neurodevelopmental delay in pediatric BD and thus serve as a possible quantifier of BD risk.

Methods

The age prediction model was based on normative development, so HCs were used to fit a model which was then tested on BD subjects. The residuals, i.e. the error that results from the difference between the predicted age and the actual age, were used as a metric of normative development, with subjects with small residuals having typical development and large residuals having atypical development.

ICV scaled and standardized volumetric data from the FreeSurfer segmentations were fed into a PCA to reduce dimensionality before being fit into a ridge regression in order to predict age. Ridge regressions are penalized linear regressions that are able to scale the weight of predictors using L2 penalties, which are calculated by taking the square root of all model coefficients and multiplying them by every coefficient (McDonald 2009). Ridge regression models are ideal for collinear datasets and their penalization adjusts coefficients to levels that are more likely to be representative of the whole population, and thus have greater potential for translation (Pagel and Lunneborg 1985).

The PCA n was selected based on the n which would produce the most accurate ridge regression predictions. The max n components, as indicated by the knee/elbow method, was 20 (Figure 16). Due to sample size limitations, the n could not be optimized using Sci-kit learn's Gridsearch cross-validation function. Withholding a portion of the sample and training with the rest resulted in underfit models with significantly lower R-squares than when the complete dataset was used (Figure 17 A). Thus, the GridSearch could not provide useful insight regarding

the optimal n of components. Instead, a pipeline was created which performed the following for each n from 1 to the max n (Figure 17 B):

1. GM data of HCs was ICV scaled and then Z-scored using the Sci-kit Learn StandardScaler.
2. A PCA with n components was fit to the standardized data.
3. PCA output was scaled before being fit into Sci-kit Learn's Ridge Regression model with leave-one-out cross-validation. In this step, the model is fit using

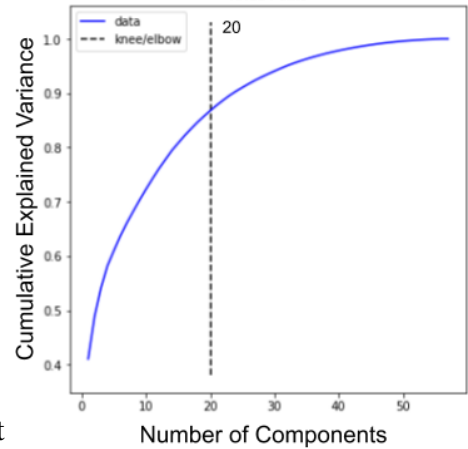


Figure 16. Result of knee/elbow method. Max $n=20$.

- all but one of the subjects. The model then predicts the age of the left-out subject. This is done until HC ages have been predicted.
4. The R^2 is calculated using the predicted and actual HC ages.

The n of components that resulted in the highest R^2 was chosen as the optimal n .

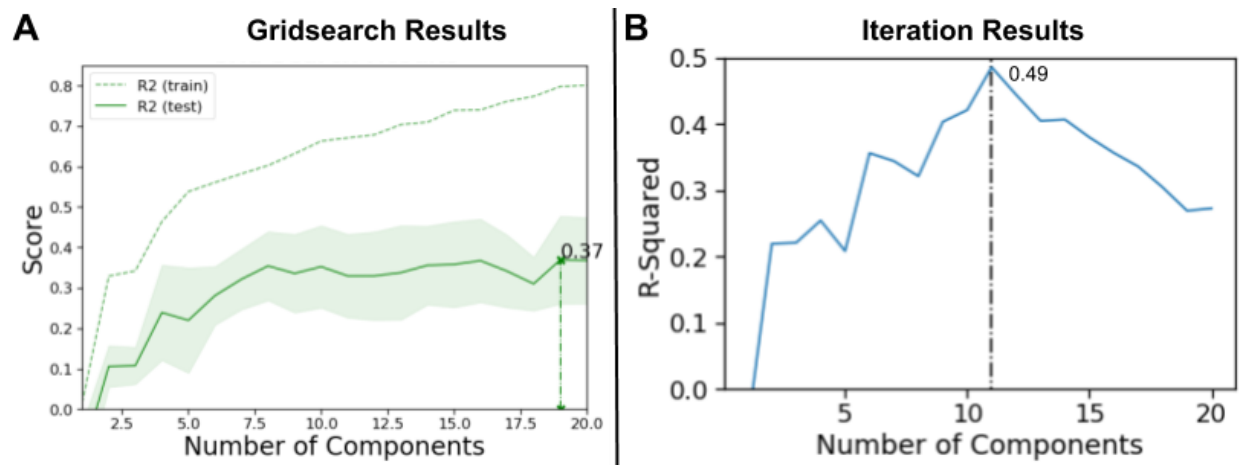


Figure 17. Comparison of Gridsearch to iteration results. Due to sample size limitations, the n could not be optimized using Sci-kit learn's Gridsearch cross-validation function. Withholding a portion of the sample and training with the rest resulted in underfit models with significantly lower R-squares than when the complete dataset was used. An iteration pipeline was built that provided more accurate estimations of R-squared values than Gridsearch.

Although this iterative algorithm does provide an empirical method for n component selection, it does not contain cross-validation at the n component step and only at the ridge

regression step. Therefore, one should be aware that this n of PCAs chosen may not be transferable to a new dataset.

In order to interpret the meaning of the components included in the final ridge regression model, for each component, the top loadings from individual brain areas were examined. To predict BD ages the final pipeline and resulting model was applied to BDs. Model performance was evaluated using residuals and the effects of age, gender, and subject group on residuals were studied using linear models and t-tests. Model generation was conducted in Python 3.7 and model performance analyses were conducted in RStudio and significance was set at $p < 0.05$.

Results

Principal Components of Ridge Regression

The knee/elbow method indicated that the max n component to be 20 (Figure 16). The first eleven components in the PCA analysis captured 74% of the variance in the data and provided the highest R^2 of the ridge regression model (0.49) with a mean absolute error (MAE) of 1.63 years (Figure 18). Table 4 shows a summary of the main loadings and ridge regression coefficients of each principal component. The table also indicates whether high component values were characteristic of younger or older subjects.

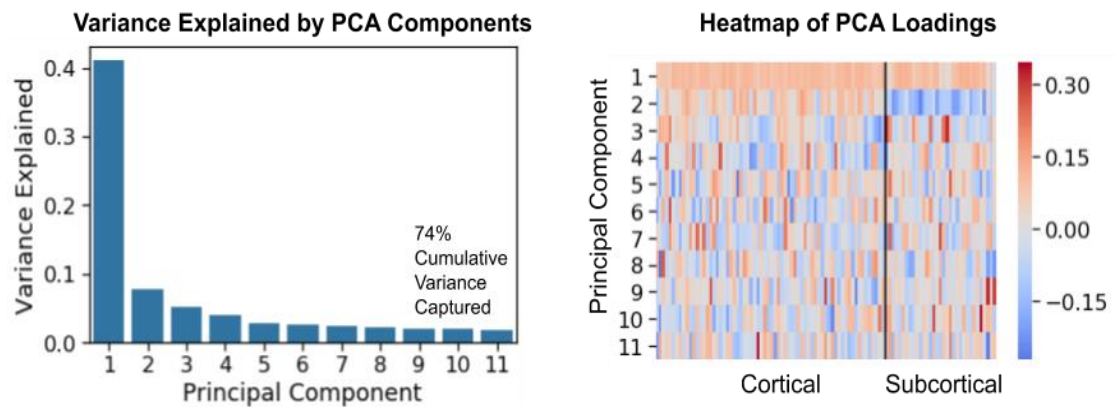
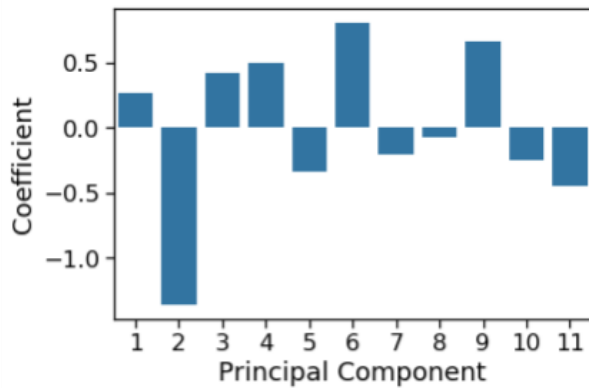


Figure 18. PCA variance explained and heatmap of loadings. In the heatmap each column represents a brain region and each row a principal component. Red bars indicate positive loadings, meaning larger values lead to a larger principal component values, while blue values indicate negative loadings.

A Coefficients of PCA Components in Ridge Regression Model



B

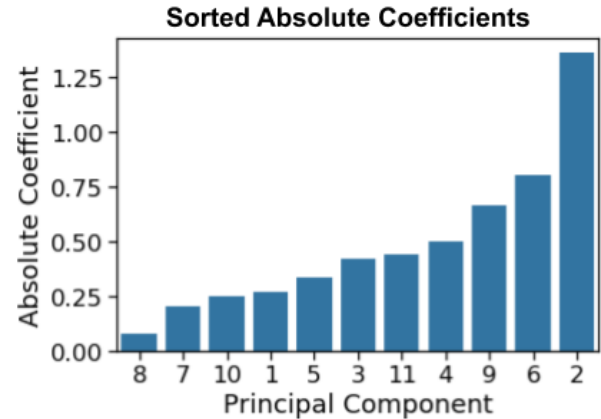


Figure 19. Coefficients of the principal components in the ridge regression model. A) Components with negative coefficients are characteristic of younger subjects, while positive coefficients are characteristic of older subjects B) The absolute value of coefficients is graphed from the largest contribution to the ridge regression model to the lowest.

Principal Component	Coefficient in Model	Summary	Characteristic of Older or Younger Subjects
2	-1.36	Measure of small brain stem, cerebellum, and ventral diencephalon	Younger
6	0.81	Measure of large entorhinal cortices.	Older
9	0.66	Measure of large bilateral pars opercularis cortices	Older
4	0.5	Measure of large cuneus, precuneus, and lingual cortices	Older
11	-0.45	Measure of large left frontal pole cortex	Younger
3	0.42	Measure of a large CSF system	Older
5	-0.37	Measure of large bilateral lateral ventricles and small bilateral fusiform cortices	Younger
1	0.27	General measure of how large areas are	Older
10	-0.25	Measure of large bilateral accumbens	Younger
7	-0.21	Measure of large frontal, temporal, and parietal cortices. Specifically the parahippocampal, entorhinal, lateral-orbitofrontal, entorhinal, and pars orbitalis cortices	Younger
8	-0.08	Measure of large caudal anterior cingulate and caudal middle frontal cortices	Small Effect

Table 4. Coefficients and interpretations of the principal components in the ridge regression model. Principal components are ranked based on the absolute value of their model coefficients. Larger absolute coefficients indicate greater contribution of a certain component to the ridge regression model. Negative coefficients are characteristic of younger subjects, positive are characteristic of older subjects. The summaries of components were generated by examining which brain areas had the largest absolute loadings for each component.

Model Performance

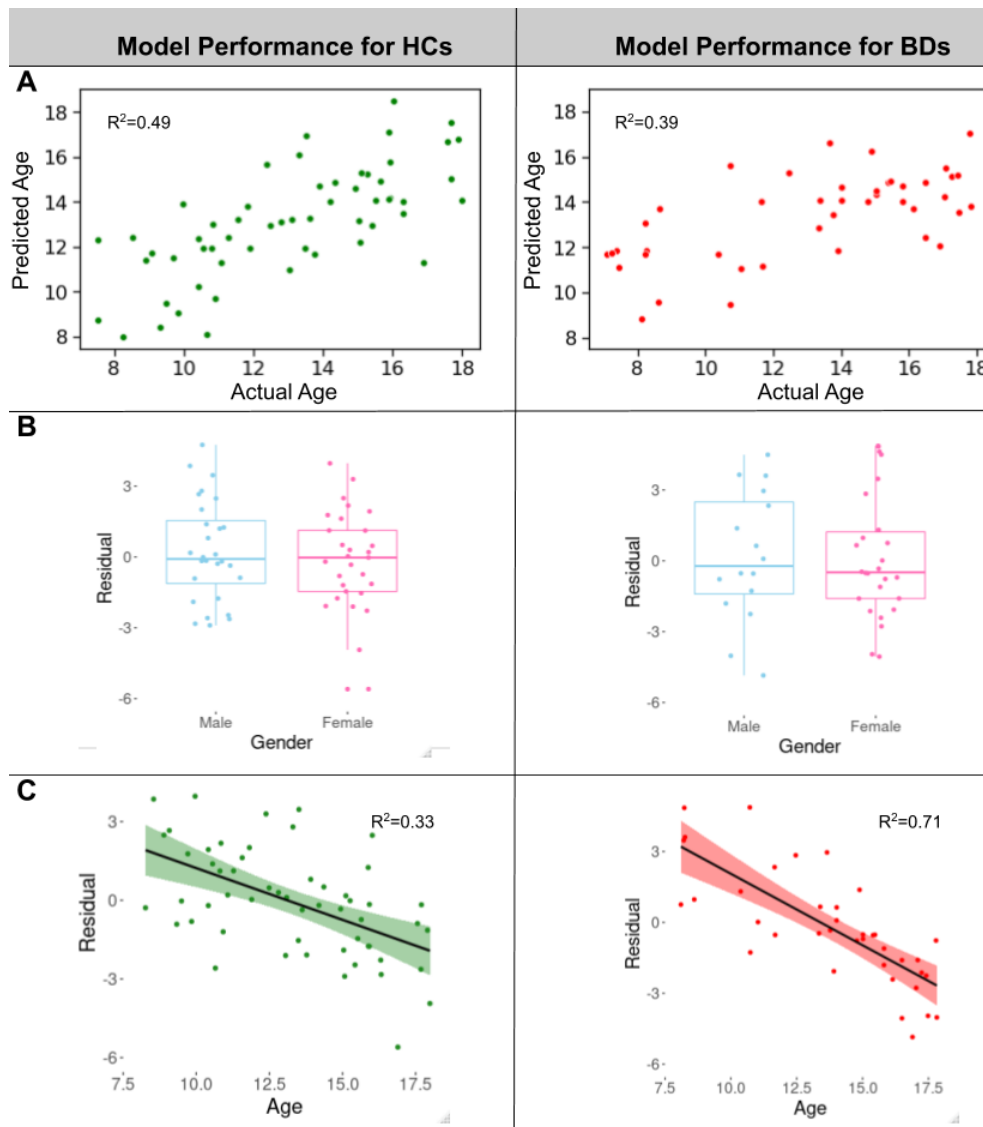


Figure 20. Summary of age prediction model performance for BDs and HCs. A) The model had a higher R^2 for HCs, indicating a better fit. B) the model performed equally well for males and females in both subject groups. C) The model performed worse for subjects at the age extremes for both subject groups.

When predicting the ages of all subjects, the final model had an R^2 of 0.44 and a mean absolute error of 1.87 years. Broken down by group, the model had R^2 values of 0.49 and 0.39, and MEs of 1.63 and 2.19 for HCs and BDs, respectively (Figure 20 A). Residuals of age predictions were not statistically different between genders for both HCs and BDs (Figure 20 B). Linear models examining the relationship between residuals and age found residuals to

decrease with age (HC: $R^2=0.33$, $p=2.91 \times 10^{-6}$; BD: $R^2=0.71$, $p=1.72 \times 10^{-12}$), meaning the model was worse at predicting subjects on the extremes of the age distribution than for subjects in the middle of the age distribution (Figure 20 C). The residuals of HCs and BDs were not significantly different (Figure 21).

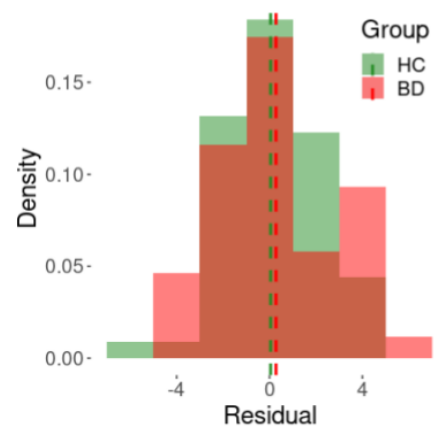


Figure 21. Distribution of residuals by group. There were no significant differences between HCs and BDs.

Discussion

The normative age prediction model identified several systems and brain regions that have been heavily implicated in neurodevelopment. The overall model had a respectable R^2 of 0.44 and was better fit to HCs than BDs (R^2 of 0.49 vs 0.39, mean absolute error of 1.63 vs 2.19 years), implying that there is more variation and heterogeneity within the BD dataset that is unaccounted for. However, on an individual level, the model could not be used to identify BD subjects at risk as there were no significant differences in residuals between HC and BDs. Therefore, it would not be possible to look at a specific subject's residual and be able to assess neurodevelopmental abnormality and BD risk.

The main component that predicted age was component 2, which was a measure of a large brainstem, cerebellum, and ventral diencephalon. These regions have direct pathways between them and play integral roles in autonomic brain functions, as well as motor control and balance (Dietrichs 1984). Thus, the PCA was able to group together a functional brain system based on the collinearity of their brain volumes across HCs. Other studies have emphasized the role of these three regions throughout development (Acton 2011; Stoodley 2016; Xie et al. 2012). Notably, a study from our group used another independent sample from NIH that contained 303 children and adolescents between the ages of 4 and 18 to develop a brain maturation index. Their resulting model placed these three brain regions within their top

predictors of maturation with the brainstem being first (Cao et al. 2015). The main CSF system component, which contained the individual parcellations of the ventricles, choroid plexus, and CSF estimates also was larger in older subjects, a pattern that is characteristic of healthy neural development (Sowell et al. 2002). The PCA also grouped together various frontal, parietal, and temporal cortical areas that tended to decrease with age. This cortical thinning has been widely reported in the developmental literature and has been associated with improvements in cognitive performance and intelligence (Squeglia et al. 2013; Vijayakumar et al. 2014; Shaw et al. 2006). There were some cortical areas, however, that were shown to increase in age, specifically the entorhinal, pars opercularis, cuneus, precuneus, and lingual cortices. Although little has been published on the developmental trajectories of these areas, a study has shown entorhinal thickness to increase during childhood and adolescence (Hasan et al. 2016). As cortical thickness and volume have been shown to have strong positive correlations with one another, it is also expected that volume would follow a similar trajectory (Storsve et al. 2014). These results give validity to the normative model as the main predictors are regions commonly cited in the neurodevelopmental literature.

When the normative model was applied to BDs, the predictions had a decrease in R^2 and an increase in mean absolute error compared to HCs. This suggests a greater degree of heterogeneity within the BD sample and more deviation from the healthy trajectory as a whole. However, once the distributions of the model's residuals were compared, they were not significantly different between HCs and BDs. This fact renders the current normative model ineffective as a way to assess neurodevelopmental deviation. While a larger sample size, such as the 303 subjects in the brain maturation index study (BMI) (Cao et al. 2015), may result in a higher R^2 (0.67 VS. 0.49), when performance is evaluated based on the mean absolute error, there was not much difference between the BMI study and the present one (1.69 years vs 1.63

years). Thus, a larger healthy training sample does not in and of itself guarantee a more precise prediction. One must take into consideration, however, that the data used was cross-sectional, and having longitudinal data may result in a more valid predictive model that can better account for intrasubject variability. However, this would still not solve the issue of intersubject variability, as large longitudinal studies still report large variation within healthy samples (Lenroot et al. 2007).

Creating a normative development model to assess BD risk has been postulated as a potential decision support tool for clinicians. However, impediments in model development may stem from the large variation within HCs themselves; there is no distinct boundary that separates BD volumes from HCs, rather the distribution of volumes across groups is largely normal. It is expected that other mental disorders may encounter similar overlaps in developmental distributions (Giedd et al. 2015). Nevertheless, the normative development model's inability to assess risk at the individual level does not discount the value of a non-supervised computational approach validating the role of many of the areas and systems which have been identified as integral to human neurodevelopment.

Chapter 6: Clustering of BD

Introduction

Illness heterogeneity is a hallmark of BD and other mental disorders. Indeed, the DSM manual indicates that there are three BD subtypes: BD, BD-I, BD-II, and BD-NOS. In brief, BD-I is marked by the occurrence of at least one manic episode, BD-II by one hypomanic and one depressive episode, and BD-NOS by alternating depressive and manic symptoms that do not meet the criteria for a BD-I or BD-II diagnosis (Glass 2009). Regarding the stability of BD subtypes, a study has shown 7.2% of BD-I subjects converting to BD-II and 7.5% of BD-II subjects converting to BD-I (Coryell et al. 1995). A longitudinal study of 263 BD children and adolescents, however, showed 20% of BD-II subjects converting to BD-I and 25% of BD-NOS subjects converting to BD I, thus implying less stability in pediatric BD subtypes (Birmaher et al. 2006). Furthermore, studies looking to identify biological differences between BD subtypes report conflicting results with little consensus in the literature (Martin Tesli et al. 2015; M. Tesli et al. 2014).

It is important to note that the DSM subtypes are defined using symptomology and not by any mechanistic, anatomic, or cognitive profiles. As symptoms may be the result of underlying pathology and various pathologies may manifest similar symptoms, exclusive use of symptoms as diagnostic criteria may limit progress towards targeted BD treatments that are tailored to the individual (Insel and Cuthbert 2015; Drysdale et al. 2017). Non-diagnostically biased clustering of patients using empirical data and machine learning algorithms has been proposed as an approach to find specific biotypes that are more clinically informative than DSM labels (Cho et al. 2019). Groundbreaking studies have found biotypes GM biotypes in adult BD, schizophrenia, and major depressive disorder that corresponded to cognitive deficits, treatment response, and clinical outcomes (Clementz et al. 2016; Drysdale et al. 2017). While pediatric biotyping studies are relatively scant, promising findings have been found using

neuropsychological data in ADHD and GM cortical thickness measures in pediatric BD (Zhang et al. 2018; Fair et al. 2012). It is still unclear, however, whether GM volumes can separate pediatric BD subjects into biotypes, and if so, how stable biotypes are over time.

The following chapter aims to be the first to investigate whether there are distinct pediatric BD biotypes, i.e. clusters, using volumes from 101 GM areas and how these clusters differ along cognitive, neuroanatomical, and clinical domains. In addition, a preliminary longitudinal sample was acquired, and clusters were tested for longitudinal stability. It was predicted that while the data-driven clusters would be associated with other empirical measures of BD severity and neurocognitive impairment, the DSM subtypes would not. This hypothesized outcome would therefore provide further evidence for transitioning to more quantitatively informed diagnostic criteria.

Methods

A 2-component PCA was conducted on the ICV and scaled GM volumes of BDs. Visualization of the PCA results indicated that there were likely 2 distinct clusters within the BD sample (Figure 22). Using Sci-kitlearn's k-means function, a 2-cluster k-means analysis was run and empirically validated using silhouette scores. T-tests and Chi-square tests were used to test for differences between clusters in age, gender, Young Mania Rating Scale scores (YMRS), medication status, Hollingshead Socioeconomic scores, and DSM subtype. DSM subtypes were also evaluated for differences in the above demographic and clinic measures using Chi-squares and GLMS.

The GM mean diffusivity (MD) of regions that played the largest roles in BD clustering were tested for group differences in a region of interest (ROI) analysis. This was done by using FSL's toolkit to create binary masks of each ROI and linearly registering the mask with the

FLIRT function to the MD maps created in Chapter 2. All registrations were checked via visual QC. The average MD value of each ROI was calculated compared across groups using GLMs. Two sets of GLMs were conducted: one comparing just the two clusters and one that added HCs as a comparison group. Both sets of GLM analyses controlled for age and gender and were FDR corrected for multiple comparisons.

Using the CANTAB battery described in Chapter 3, differences in neurocognition were evaluated for both BD clusters and BD DSM subtypes using GLMs that controlled for age and gender. Results were FDR corrected for multiple comparisons.

The SVM from Chapter Four's classification performance was re-evaluated using clusters in order to determine whether the model performed better for one cluster compared to another. Using the same methodology described in Chapter 4, a three-class SVM was used to classify subjects as belonging to Cluster 1, Cluster 2, or the HC group.

Seven of the BD subjects, three from Cluster 1 and four from Cluster 2, were brought back for a one-year follow-up. Follow-up T1 scans were acquired and processed using the same protocol detailed in Chapter 2. The baseline ICV-corrected GM volumes of BD subjects were fit using Sci-kit Learn's StandardScaler, 2-Component PCA, and 2 cluster K-means algorithms. The three fitted functions were then applied to the ICV-corrected GM volumes of the follow-up scans in order to predict cluster classification.

The clustering pipeline was performed in Python 3.7 while the t-test, chi-square, and GLM comparisons were performed in RStudio. The significance threshold for all t-tests, Chi-squares, and GLMs was set at $p < 0.05$.

Results

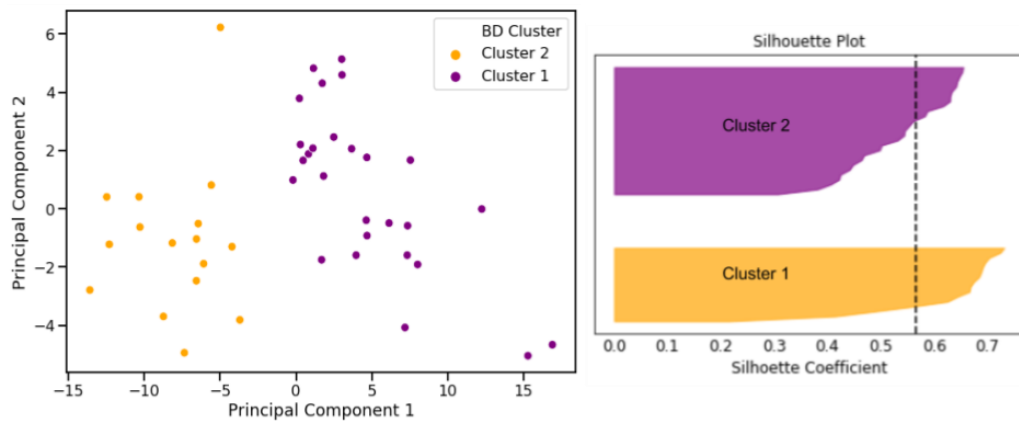


Figure 22. K-means clustering results

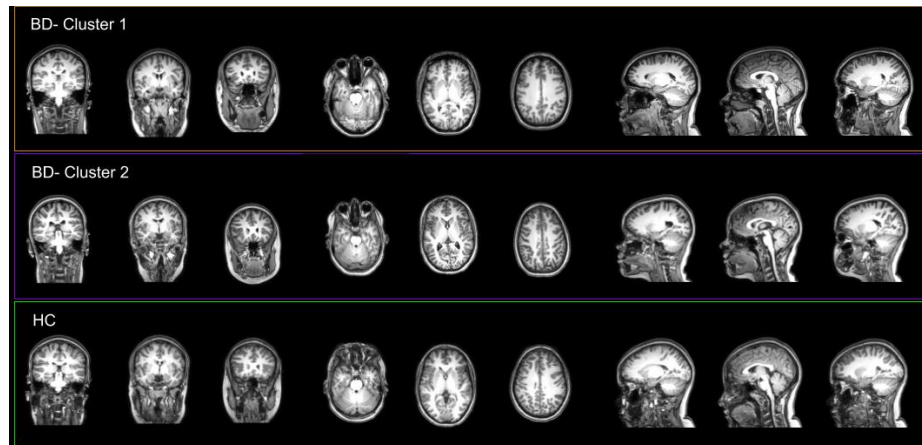


Figure 23. Example subjects from Cluster 1, Cluster 2, and HC groups. The figure illustrates that while the PCA and clustering analyses were able to detect distinct GM differences between the BD clusters, there are no gross anatomical abnormalities that are identifiable with the naked eye.

The first two principal components accounted for 60% of the variance in the data and separated the BD subjects into two clusters: 16 subjects in Cluster 1 and 27 in Cluster 2. The average silhouette score of the two clusters was 0.56 (Figure 22). Looking at the individual components, principal component 1 (PC 1) was a summary measure of small GM volumes, both cortical and subcortical. PC 2, on the other hand, was an indicator of small CSF system size (Figure 24, A and B). Thus, Cluster 2 was shown to have smaller brains than Cluster 1 and a smaller CSF system (Figure 24, C and D).

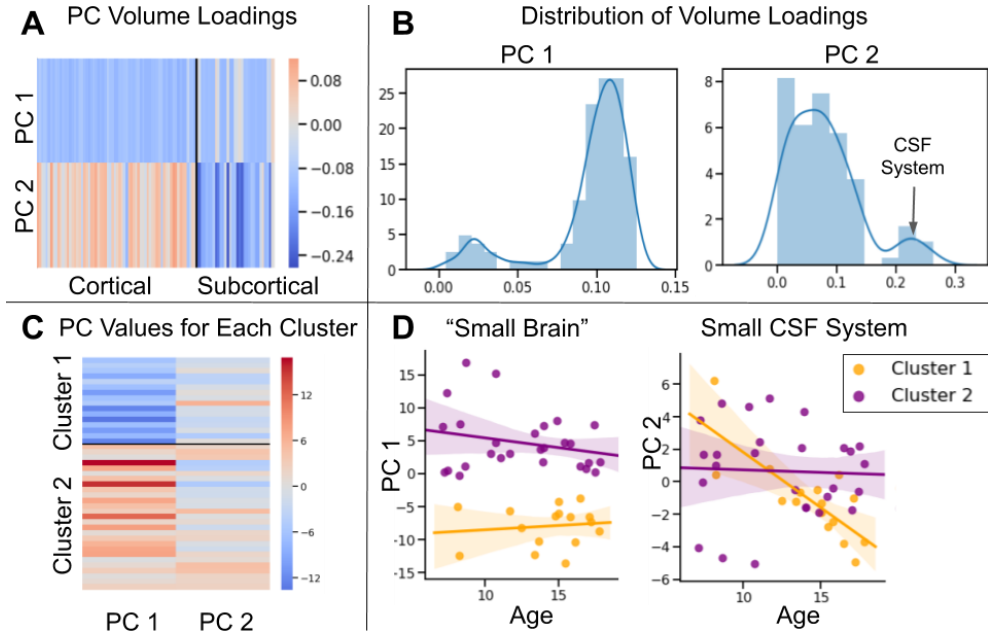


Figure 24. Principal components of K-means analysis. A) Principal component 1 is a summary “Small Brain” measure and principal component 2 is largely an indicator of a small CSF system. B) Distribution of the loadings of each of the components show that principal component 1 is a whole brain measure. The schism between top loadings and the rest is much more pronounced in principal component 2, with the CSF system having the highest loadings. C) A heatmap that shows the principal component values for each BD subject. It is clear that principal component 1 is the main divider of the two clusters, with Cluster 1 having low values and Cluster 2 having high. D) Principal components over age. The differences in principal component one between clusters are large and remain constant over age. Cluster 1 subjects, however, appear to have more rapid CSF system growth than Cluster 2.

The principal components were fit to HC and the component values were compared between the three groups (Figure 25). For principal component 1, there was a significant effect of group [$F(2, 88)=38.46$, $p= 9.93 \times 10^{-13}$], with Cluster 1 having smaller values, and thus larger brain areas, than HCs and Cluster 2s ($p= 2.07 \times 10^{-9}$, $p= 7.19 \times 10^{-12}$). HCs also had smaller principal component 1 values than Cluster 2 ($p= 4.69 \times 10^{-3}$). Principal component 2, the CSF system measures, had a significant effect of group [$F(2,88)=6.15$, $p= 3.17 \times 10^{-3}$] and a significant age X group interaction [$F(1,88)=18.52$, $p= 4.33 \times 10^{-5}$]. Post-hoc analyses confirmed that Cluster 2 has slower rate of CSF system growth than Cluster 1 and HCs ($p=0.02$; $p=0.03$). CSF system growth trended faster in Cluster 1 than HCs as well ($p=0.06$),

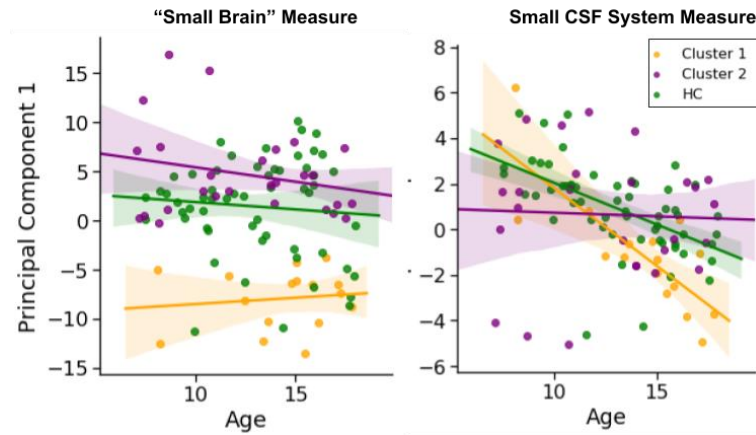


Figure 25. Three-group comparison of principal components over age. For Principal component 1, Cluster 1 had significantly larger brain areas compared to HCs and Cluster 2. Cluster 2 also had significantly smaller brain areas than HCs. In principal component 2, Cluster 2 is marked by a slower increase in CSF system size than HCs and Cluster 1s. CSF system growth in Cluster 1 also trended faster than HCs.

Characterizing Clusters

Cluster 1 and Cluster 2 did not have significant differences in medication status, Hollingshead Socioeconomic scores, or DSM subtypes (Table 6). While age and gender differences between the two groups did not reach significance, they did trend, with

Table 5. Gender of BD clusters

	Male	Female
Cluster 1	3	13
Cluster 2	13	14

Table 6. DSM subtypes of BD clusters

	I	II	NOS
Cluster 1	9	2	6
Cluster 2	19	2	6

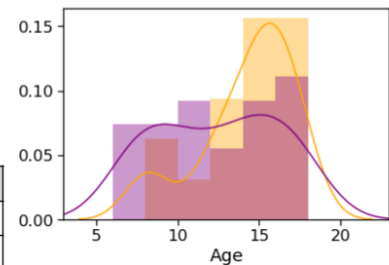


Figure 26. Age distribution of BD clusters. Cluster 1 trends older

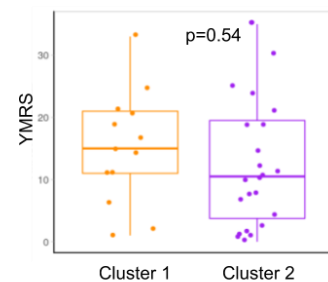
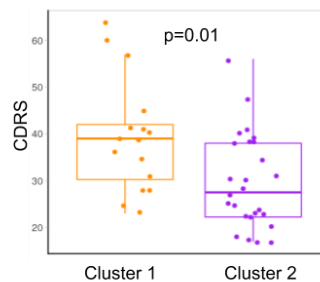


Figure 27. CDRS and YMRS differences between clusters. Cluster 1 was significantly more depressed than Cluster 2.

Cluster 1 having a larger proportion of females than males ($X^2=3.3$, $p=0.07$, Table 5) and older subjects (Cluster 1: $M=14.27 \pm 2.93$ years, Cluster 2: $M=12.43 \pm 3.70$, $p=0.08$, (Figure 26). Cluster 1 had higher CDRS scores, but there were no significant differences in YMRS scores (Figure 27). DSM subtypes were not significantly different in any clinical measure tested.

Diffusivity Differences Between GM areas of Clusters

Principal component one, the summary “small brain” measure was the main separator of the BD clusters. In order to constrain the diffusivity analysis to the most significant ROIs, the distribution of the loadings in principal component one was plotted, and any area with loadings larger than the peak were included in the analysis (Figure 28). Of the 38

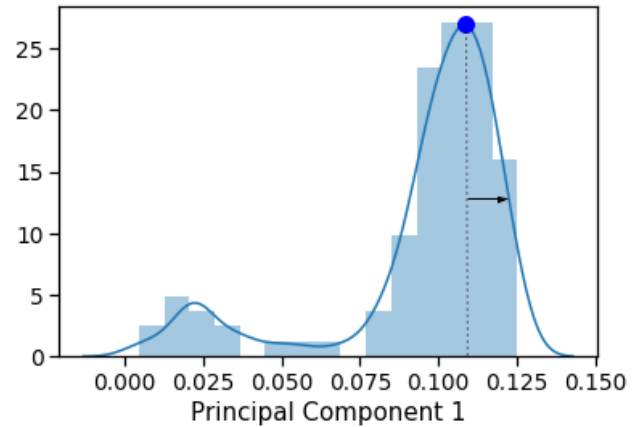


Figure 28. Distribution plot of principal component 1 loadings. In order to limit the MD analysis to the GM volumes that separate the BD clusters the most, only areas to the right of the peak were included.

brain regions that were then included in the analysis, only the bilateral amygdala was found to have higher MD in Cluster 1 than Cluster 2 (Right:

$t=2.30$, $p=0.03$; Left: $t=2.37$, $p=0.02$). This

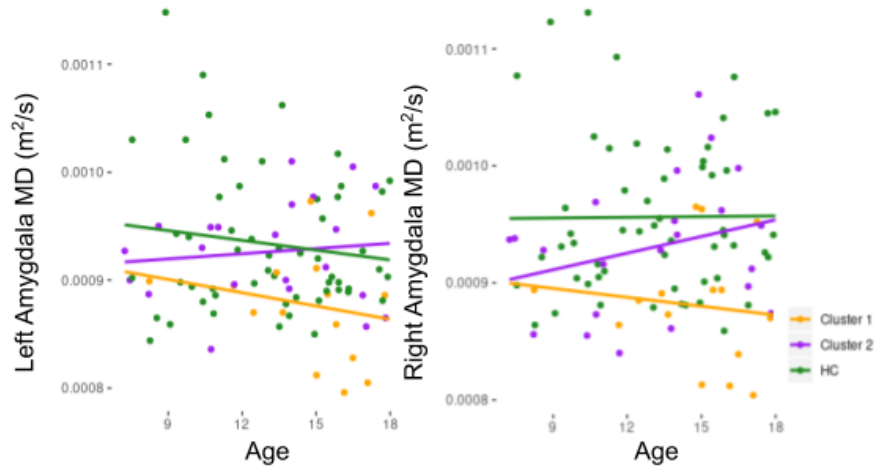


Figure 29. Differences in Amygdala MD between Cluster 1, Cluster 2, and HCs. Cluster 1 had lower MD than Cluster 2 and HC groups. Although Cluster 2 subjects appear to follow a different trajectory, interactions between age and group were not significant.

did not survive FDR correction. When HCs were included in the GLM group comparisons, Cluster 1 was also shown to have higher mean diffusivity than HCs (Right: $t=3.91$, $p=0.0002$; Left: $t=3.02$, $p=0.003$) (Figure 29). There were no differences between HCs and subjects in Cluster 2 and there was no significant interaction between age and group.

Cognitive Differences Between Clusters

GLMs comparing the cognitive performance of the two clusters found Cluster 2 subjects took more moves on average than Cluster 1 to complete the 3-move SOC tasks ($p=0.04$) and made more errors of omission in the AGN positive non-shift trials ($p=0.05$) (Figure 30). These results did not

survive FDR

correction. When

HCs were included

as a comparison

group in the GLMs,

ten CANTAB

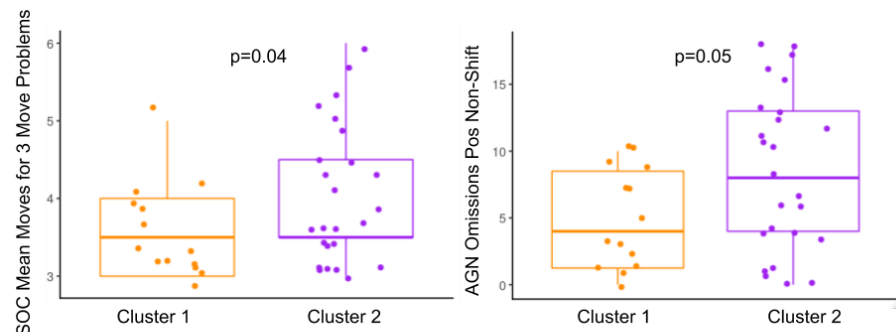


Figure 30. Cognitive differences between Cluster 1 and Cluster 2. Cluster 2, compared to Cluster 1, has impaired spatial reasoning and planning (SOC task) as well as affective processing (AGN task).

measures were significantly different between groups. Two measures, errors of omission on AGN negative shift trials and CGT deliberation time, survived FDR correction. Results are summarized in Figure 31. There were no cognitive differences between DSM subtypes.

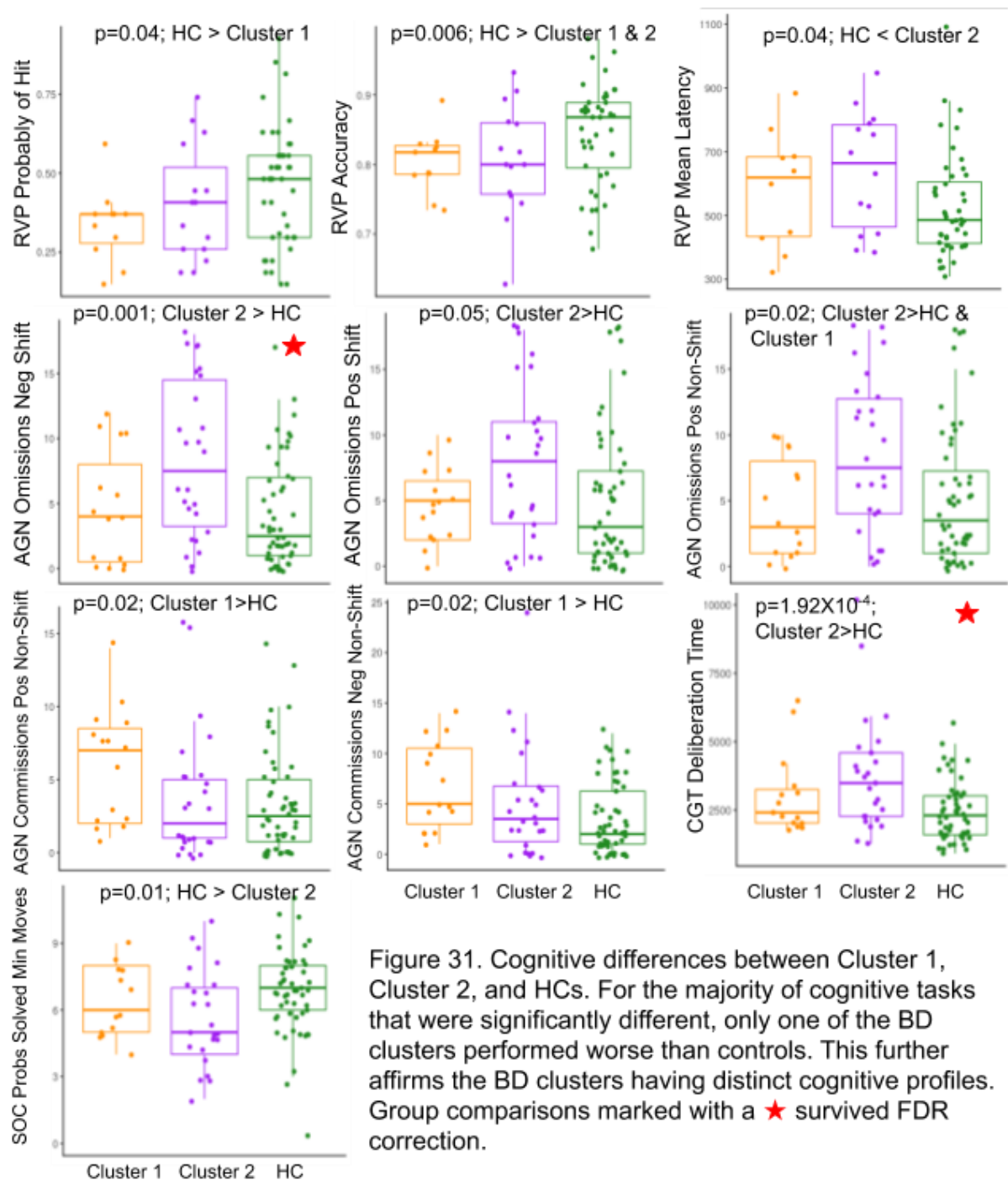


Figure 31. Cognitive differences between Cluster 1, Cluster 2, and HCs. For the majority of cognitive tasks that were significantly different, only one of the BD clusters performed worse than controls. This further affirms the BD clusters having distinct cognitive profiles. Group comparisons marked with a ★ survived FDR correction.

SVM Classification Performance Differences Between BD Clusters

As shown in Chapter 4, the SVM accuracy for the whole sample was 63%, with 84.2% accuracy for HCs and 34.8% for BDs. Evaluating performance by cluster, however, revealed that the SVM performs better for Cluster 1 than Cluster 2 (62.5% accuracy VS. 18.5%) (Figure

32 B). A three-class SVM was run to classify subjects as members of Cluster 1, Cluster 2, or HC groups (Figure 32 C). The model had an accuracy score of 57%, and the accuracy scores for each subject group were as follows: Cluster 1- 68.8%, Cluster 2- 40.7%, HC- 61.4%.

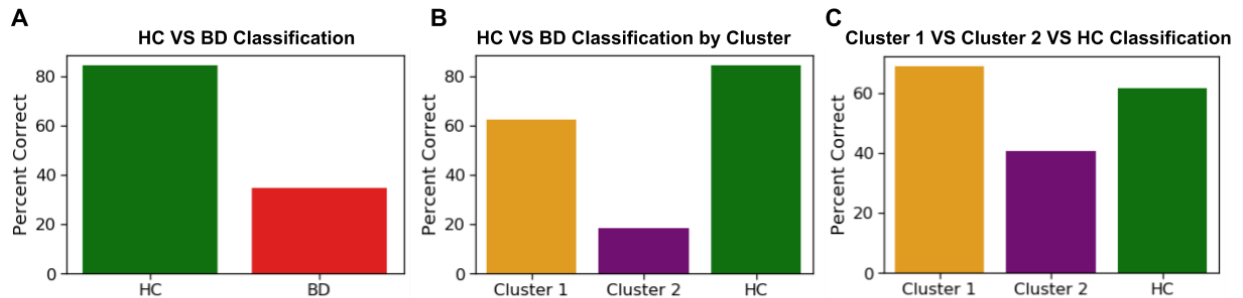


Figure 32. Revisiting SVM classification. A) The original SVM classification model performed much worse for BDs than HCs. B) When the original SVM predictions were broken down by cluster, it was clear that the model performed worse for Cluster 2 than Cluster 1. Cluster 2 subjects were more likely to be labeled as HCs because they in fact were closer to the GM profile of HCs than Cluster 1 subjects. As the Cluster 1 subject profile was distinctly different than the other two groups, the model attributed the abnormalities of Cluster 1 to the whole BD dataset. C) When a three-class SVM was conducted, differences in accuracy of classification between the three groups were not as stark.

Longitudinal Stability of BD Clusters

Five of the seven BD subjects who were rescanned a year later remained in the same cluster (Figure 33). Two subjects, one from Cluster 1 and one from Cluster 2, were classified in a different cluster.

Discussion

Using whole brain GM parcellations and PCA + K-means clustering, two anatomically distinct BD clusters were identified which had unique depressive, cognitive, and MD characteristics. As predicted, DSM subtypes did

not map on to the data-driven clusters and did not have any significant clinical, cognitive, or anatomical associations. In addition, preliminary longitudinal data suggest substantial cluster

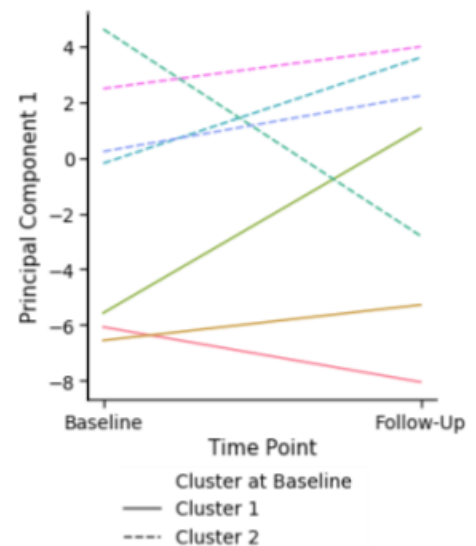


Figure 33. Cluster classification of BD subjects at two time points. Five of the seven BD subjects with follow-up scans remained in the same cluster. Changes in principal component 1, the "small brain measure" that was the main separator of clusters, are visualized.

stability. These results shed light on the heterogeneity within pediatric BD and highlight the potential pitfalls of relying on DSM labels and assuming group homogeneity.

Cluster 1 was significantly more depressed than Cluster 2 and trended older and more female. BD females tend to score higher than males on depression severity measures and report more comorbidities (F. Benazzi 1999; Saunders et al. 2012; Parker et al. 2014). Older BD youth have also been reported to be more depressed than their younger counterparts (Duax and Youngstrom 2007). Thus, it is important to parse whether the depressive profile of Cluster 1 is a result of subject demographics. Indeed, the GLM that assessed depression severity of the clusters controlled for age and gender and yielded a significant difference. For additional assurance, however, a linear model tested the relationship between age and CDRS scores and a t-test looked for CDRS score differences between genders. Both tests were null. Therefore, the link between the GM profile of Cluster 1 and depression severity is not predicated on demographics but may be a product of GM abnormality.

GM microstructure was explored in an ROI analysis using MD maps. The bilateral amygdala was found to have lower MD values in Cluster 1 than in Cluster 2 and the HC group. Lower MD values indicate less movement of water, which in GM areas is often interpreted as a sign of greater neuronal and glial density (Pierpaoli et al. 1996; Juranek et al. 2012). The clinical impact of decreased MD in Cluster 1 is difficult to deduce as few studies have been published on amygdalar MD and none of which examined a BD sample (Juranek et al. 2012; Neuner et al. 2011). However, Cluster 1 also had larger ICV-scaled amygdalar volumes. Interestingly, this trait is reported in pediatric autism research while smaller amygdalar volume is characteristic of both major depressive disorder and BD (Usher et al. 2010; Kempton 2011; Pfiefer et al. 2008; C. M. Schumann 2004; Cynthia Mills Schumann et al. 2009). Greater amygdalar volume in autism has also been correlated with increased feelings of depression and anxiety (Juranek et al.

2006). Therefore, it is possible that the higher depression score in Cluster 1 may be linked to higher amygdala volumes and MD. Further study of the role amygdalar microstructure may play in mental disorders is necessary to substantiate this finding.

The neurocognitive data provided evidence for distinct deficits for each BD cluster. Two-group comparisons highlighted differences in AGN and SOC performance, which test affective information processing and spatial planning and problem solving, respectively. Cluster 2 performed worse in both measures, failing to identify words, such as “joy” or “happiness,” as positive in the AGN task and taking more moves on average to solve certain SOC problems. When the HC group was added as a comparison group, nine cognitive measures were found to be significantly different between the three groups. Post-hoc pairwise comparisons indicated that while Cluster 1 and Cluster 2 both had significantly lower RVP accuracies than HCs, implying deficits in attention, the BD clusters exhibited differing patterns in all other significant cognitive measures. The AGN task, for example, was the most noteworthy of all the CANTAB assessments, making up five out of the nine significant measures. It was observed that Cluster 1 subjects made significantly more AGN errors of commission than HCs, indicating impaired response control, while Cluster 2 made significantly more errors of omission, thus failing to respond to the target stimulus (Seymour et al. 2015; McCormack et al. 2016). Therefore, while both clusters showed deficits in affective processing compared to HCs, a commonly reported characteristic of BD (McCormack et al. 2016), their specific impairments were in fact different. Overall, Cluster 1 and Cluster 2 appear to differ in affective processing, attention, spatial reasoning, and decision making. While in some of these domains, the difference between clusters did not reach significance and/or did not survive FDR correction, one must acknowledge the limited power provided by only having 16 subjects in Cluster 1. Nevertheless,

the HC VS Cluster contrasts provide compelling evidence of consistently unique cluster deficits that can guide future studies.

Re-examining the SVM BD vs HC classifications performed in Chapter 4 with clusters revealed the crucial role heterogeneity of BD played in model performance. In the original classifier, only 34.8% of BDs were classified correctly. However, when accuracy was broken down by cluster, the model had a much higher classification accuracy for Cluster 1 than Cluster 2. It is clear that due to Cluster 1 having a very distinct GM profile, it had an outsized influence on the hyperplane separating BDs from HCs. In addition, although Cluster 2 did have significantly smaller grey matter areas than HC group, evaluation of the principal component scores indicated that the differences between Cluster 1 and Cluster 2 were in fact much larger than Cluster 2 and HCs. Therefore, Cluster 2 subjects were often mislabeled as HCs. The three-class SVM, however, had a less skewed distribution of accuracy scores across Cluster 1, Cluster 2, and HCs. The overall model accuracy of the three-class SVM was 57% while the original two-class SVM was 63%. While this may not appear at first glance to be an improvement in classification, one must note that chance accuracy for a three-class SVM is 33.3% and 50% for a two-class SVM. Therefore, to quantify how much better model performance was from chance, the following formula was applied: $\frac{Model\ Accuracy - Chance\ Accuracy}{Chance\ Accuracy} \times 100$. The three-class SVM performed 71.8% better than chance while the two class SVM was only 26.0% better. Despite this improvement, however, the three-class model has too low of an accuracy score to be useful in a clinical setting. Nevertheless, this experiment provides evidence for the importance of accounting for heterogeneity in classification models. Should future studies employ similar methodologies but with larger samples and training datasets, it is likely that model prediction accuracy will improve. This will then increase the potential for the development of decision support tools using GM volume parcellations.

Our preliminary longitudinal data contained scans from seven subjects. While five subjects remained in the same cluster a year later, more longitudinal data is needed for clearer estimations of cluster stability. Furthermore, following subjects over the course of their development will provide vital information as to whether the clusters differ in their prognosis, e.g. suicidality, treatment resistance, hospitalizations, number of manic/depressive/hypomanic episodes, etc. Future studies should also attempt to replicate using pediatric BD data from other sites in order to better determine whether the clusters that emerged within this analysis may have been a result of sampling and/or recruitment or whether the clusters have the potential to be found in the general population.

Chapter 7: Cortical Thickness and Gyrification of BD Clusters

Introduction

The main volumetric signature of Cluster 1 was larger ICV-scaled GM volumes. An exploration of the GM characteristics led to the discovery that the subjects from Cluster 1 had significantly smaller ICVs than Cluster 2s and HCs. Interestingly, while Cluster 1’s average total GM volume was also less than the other two groups, the difference was not as large (Figure 34, Table 7). Thus, dividing

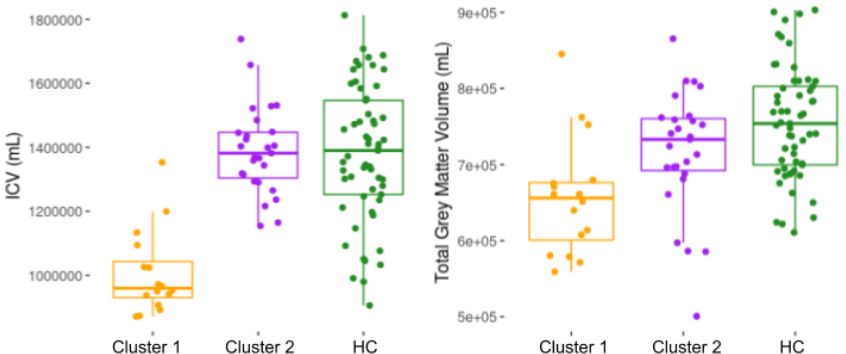


Figure 34. ICV and total grey matter differences between Cluster 1, Cluster 2, and HCs.

Cluster 1’s GM volumes by their relatively small ICVs, led to large ICV scaled areas. The question then arose as to how

	Cluster 1 VS Cluster 2	Cluster 1 VS HC
ICV	-2.86	-1.90
Total Grey Matter Volume	-0.74	-1.33

Table 7. Cohen's Differences in ICV and Total Grey Matter Volumes between groups.

Cluster 1 brains can accommodate proportionally larger GM regions. The two hypotheses that were the most anatomically compelling were that 1) Cluster 1 brains had greater cortical folding, i.e. gyrification, than Cluster 2 and HCs, and 2) Cluster 1 brains had greater cortical thickness than Cluster 2 and HCs, which would allow cortical GM regions to occupy a larger space.

Gyrification and cortical thickness follow similar patterns throughout development and have both been shown to decrease over a person’s lifespan, with childhood and adolescence experiencing the most rapid declines (Cao et al. 2017; Giedd and Rapoport 2010). Decreased cortical thickness has been consistently reported in adult and pediatric BD (Janssen et al. 2014; Rimol et al. 2010, 2012). Research on gyrification in BD, however, is more mixed. While the

majority of gyrification studies report BD subjects to have altered gyrification trajectories with less gyrification in adulthood than HCs (Cao et al. 2017; Palaniyappan and Liddle 2014; Nanda et al. 2014), one study found no differences between BD and HC gyrification (Mirakhur et al. 2009), and one has reported increased gyrification in BD (Nenadic et al. 2015). Heterogeneity in the BD population may account for the contradictory results reported in the literature.

The following chapter will test whether cortical morphology of the BD clusters is one of the driving factors behind their GM differences. Provided Cluster 1 subjects exhibit thicker cortices and/or greater gyrification, this analysis will further substantiate the proposition that the two BD clusters should be viewed and analyzed as separate patient groups.

Methods

Cortical thickness and gyrification measures were obtained using FreeSurfer's recon-all processing pipeline (B. Fischl and Dale 2000; Schaer et al. 2012). Cortical thickness is measured in mm while gyrification is measured using the local gyrification index (LGI), where higher LGI values indicate greater cortical folding. Average cortical thickness and gyrification were calculated across the whole cortex. GLMs were used to test for differences between Cluster 1, Cluster 2, and HC groups for both measures and accounted for any group X age interactions. Post-hoc was conducted using pairwise t-tests that were Bonferroni corrected for multiple group comparisons.

In order to identify specific regions that may differ in cortical morphology, vertex-wise comparisons of cortical thickness and gyrification were done using Qdec (Query, Design, Estimate, Contrast) from FreeSurfer. Qdec is a single-binary application that uses whole brain GLMs to perform two-group comparisons on surface data generated from the FreeSurfer pipeline and is the recommended software for analyzing FreeSurfer generated gyrification

measures (<https://surfer.nmr.mgh.harvard.edu/fswiki/Qdec>). Analyses of cortical thickness and gyrification were done with 20 FWHM smoothing. All other parameters remained at default and significance was set $p < 0.05$. Preliminary analysis indicated significant relationships between age and both cortical thickness and gyrification. Gender, however, did not have significant differences for either cortical measure. Therefore, Qdec analyses only controlled for age. Qdec output consists of an average subject map with colored voxels indicating where groups are significantly different.

Regions that were within areas highlighted as significant in the Cluster 1 VS Cluster 2 output maps were included in a follow-up analysis intended to characterize cortical morphology trajectories. GLMs were used to test for interactions between group and age. Post-hoc pairwise t-tests with age interactions were also conducted. Significance for all analyses was set at $p < 0.05$.

Results

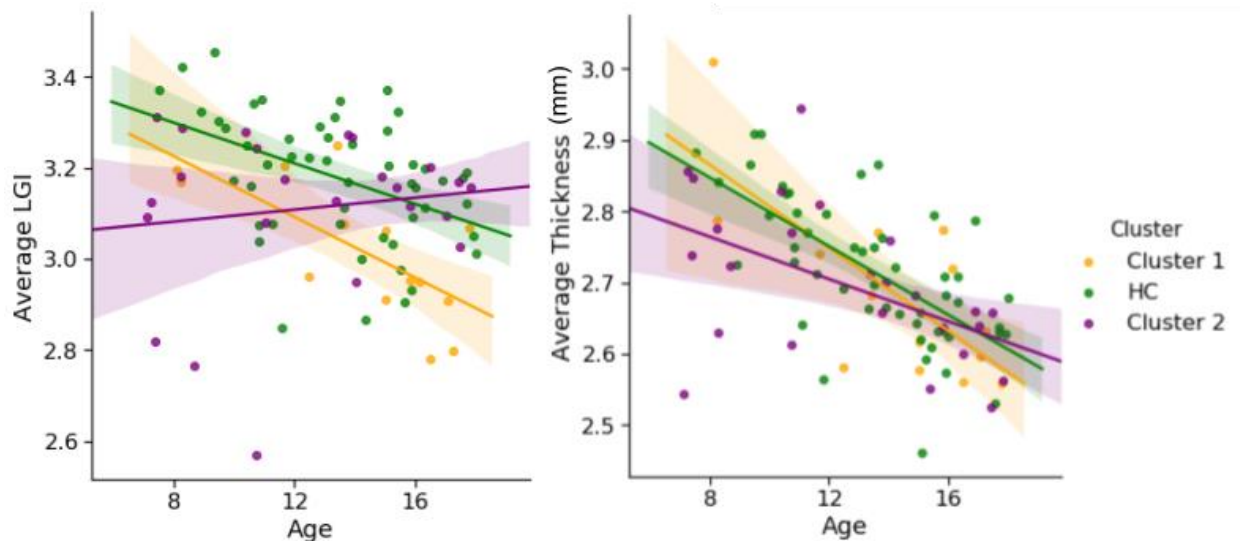


Figure 35. Average gyrification and cortical thickness in Cluster 1, Cluster 2, and HCs.

Average cortical thickness was not significantly different between groups. There were significant differences in average gyrification between groups [$F(2,83)=7.14$, $p=1.4 \times 10^{-3}$].

Cluster 1 was found to have less gyrification than HCs ($p=0.003$). There were no differences

between Cluster 2 and the other two groups. There was, however, a significant interaction between age and group, with gyrification decreasing with age in the Cluster 1 and HC groups, while gyrification slightly increased with age for Cluster 2 [$F(2,83)=5.14$, $p=7.8 \times 10^{-3}$] (Figure 36 A). This observed increase in Cluster 2 was largely dependent on three subjects with low gyrification indices. When the analysis was repeated without these outliers, the interaction only trended towards significance [$F(2,80)=2.3$, $p=0.10$] (Figure 36 B). In addition, average gyrification differences between Cluster 1 and Cluster 2 reached significance, with Cluster 1 having less gyrification.

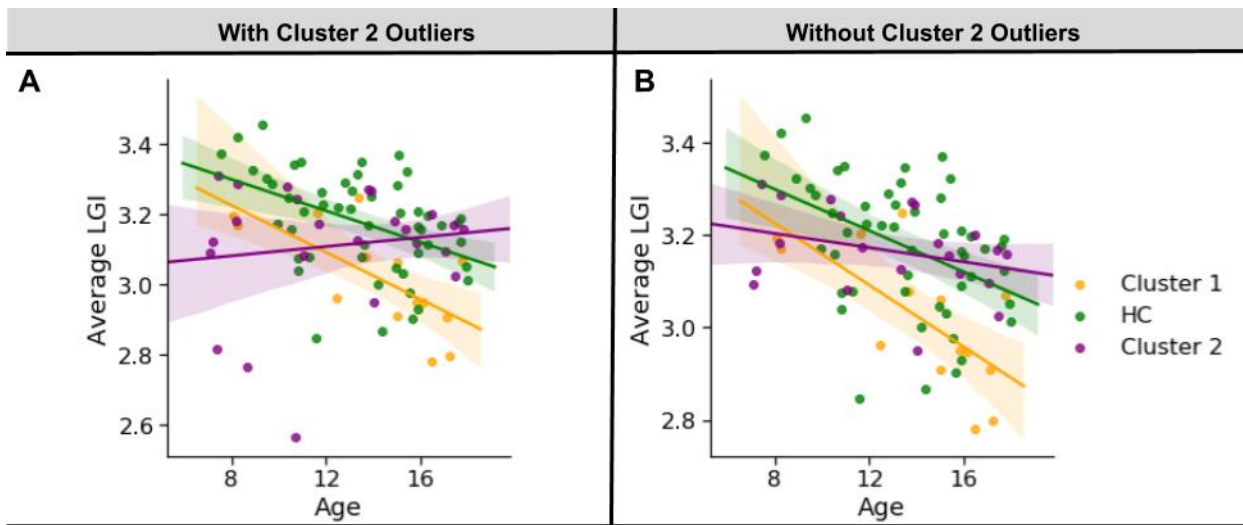


Figure 36. Comparison of the effects of three Cluster 2 outliers on average gyrification results

Qdec surface analysis is shown in Figure 37. Qdec analysis of gyrification data showed Cluster 1 and HCs to have minimal differences. Cluster 2, however, had significantly less gyrification than Cluster 1, predominantly in the frontal and occipital cortices. Once the three Cluster 2 outliers identified using average gyrification across the whole brain were removed, however, significant areas were predominantly in the frontal cortex (Figure 38). Cluster 2 also had less gyrification than HCs in much of the frontal, parietal, and temporal cortices. Regarding cortical thickness, Cluster 1 subjects had greater cortical thickness than HCs and Cluster 2s in areas within the parietal and temporal cortices. While Cluster 2 subjects also had thinner

cortices than HCs throughout the temporal and parietal cortices, they also had thicker cortices than HCs in areas within the prefrontal cortex and the occipital lobe.

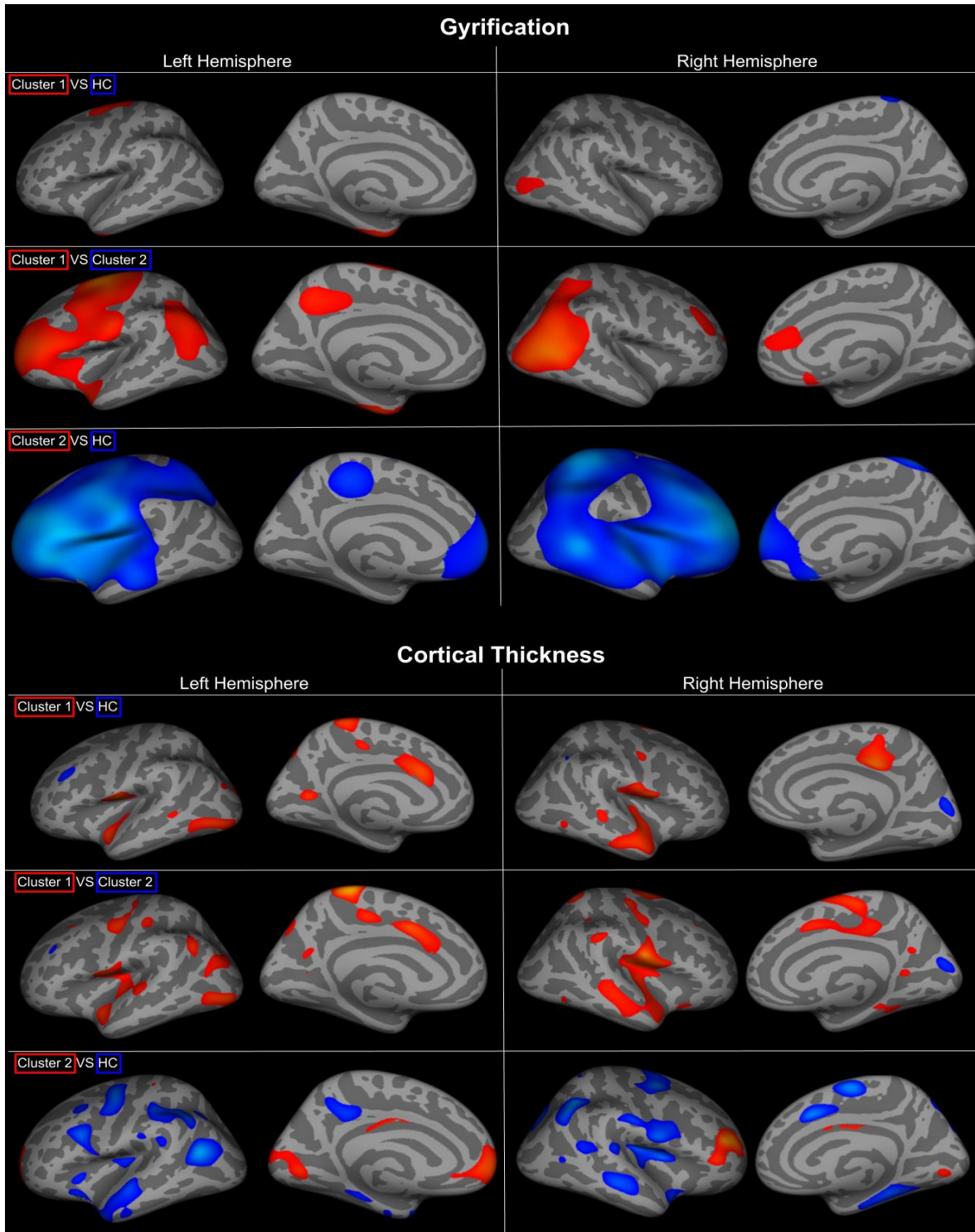
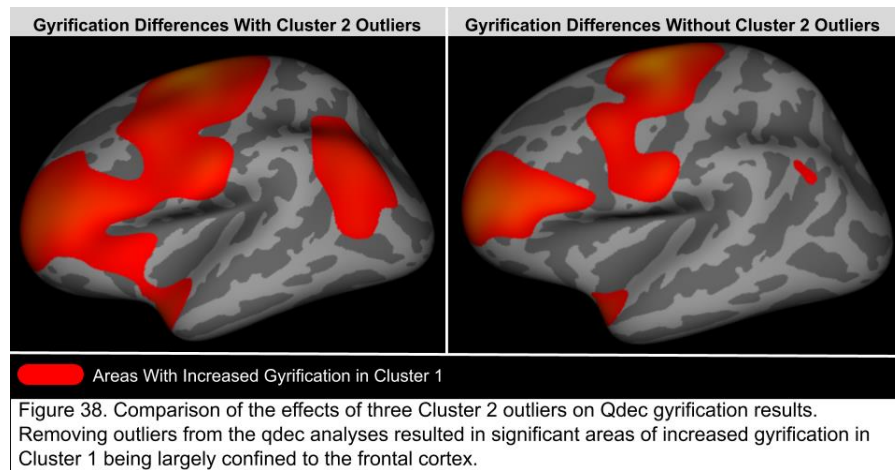


Figure 37. Qdec analyses of gyrification and cortical thickness. Each pairwise comparison is visualized. Comparisons are labeled with each subject group outlined in either red or blue. Red areas on the map indicate higher values for the group outlined in red and blue areas for the group outlined in blue.



For both gyrification and cortical thickness measures, the observed group differences were largely driven by group X age interactions.

Table 8 contains the results of interactions between group X age for areas that were highlighted in the Qdec Cluster 1 VS Cluster 2 gyrification analysis. Of the 12 regions tested, all but four had a significant interaction between age and group (Figure 39). Three, however, trended towards significance. In all eight areas with significant interactions, HCs had a more rapid, i.e. steeper, decline in gyrification than Cluster 2. Cluster 1 had a steeper decline in four areas and trended steeper in three. HCs and Cluster 1s trajectories did not differ in any region. In the left precentral cortex, however, Cluster 1 trended steeper than HCs ($p=0.09$).

Region	$F_{\text{interaction}} (2,83)$	$P_{\text{interaction}}$	Cluster 1 Steeper than Cluster 2 (p)	Cluster 1 Steeper than HC (p)	HC Steeper than Cluster 2 (p)
Superior frontal					
<i>Left</i>	2.41	0.10	0.10	-	0.10
<i>Right</i>	2.42	0.10	-	-	0.06
Precentral					
<i>Left</i>	9.01	2.9×10^{-4}	1.6×10^{-3}	0.09	2.9×10^{-3}
<i>Right</i>	3.26	0.04	0.10	-	0.04
Postcentral					
<i>Left</i>	6.38	2.6×10^{-3}	9.0×10^{-3}	-	7.5×10^{-3}
<i>Right</i>	3.635	0.03	-	-	0.02
Rostral middle frontal					
<i>Left</i>	8.63	4.0×10^{-4}	0.01	-	7.9×10^{-4}
<i>Right</i>	5.07	8.4×10^{-3}	0.07	-	6.2×10^{-3}
Frontal pole					
<i>Left</i>	3.97	0.02	0.06	-	0.02
<i>Right</i>	0.93	0.4	-	-	0.02
Pars Triangularis					
<i>Left</i>	7.63	9.1×10^{-4}	0.01	-	1.7×10^{-3}
<i>Right</i>	2.97	0.06	-	-	-

Table 8. Group X Age interactions for gyrification results. Significant p-values are highlighted in yellow and trending p-values (≤ 0.1), highlighted in orange. If the group interaction was significant, a post-hoc was conducted. Steeper references a more rapid decline in gyrification throughout development.

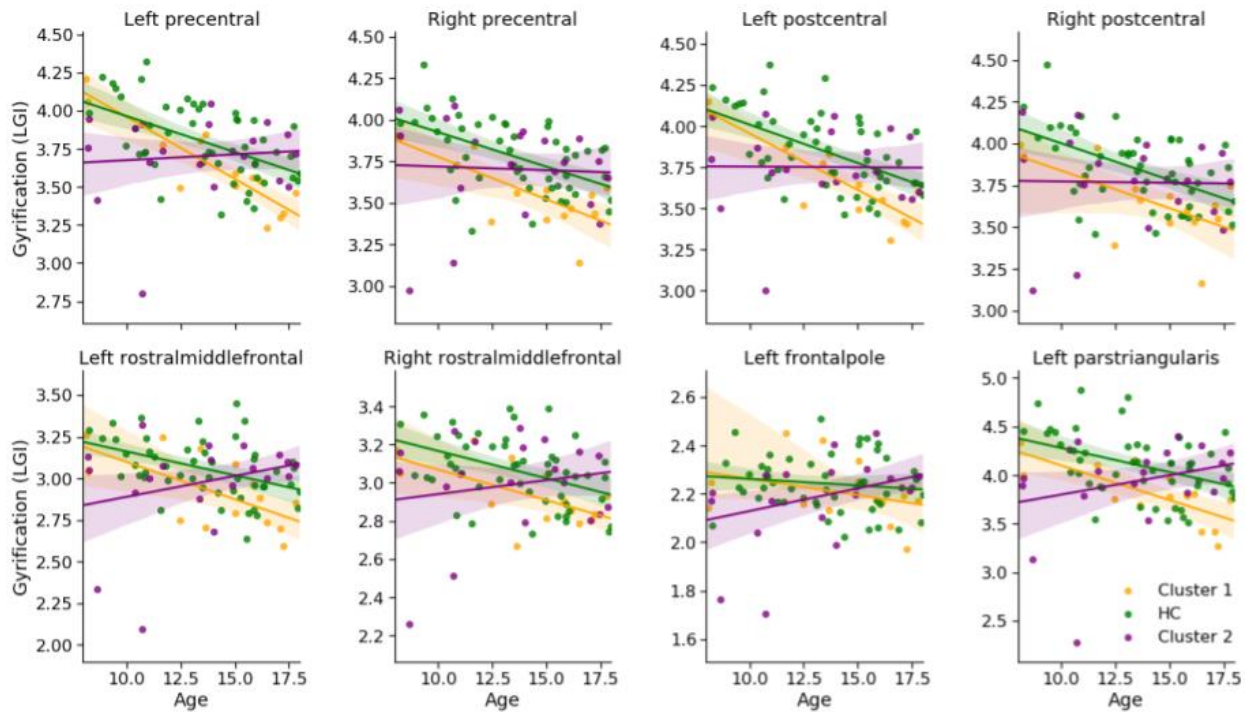


Figure 39. Qdec gyrification areas with significant age X group interactions.

Five of the 8 cortical regions identified by the Qdec Cluster 1 VS Cluster 2 cortical thickness analysis had significant interactions between age and group (Table 9). In four of the five areas with significant interactions, Cluster 1 had a steeper decline in thickness than Cluster 2 (Figure 40). Cluster 1 trended steeper than Cluster 2 in the right posterior cingulate. Cluster 1 had a steeper decline in thickness than HCs in the left paracentral area and the right posterior cingulate. HCs had a significantly steeper decline than Cluster 2 in the bilateral superior temporal and left paracentral areas. The left inferior parietal area trended steeper in HCs compared to Cluster 2.

Region	$F_{\text{interaction}} (2,83)$	$P_{\text{interaction}}$	Cluster 1 Steeper than Cluster 2 (p)	Cluster 1 Steeper than HC (p)	HC Steeper than Cluster 2 (p)
Superior temporal					
<i>Left</i>	5.34	6.54X10 ⁻³	0.03	-	0.02
<i>Right</i>	4.02	0.01	0.03	-	0.05
Paracentral					
<i>Left</i>	4.06	0.02	6.10X10 ⁻³	0.02	0.01
<i>Right</i>	2.11	0.13	-	-	-
Inferior Parietal					
<i>Left</i>	3.38	0.04	0.04	-	0.08
<i>Right</i>	0.99	0.38	-	-	-
Posterior Cingulate					
<i>Left</i>	0.08	0.92	-	-	-
<i>Right</i>	3.51	0.03	0.10	8.79X10 ⁻³	-

Table 9. Group X Age interactions for cortical thickness results. Significant p-values are highlighted in yellow and trending p-values (≤ 0.1), highlighted in orange. If the group interaction was significant, a post-hoc was conducted. Steeper references a more rapid cortical thinning throughout development.

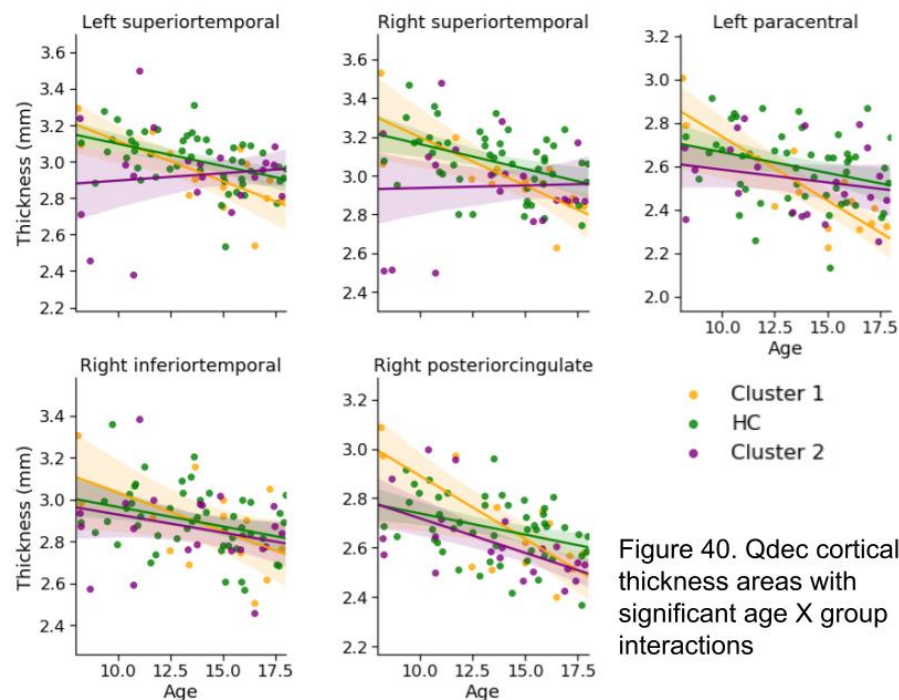


Figure 40. Qdec cortical thickness areas with significant age X group interactions

Discussion

When examining differences between Cluster 1, Cluster 2, and HCs, the averaging of gyrification and cortical thickness measures across the whole cortex resulted in markedly different results than the Qdec vertex-wise analysis. Average gyrification was lower in Cluster 1 compared to the other two groups. When gyrification was examined in Qdec, however, Cluster 1 subjects had more gyrification than Cluster 2 in the frontal cortex and gyrification differences between Cluster 1 and HCs were minimal. There were no significant differences in average thickness between groups, yet Qdec analysis identified parietal and temporal regions where Cluster 1 had thicker cortices than both Cluster 2s and HCs. Post-hoc testing and visualization of significant areas highlighted in the Qdec Cluster 1 VS Cluster 2 analysis attributed the patterns observed to be largely due to differences in age trajectories for gyrification and cortical thickness between clusters. These results underscore how data averaged across the whole brain may efface patterns in specific regions and thus not be representative of the complete picture (Raznahan et al. 2010). The regional results, however, highlighted specific neurodevelopmental abnormalities in both BD clusters.

Visualization of areas with significant cortical thickness and gyrification differences between clusters suggests that Cluster 2 subjects experience slowed cortical development compared to Cluster 1 and HCs. Healthy trajectories show prominent decreases in both gyrification and thickness during childhood and adolescence, which was exhibited by the HC and Cluster 1 groups (Piccolo et al. 2016). This may impact Cluster 2 subjects' cognitive development as slower rates of cortical thinning have been linked to increased hyperactivity and impulsivity (Giedd and Rapoport 2010).

In significant regions identified by the Qdec analyses, Cluster 1 neurodevelopment followed a distinctive pattern: Younger subjects had high levels of gyrification and thick

cortices, which was then followed by marked decreases in both morphological measures throughout adolescence. This trajectory is also exhibited in HCs. However, in the right posterior cingulate and left paracentral areas Cluster 1 exhibited more rapid cortical thinning than HCs, and in the left precentral cortex, decreases in gyrification trended faster in Cluster 1 than HCs. Interestingly, this pattern of accelerated thinning and de-gyrification during adolescence has been reported in autistic spectrum disorder (Zielinski et al. 2014). Thus, the cortical development of Cluster 1 may point to disruptions in the typical synaptic pruning process.

A 2018 study by Zhang et al examining cortical thickness in pediatric BD found two biotypes similar to the clusters reported in this study: Compared to HCs, biotype 1, which contained 16 out of the 52 subjects, had thicker cortices in the temporal and parietal lobes while biotype 2 had thinner cortices in the same cortical areas (Zhang et al. 2018). Beyond their cortical thickness differences, the two biotypes also responded differentially to quetiapine, an atypical antipsychotic. There is notable consonance between the Zhang et al study and the results presented in this dissertation, specifically in the cortical thickness differences that characterize the biotypes as well as the biotype to BD sample ratios. This, in turn, provides preliminary evidence for the reproducibility of the clusters and suggests that these clusters may have therapeutic relevance. Indeed, two studies are not sufficient for complete validation of the clusters as unique BD biotypes and more corroborating research is needed. Nevertheless, the identification of these two pediatric BD clusters that exhibit unique GM morphology and neurocognitive deficits is a major step forward in the quest to understand and ultimately treat BD heterogeneity.

Chapter 8: Discussion

The present dissertation provided an exploration of two potential approaches for the development of biologically informed decision support tools for the characterization of pediatric BD. While the normative development model was unable to quantify BD risk at an individual level, it nevertheless emphasized the heterogeneous nature of both healthy and BD development. The clustering analysis, however, was successful at parsing the variance in the BD sample which resulted in the identification of two distinct BD biotypes. Whereas the BD clusters mapped onto specific anatomical and neurocognitive patterns, symptom-based DSM subtypes were not associated with any empirical measures of mental health. Hence, these findings are testament to the potential of unsupervised and unbiased computational methods in the future of BD diagnostics.

The results of the classical statistical analysis from Chapter 3 and the SVM classification in Chapter 4 serve as an example of how assumptions of BD homogeneity can fail to identify meaningful patterns in the data. Using standard methods of diagnosis, ICV correction, and tests for group differences, the results were largely in accordance with the BD literature and would have further entrenched expectations of a specific BD profile. As was shown in the clustering analysis from Chapter 6, this approach effectively overlooked Cluster 1, which had a unique biotype with stark anatomical differences and accounted for over a third of our BD sample. Indeed, Cluster 1 subjects received the same diagnosis and were medicated similarly to Cluster 2. This highlights how symptom-based diagnostics often fail to reflect neurobiological mechanisms. It is highly likely that distinct biological abnormalities may require different methods of treatment (Yun and Lee 2019). Therefore, Cluster 1 subjects may not respond to therapies that are developed and tested within the typical symptomologic framework. Thus, in order to transition psychiatry into the field of precision medicine, research must refrain from

making assumptions of homogeneity, otherwise specific biotypes will be overlooked and averaged out (Wolfers et al. 2018).

In the examinations of the anatomical and neurocognitive differences between the BD clusters, Cluster 2 most frequently adhered to the patterns reported in the BD literature. The Cluster 2 profile included smaller GM areas, impaired executive planning, attention, and spatial reasoning, and abnormal cortical thickness and gyrification trajectories. As Cluster 2 accounted for the majority of BD subjects within the sample, with 27 of the 43 subjects, it is likely that Cluster 2 contains subjects who mostly match the common clinical characterization of BD. Cluster 1, on the other hand, had unique abnormalities that did not fit the typical BD neuroanatomical profile. Rather, some of the identified Cluster 1 patterns, such as enlarged amygdalae and accelerated cortical thinning in adolescence, have been reported in autistic spectrum disorder (ASD) (Kohli et al. 2019; Ulay and Ertuğrul 2009). While none of the BD subjects had comorbid ASD diagnoses, it is important to note the high rates of BD/ASD comorbidity and the possible etiological parallels between the two disorders (Munesue et al. 2008; Kerbeshian and Burd 1996).

Future studies should address not only the heterogeneity that exists within disorders but also the blurred boundaries between them. Many symptomological criteria are shared among mental disorders, which increases subjectivity in current clinical practices. Consequently, sole reliance on observation makes psychiatry susceptible to increased diagnostic errors and decreased rates of treatment response (Saeed 2018; Gillan et al. 2017). Studies have shown clinical overlap and/or high rates of comorbidity between BD and major depressive disorder, ADHD, schizophrenia, and ASD (Franco Benazzi 2003), and the similarities between BD Cluster 1 subjects and youth with ASD are a testament to this fact. Psychiatric diagnosis must evolve to center around the distinct deficits exhibited rather than non-mechanistically specific

labels in order to increase therapeutic efficacy. Thus, future studies may build upon these findings to develop targeted treatments based on a pediatric patient's GM biotype, making personalized care accessible during a critical period of mental development.

References

- Alonso J¹, Petukhova M, Vilagut G, Chatterji S, Heeringa S, Üstün TB, Alhamzawi AO, Viana MC, Angermeyer M, Bromet E, Bruffaerts R, de Girolamo G, Florescu S, Gureje O, Haro JM, Hinkov H, Hu CY, Karam EG, Kovess V, Levinson D, Medina-Mora ME, Nakamura Y, Ormel J, Posada-Villa J, Sagar R, Scott KM, Tsang A, Williams DR, Kessler RC. 2011. “Days out of Role due to Common Physical and Mental Conditions: Results from the WHO World Mental Health Surveys.” *Molecular Psychiatry* 16 (12): 1234–46.
- Arnone, Danilo, J. Cavanagh, D. Gerber, S. M. Lawrie, K. P. Ebmeier, and A. M. McIntosh. 2009. “Magnetic Resonance Imaging Studies in Bipolar Disorder and Schizophrenia: Meta-Analysis.” *The British Journal of Psychiatry: The Journal of Mental Science* 195 (3): 194–201.
- Atmaca, Murad, Huseyin Ozdemir, and Hanefi Yildirim. 2007. “Corpus Callosum Areas in First-Episode Patients with Bipolar Disorder.” *Psychological Medicine* 37 (5): 699–704.
- Bak, N., B. H. Ebdrup, B. Oranje, B. Fagerlund, M. H. Jensen, S. W. Düring, M. Ø. Nielsen, B. Y. Glenthøj, and L. K. Hansen. 2017. “Two Subgroups of Antipsychotic-Naive, First-Episode Schizophrenia Patients Identified with a Gaussian Mixture Model on Cognition and Electrophysiology.” *Translational Psychiatry* 7 (4): e1087.
- Barnea-Goraly, Naama, Kiki D. Chang, Asya Karchemskiy, Meghan E. Howe, and Allan L. Reiss. 2009. “Limbic and Corpus Callosum Aberrations in Adolescents with Bipolar Disorder: A Tract-Based Spatial Statistics Analysis.” *Biological Psychiatry* 66 (3): 238–44.
- Barnea-Goraly, Naama, Vinod Menon, Mark Eckert, Leanne Tamm, Roland Bammer, Asya Karchemskiy, Christopher C. Dant, and Allan L. Reiss. 2005. “White Matter Development during Childhood and Adolescence: A Cross-Sectional Diffusion Tensor Imaging Study.” *Cerebral Cortex* 15 (12): 1848–54.

- Bearden, Carrie E., Michelle Woogen, and David C. Glahn. 2010. "Neurocognitive and Neuroimaging Predictors of Clinical Outcome in Bipolar Disorder." *Current Psychiatry Reports* 12 (6): 499–504.
- Bellani, Marcella, Ping-Hong Yeh, Michele Tansella, Matteo Balestrieri, Jair C. Soares, and Paolo Brambilla. 2009. "DTI Studies of Corpus Callosum in Bipolar Disorder." *Biochemical Society Transactions*. <https://doi.org/10.1042/bst0371096>.
- Benazzi, F. 1999. "Gender Differences in Bipolar II and Unipolar Depressed Outpatients: A 557-Case Study." *Annals of Clinical Psychiatry: Official Journal of the American Academy of Clinical Psychiatrists* 11 (2): 55–59.
- Benazzi, Franco. 2003. "Bipolar II Disorder and Major Depressive Disorder: Continuity or Discontinuity?" *The World Journal of Biological Psychiatry*. <https://doi.org/10.1080/15622970310029914>.
- "Bipolar Disorder." 2016. *The Lancet* 387 (10027): 1561–72.
- Boris Birmaher, MD, David Axelson, MD, Michael Strober, PhD, Mary Kay Gill, MSN, Sylvia Valeri, PhD, Laurel Chiappetta, MS, Neal Ryan, MD, Henrietta Leonard, MD, Jeffrey Hunt, MD, Satish Iyengar, PhD, and Martin Keller, MD. 2006. "Clinical Course of Children and Adolescents with Bipolar Spectrum Disorders." *Archives of General Psychiatry* 63 (2): 175–83.
- Brambilla, Paolo, Mark A. Nicoletti, Roberto B. Sassi, Alan G. Mallinger, Ellen Frank, David J. Kupfer, Matcheri S. Keshavan, and Jair C. Soares. 2003. "Magnetic Resonance Imaging Study of Corpus Callosum Abnormalities in Patients with Bipolar Disorder." *Biological Psychiatry*. [https://doi.org/10.1016/s0006-3223\(03\)00070-2](https://doi.org/10.1016/s0006-3223(03)00070-2).
- Brouwer, Rachel M., René C. W. Mandl, Hugo G. Schnack, Inge L. C. van Soelen, G. Caroline van Baal, Jiska S. Peper, René S. Kahn, Dorret I. Boomsma, and H. E. Hulshoff Pol. 2012.

- “White Matter Development in Early Puberty: A Longitudinal Volumetric and Diffusion Tensor Imaging Twin Study.” *PloS One* 7 (4): e32316.
- Bzdok, Danilo, and Andreas Meyer-Lindenberg. 2018. “Machine Learning for Precision Psychiatry: Opportunities and Challenges.” *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging* 3 (3): 223–30.
- Caetano, S. C., Silveira, C. M., Kaur, S., Nicoletti, M., Hatch, J. P., Brambilla, P., Sassi, R., Axelson, D., Keshavan, M. S., Ryan, N. D., Birmaher, B., & Soares, J. C. 2008. “Abnormal Corpus Callosum Myelination in Pediatric Bipolar Patients.” *Journal of Affective Disorders* 108 (3): 297–301.
- Cao, Bo, Benson Mwangi, Khader M. Hasan, Sudhakar Selvaraj, Cristian P. Zeni, Giovana B. Zunta-Soares, and Jair C. Soares. 2015. “Development and Validation of a Brain Maturation Index Using Longitudinal Neuroanatomical Scans.” *NeuroImage* 117 (August): 311–18.
- Cao, Bo, Benson Mwangi, Ives Cavalcante Passos, Mon-Ju Wu, Zafer Keser, Giovana B. Zunta-Soares, Dianping Xu, Khader M. Hasan, and Jair C. Soares. 2017. “Lifespan Gyrfication Trajectories of Human Brain in Healthy Individuals and Patients with Major Psychiatric Disorders.” *Scientific Reports* 7 (1): 511.
- Cascio, Carissa J., Guido Gerig, and Joseph Piven. 2007. “Diffusion Tensor Imaging: Application to the Study of the Developing Brain.” *Journal of the American Academy of Child and Adolescent Psychiatry* 46 (2): 213–23.
- Catani, Marco, Flavio Dell’acqua, and Michel Thiebaut de Schotten. 2013. “A Revised Limbic System Model for Memory, Emotion and Behaviour.” *Neuroscience and Biobehavioral Reviews* 37 (8): 1724–37.
- “Changes in White Matter Microstructure in the Developing brain—A Longitudinal Diffusion

- Tensor Imaging Study of Children from 4 to 11 Years of Age.” 2016. *NeuroImage* 124 (January): 473–86.
- Chekroud, Adam Mourad, Ryan Joseph Zotti, Zarrar Shehzad, Ralitza Gueorguieva, Marcia K. Johnson, Madhukar H. Trivedi, Tyrone D. Cannon, John Harrison Krystal, and Philip Robert Corlett. 2016. “Cross-Trial Prediction of Treatment Outcome in Depression: A Machine Learning Approach.” *The Lancet. Psychiatry* 3 (3): 243–50.
- Chen, Ying, Sumayya J. Almarzouqi, Michael L. Morgan, and Andrew G. Lee. 2015. “T1-Weighted Image.” *Encyclopedia of Ophthalmology*. https://doi.org/10.1007/978-3-642-35951-4_1228-1.
- Clementz, Brett A., John A. Sweeney, Jordan P. Hamm, Elena I. Ivleva, Lauren E. Ethridge, Godfrey D. Pearlson, Matcheri S. Keshavan, and Carol A. Tamminga. 2016. “Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers.” *The American Journal of Psychiatry* 173 (4): 373–84.
- Courchesne, E., H. J. Chisum, J. Townsend, A. Cowles, J. Covington, B. Egaas, M. Harwood, S. Hinds, and G. A. Press. 2000. “Normal Brain Development and Aging: Quantitative Analysis in Vivo MR Imaging in Healthy Volunteers.” *Radiology* 216 (3): 672–82.
- Dacquino, Claudia, Pietro De Rossi, and Gianfranco Spalletta. 2015. “Schizophrenia and Bipolar Disorder: The Road from Similarities and Clinical Heterogeneity to Neurobiological Types.” *Clinica Chimica Acta; International Journal of Clinical Chemistry* 449 (September): 49–59.
- Dickstein, Daniel P., and Ellen Leibenluft. 2006. “Emotion Regulation in Children and Adolescents: Boundaries between Normalcy and Bipolar Disorder.” *Development and Psychopathology* 18 (4): 1105–31.
- Dietrichs, E. 1984. “Cerebellar Autonomic Function: Direct Hypothalamo Cerebellar Pathway.”

Science 223 (4636): 591–93.

- Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, Fetcho RN, Zebley B, Oathes DJ, Etkin A, Schatzberg AF, Sudheimer K, Keller J, Mayberg HS, Gunning FM, Alexopoulos GS, Fox MD, Pascual-Leone A, Voss HU, Casey BJ, Dubin MJ, Liston C. 2017. “Resting-State Connectivity Biomarkers Define Neurophysiological Subtypes of Depression.” *Nature Medicine* 23 (1): 28–38.
- Duax, Jeanne M., and Eric A. Youngstrom. 2007. “Sex Differences in Pediatric Bipolar Disorder.” *The Journal of Clinical Psychiatry*. <https://doi.org/10.4088/jcp.v68n1016>.
- Fair, Damien A., Deepti Bathula, Molly A. Nikolas, and Joel T. Nigg. 2012. “Distinct Neuropsychological Subgroups in Typically Developing Youth Inform Heterogeneity in Children with ADHD.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (17): 6769–74.
- Fischl, B., and A. M. Dale. 2000. “Measuring the Thickness of the Human Cerebral Cortex from Magnetic Resonance Images.” *Proceedings of the National Academy of Sciences of the United States of America* 97 (20): 11050–55.
- Bruce Fischl, André van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H. Salat, Evelina Busa, Larry J. Seidman, Jill Goldstein, David Kennedy, Verne Caviness, Nikos Makris, Bruce Rosen, Anders M. Dale. 2004. “Automatically Parcellating the Human Cerebral Cortex.” *Cerebral Cortex* 14 (1): 11–22.
- Giedd, Jay N., and Judith L. Rapoport. 2010. “Structural MRI of Pediatric Brain Development: What Have We Learned and Where Are We Going?” *Neuron* 67 (5): 728–34.
- Gillan, Claire M., Claire M. . Gillan, and Robert Whelan. 2017. “What Big Data Can Do for Treatment in Psychiatry.” *Current Opinion in Behavioral Sciences*.
<https://doi.org/10.1016/j.cobeha.2017.07.003>.

- Glass, Richard M. 2009. "Diagnostic and Statistical Manual of Mental Disorders (DSM)." *AMA Manual of Style*. <https://doi.org/10.1093/jama/9780195176339.022.529>.
- Goswami, Utpal, Aditya Sharma, Udayan Khastigir, Ian Nicol Ferrier, Allan H. Young, Peter Gallagher, Jill M. Thompson, and P. Brian Moore. 2006. "Neuropsychological Dysfunction, Soft Neurological Signs and Social Disability in Euthymic Patients with Bipolar Disorder." *The British Journal of Psychiatry: The Journal of Mental Science* 188 (April): 366–73.
- Grimm, Oliver, Sebastian Pohlack, Raffaele Cacciaglia, Tobias Winkelmann, Michael M. Plichta, Traute Demiralp, and Herta Flor. 2015. "Amygdalar and Hippocampal Volume: A Comparison between Manual Segmentation, Freesurfer and VBM." *Journal of Neuroscience Methods* 253 (September): 254–61.
- Groeschel, S., B. Vollmer, M. D. King, and A. Connelly. 2010. "Developmental Changes in Cerebral Grey and White Matter Volume from Infancy to Adulthood." *International Journal of Developmental Neuroscience: The Official Journal of the International Society for Developmental Neuroscience* 28 (6): 481–89.
- Haldane, Morgan, Giles Cunningham, Chris Androustos, and Sophia Frangou. 2008. "Structural Brain Correlates of Response Inhibition in Bipolar Disorder I." *Journal of Psychopharmacology* 22 (2): 138–43.
- Hasan, Khader M., Benson Mwangi, Bo Cao, Zafer Keser, Nicholas J. Tustison, Peter Kochunov, Richard E. Frye, Mirjana Savatic, and Jair Soares. 2016. "Entorhinal Cortex Thickness across the Human Lifespan." *Journal of Neuroimaging: Official Journal of the American Society of Neuroimaging* 26 (3): 278–82.
- Heng, Serene, Allen W. Song, and Kang Sim. 2010. "White Matter Abnormalities in Bipolar Disorder: Insights from Diffusion Tensor Imaging Studies." *Journal of Neural*

Transmission 117 (5): 639–54.

Ivleva, E. I., Clementz, B. A., Dutcher, A. M., Arnold, S. J. M., Jeon-Slaughter, H., Aslan, S., Witte, B., Poudyal, G., Lu, H., Meda, S. A., Pearlson, G. D., Sweeney, J. A., Keshavan, M. S., & Tamminga, C. A. 2017. “Brain Structure Biomarkers in the Psychosis Biotypes: Findings From the Bipolar-Schizophrenia Network for Intermediate Phenotypes.” *Biological Psychiatry* 82 (1): 26–39.

Janssen J, Alemán-Gómez Y, Schnack H, Balaban E, Pina-Camacho L, Alfaro-Almagro F, Castro-Fornieles J, Otero S, Baeza I, Moreno D, Bargalló N, Parellada M, Arango C, Desco M . 2014. “Cortical Morphology of Adolescents with Bipolar Disorder and with Schizophrenia.” *Schizophrenia Research*. <https://doi.org/10.1016/j.schres.2014.06.040>.

Juranek, Jenifer, Pauline A. Filipek, Gholam R. Berenji, Charlotte Modahl, Kathryn Osann, and M. Anne Spence. 2006. “Association between Amygdala Volume and Anxiety Level: Magnetic Resonance Imaging (MRI) Study in Autistic Children.” *Journal of Child Neurology* 21 (12): 1051–58.

Juranek, Jenifer, Chad P. Johnson, Mary R. Prasad, Larry A. Kramer, Ann Saunders, Pauline A. Filipek, Paul R. Swank, Charles S. Cox Jr, and Linda Ewing-Cobbs. 2012. “Mean Diffusivity in the Amygdala Correlates with Anxiety in Pediatric TBI.” *Brain Imaging and Behavior* 6 (1): 36–48.

Kempton, Matthew J. 2011. “Structural Neuroimaging Studies in Major Depressive Disorder.” *Archives of General Psychiatry*. <https://doi.org/10.1001/archgenpsychiatry.2011.60>.

Kerbeshian, J., and L. Burd. 1996. “Case Study: Comorbidity among Tourette’s Syndrome, Autistic Disorder, and Bipolar Disorder.” *Journal of the American Academy of Child and Adolescent Psychiatry* 35 (5): 681–85.

Khundrakpam, Budhachandra S., Jussi Tohka, Alan C. Evans, and Brain Development

- Cooperative Group. 2015. "Prediction of Brain Maturity Based on Cortical Thickness at Different Spatial Resolutions." *NeuroImage* 111 (May): 350–59.
- Kohli, Jiwandeep S., Mikaela K. Kinnear, Christopher H. Fong, Inna Fishman, Ruth A. Carper, and Ralph-Axel Müller. 2019. "Local Cortical Gyrification Is Increased in Children With Autism Spectrum Disorders, but Decreases Rapidly in Adolescents." *Cerebral Cortex* 29 (6): 2412–23.
- Lebel, Catherine, and Christian Beaulieu. 2011. "Longitudinal Development of Human Brain Wiring Continues from Childhood into Adulthood." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31 (30): 10937–47.
- Leibenluft, Ellen, Dennis S. Charney, and Daniel S. Pine. 2003. "Researching the Pathophysiology of Pediatric Bipolar Disorder." *Biological Psychiatry* 53 (11): 1009–20.
- Lenroot, Rhoshel K., and Jay N. Giedd. 2006. "Brain Development in Children and Adolescents: Insights from Anatomical Magnetic Resonance Imaging." *Neuroscience and Biobehavioral Reviews* 30 (6): 718–29.
- Lenroot, R. K., Gogtay, N., Greenstein, D. K., Wells, E. M., Wallace, G. L., Clasen, L. S., Blumenthal, J. D., Lerch, J., Zijdenbos, A. P., Evans, A. C., Thompson, P. M., & Giedd, J. N. 2007. "Sexual Dimorphism of Brain Developmental Trajectories during Childhood and Adolescence." *NeuroImage* 36 (4): 1065–73.
- Leow, A., Ajilore, O., Zhan, L., Arienzo, D., GadElkarim, J., Zhang, A., Moody, T., Van Horn, J., Feusner, J., Kumar, A., Thompson, P., & Altshuler, L. 2013. "Impaired Inter-Hemispheric Integration in Bipolar Disorder Revealed with Brain Network Analyses." *Biological Psychiatry* 73 (2): 183–93.
- Leverich GS, Post RM, Keck PE Jr, Altshuler LL, Frye MA, Kupka RW, Nolen WA, Suppes T, McElroy SL, Grunze H, Denicoff K, Moravec MK, Luckenbaugh D. 2007. "The Poor

Prognosis of Childhood-Onset Bipolar Disorder.” *The Journal of Pediatrics* 150 (5): 485–90.

Lima, Isabela M. M., Andrew D. Peckham, and Sheri L. Johnson. 2018. “Cognitive Deficits in Bipolar Disorders: Implications for Emotion.” *Clinical Psychology Review* 59 (February): 126–36.

Lin, Fuchun, Shenhong Weng, Baojun Xie, Guangyao Wu, and Hao Lei. 2011. “Abnormal Frontal Cortex White Matter Connections in Bipolar Disorder: A DTI Tractography Study.” *Journal of Affective Disorders* 131 (1-3): 299–306.

Lin, Ping-I, Melvin G. McInnis, James B. Potash, Virginia Willour, Dean F. MacKinnon, J. Raymond DePaulo, and Peter P. Zandi. 2006. “Clinical Correlates and Familial Aggregation of Age at Onset in Bipolar Disorder.” *The American Journal of Psychiatry* 163 (2): 240–46.

“Long-Term Stability of Polarity Distinctions in the Affective Disorders.” 1995. *American Journal of Psychiatry*. <https://doi.org/10.1176/ajp.152.3.385>.

López-Larson, Melissa P., Melissa P. DelBello, Molly E. Zimmerman, Michael L. Schwiers, and Stephen M. Strakowski. 2002. “Regional Prefrontal Gray and White Matter Abnormalities in Bipolar Disorder.” *Biological Psychiatry* 52 (2): 93–100.

Marquand, Andre F., Ieab Rezek, Jan Buitelaar, and Christian F. Beckmann. 2016. “Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies.” *Biological Psychiatry* 80 (7): 552–61.

Marquand, Andre F., Thomas Wolfers, Maarten Mennes, Jan Buitelaar, and Christian F. Beckmann. 2016. “Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders.” *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging* 1 (5): 433–47.

- McCormack C, Green MJ, Rowland JE, Roberts G, Frankland A, Hadzi-Pavlovic D, Joslyn C, Lau P, Wright A, Levy F, Lenroot RK, Mitchell PB. 2016. “Neuropsychological and Social Cognitive Function in Young People at Genetic Risk of Bipolar Disorder.” *Psychological Medicine* 46 (4): 745–58.
- McDonald, Gary C. 2009. “Ridge Regression.” *Wiley Interdisciplinary Reviews: Computational Statistics* 1 (1): 93–100.
- “Meta-Analysis of Amygdala Volumes in Children and Adolescents With Bipolar Disorder.” 2008. *Journal of the American Academy of Child and Adolescent Psychiatry* 47 (11): 1289–98.
- Minckwitz G von, Untch M, Nüesch E, Loibl S, Kaufmann M, Kümmel S, Fasching PA, Eiermann W, Blohmer JU, Costa SD, Mehta K, Hilfrich J, Jackisch C, Gerber B, du Bois A, Huober J, Hanusch C, Konecny G, Fett W, Stickeler E, Harbeck N, Müller V, Jüni P. 2011. “Impact of Treatment Characteristics on Response of Different Breast Cancer Phenotypes: Pooled Analysis of the German Neo-Adjuvant Chemotherapy Trials.” *Breast Cancer Research and Treatment* 125 (1): 145–56.
- Mirakhur, Ajay, T. William J. Moorhead, Andrew C. Stanfield, James McKirdy, Jessika E. D. Sussmann, Jeremy Hall, Stephen M. Lawrie, Eve C. Johnstone, and Andrew M. McIntosh. 2009. “Changes in Gyrification over 4 Years in Bipolar Disorder and Their Association with the Brain-Derived Neurotrophic Factor valine(66) Methionine Variant.” *Biological Psychiatry* 66 (3): 293–97.
- Munesue, T., Y. Ono, K. Mutoh, K. Shimoda, H. Nakatani, and M. Kikuchi. 2008. “High Prevalence of Bipolar Disorder Comorbidity in Adolescents and Young Adults with High-Functioning Autism Spectrum Disorder: A Preliminary Study of 44 Outpatients.” *Journal of Affective Disorders*. <https://doi.org/10.1016/j.jad.2008.02.015>.

- Mwangi, Benson, Tian Siva Tian, and Jair C. Soares. 2014. "A Review of Feature Reduction Techniques in Neuroimaging." *Neuroinformatics* 12 (2): 229–44.
- Mwangi, Benson, Mon-Ju Wu, Isabelle E. Bauer, Haina Modi, Cristian P. Zeni, Giovana B. Zunta-Soares, Khader M. Hasan, and Jair C. Soares. 2015. "Predictive Classification of Pediatric Bipolar Disorder Using Atlas-Based Diffusion Weighted Imaging and Support Vector Machines." *Psychiatry Research* 234 (2): 265–71.
- Nanda, Pranav, Neeraj Tandon, Ian T. Mathew, Christoforos I. Giakoumatos, Hulegar A. Abhishekh, Brett A. Clementz, Godfrey D. Pearlson, John Sweeney, Carol A. Tamminga, and Matcheri S. Keshavan. 2014. "Local Gyrfication Index in Probands with Psychotic Disorders and Their First-Degree Relatives." *Biological Psychiatry* 76 (6): 447–55.
- Narum, Shawn R. 2006. "Beyond Bonferroni: Less Conservative Analyses for Conservation Genetics." *Conservation Genetics*. <https://doi.org/10.1007/s10592-006-9189-7>.
- Nenadic, Igor, Raka Maitra, Maren Dietzek, Kerstin Langbein, Stefan Smesny, Heinrich Sauer, and Christian Gaser. 2015. "Prefrontal Gyrfication in Psychotic Bipolar I Disorder vs. Schizophrenia." *Journal of Affective Disorders* 185 (October): 104–7.
- Neuner, Irene, Yuliya Kupriyanova, Tony Stöcker, Ruiwang Huang, Oleg Posnansky, Frank Schneider, and N. Jon Shah. 2011. "Microstructure Assessment of Grey Matter Nuclei in Adult Tourette Patients by Diffusion Tensor Imaging." *Neuroscience Letters* 487 (1): 22–26.
- Orrù, Graziella, William Pettersson-Yeo, Andre F. Marquand, Giuseppe Sartori, and Andrea Mechelli. 2012. "Using Support Vector Machine to Identify Imaging Biomarkers of Neurological and Psychiatric Disease: A Critical Review." *Neuroscience and Biobehavioral Reviews* 36 (4): 1140–52.
- Pagel, Mark D., and Clifford E. Lunneborg. 1985. "Empirical Evaluation of Ridge Regression."

Psychological Bulletin. <https://doi.org/10.1037/0033-2909.97.2.342>.

Palaniyappan, Lena, and Peter F. Liddle. 2014. "Diagnostic Discontinuity in Psychosis: A Combined Study of Cortical Gyrification and Functional Connectivity." *Schizophrenia Bulletin* 40 (3): 675–84.

Parker, Gordon, Kathryn Fletcher, Amelia Paterson, Josephine Anderson, and Michael Hong. 2014. "Gender Differences in Depression Severity and Symptoms across Depressive Sub-Types." *Journal of Affective Disorders* 167 (June): 351–57.

Paus, Tomáš. 2005. "Mapping Brain Maturation and Cognitive Development during Adolescence." *Trends in Cognitive Sciences*. <https://doi.org/10.1016/j.tics.2004.12.008>.

Perlaki, Gabor, Reka Horvath, Szilvia Anett Nagy, Peter Bogner, Tamas Doczi, Jozsef Janszky, and Gergely Orsi. 2017. "Comparison of Accuracy between FSL's FIRST and Freesurfer for Caudate Nucleus and Putamen Segmentation." *Scientific Reports*. <https://doi.org/10.1038/s41598-017-02584-5>.

Pierpaoli, C., P. Jezzard, P. J. Basser, A. Barnett, and G. Di Chiro. 1996. "Diffusion Tensor MR Imaging of the Human Brain." *Radiology*. <https://doi.org/10.1148/radiology.201.3.8939209>.

Pintzka, Carl W. S., Tor I. Hansen, Hallvard R. Evensmoen, and Asta K. Håberg. 2015. "Marked Effects of Intracranial Volume Correction Methods on Sex Differences in Neuroanatomical Structures: A HUNT MRI Study." *Frontiers in Neuroscience* 9 (July): 238.

Poletti, Sara, Irene Bollettini, Elena Mazza, Clara Locatelli, Daniele Radaelli, Benedetta Vai, Enrico Smeraldi, Cristina Colombo, and Francesco Benedetti. 2015. "Cognitive Performances Associate with Measures of White Matter Integrity in Bipolar Disorder." *Journal of Affective Disorders* 174 (March): 342–52.

- Post RM, Leverich GS, Kupka RW, Keck PE Jr, McElroy SL, Altshuler LL, Frye MA, Luckenbaugh DA, Rowe M, Grunze H, Suppes T, Nolen WA. 2010. "Early-Onset Bipolar Disorder and Treatment Delay Are Risk Factors for Poor Outcome in Adulthood." *The Journal of Clinical Psychiatry* 71 (7): 864–72.
- Qiu, Deqiang, Li-Hai Tan, Ke Zhou, and Pek-Lan Khong. 2008. "Diffusion Tensor Imaging of Normal White Matter Maturation from Late Childhood to Young Adulthood: Voxel-Wise Evaluation of Mean Diffusivity, Fractional Anisotropy, Radial and Axial Diffusivities, and Correlation with Reading Development." *NeuroImage* 41 (2): 223–32.
- Reddan, Marianne C., Martin A. Lindquist, and Tor D. Wager. 2017. "Effect Size Estimation in Neuroimaging." *JAMA Psychiatry* 74 (3): 207–8.
- Rimol LM, Hartberg CB, Nesvåg R, Fennema-Notestine C, Hagler DJ Jr, Pung CJ, Jennings RG, Haukvik UK, Lange E, Nakstad PH, Melle I, Andreassen OA, Dale AM, Agartz I. 2010. "Cortical Thickness and Subcortical Volumes in Schizophrenia and Bipolar Disorder." *Biological Psychiatry* 68 (1): 41–50.
- Rimol LM¹, Nesvåg R, Hagler DJ Jr, Bergmann O, Fennema-Notestine C, Hartberg CB, Haukvik UK, Lange E, Pung CJ, Server A, Melle I, Andreassen OA, Agartz I, Dale AM. 2012. "Cortical Volume, Surface Area, and Thickness in Schizophrenia and Bipolar Disorder." *Biological Psychiatry* 71 (6): 552–60.
- Robinson, Lucy J., Jill M. Thompson, Peter Gallagher, Utpal Goswami, Allan H. Young, I. Nicol Ferrier, and P. Brian Moore. 2006. "A Meta-Analysis of Cognitive Deficits in Euthymic Patients with Bipolar Disorder." *Journal of Affective Disorders* 93 (1-3): 105–15.
- Roybal, Donna J., Manpreet K. Singh, Victoria E. Cosgrove, Meghan Howe, Ryan Kelley, Naama Barnea-Goraly, and Kiki D. Chang. 2012. "Biological Evidence for a Neurodevelopmental Model of Pediatric Bipolar Disorder." *The Israel Journal of*

Psychiatry and Related Sciences 49 (1): 28–43.

Sanfilipo, Michael P., Ralph H. B. Benedict, Robert Zivadinov, and Rohit Bakshi. 2004.

“Correction for Intracranial Volume in Analysis of Whole Brain Atrophy in Multiple Sclerosis: The Proportion vs. Residual Method.” *NeuroImage* 22 (4): 1732–43.

Sarnicola, Antonio, Matthew Kempton, Cristina Germanà, Morgan Haldane, Michael Hadjulis,

Tessa Christodoulou, Athanasios Koukopoulos, Paolo Girardi, Roberto Tatarelli, and

Sophia Frangou. 2009. “No Differential Effect of Age on Brain Matter Volume and

Cognition in Bipolar Patients and Healthy Individuals.” *Bipolar Disorders* 11 (3): 316–22.

Satopaa, Ville, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011. “Finding a

‘Kneedle’ in a Haystack: Detecting Knee Points in System Behavior.” *2011 31st*

International Conference on Distributed Computing Systems Workshops.

<https://doi.org/10.1109/icdcsw.2011.20>.

Saunders, Erika F. H., Kate D. Fitzgerald, Peng Zhang, and Melvin G. McInnis. 2012. “Clinical

Features of Bipolar Disorder Comorbid with Anxiety Disorders Differ between Men and

Women.” *Depression and Anxiety* 29 (8): 739–46.

Schaer, Marie, Meritxell Bach Cuadra, Nick Schmansky, Bruce Fischl, Jean-Philippe Thiran,

and Stephan Eliez. 2012. “How to Measure Cortical Folding from MR Images: A Step-by-

Step Tutorial to Compute Local Gyrfication Index.” *Journal of Visualized Experiments*:

JoVE, no. 59 (January): e3417.

Schulenberg, John E., Arnold J. Sameroff, and Dante Cicchetti. 2004. “The Transition to

Adulthood as a Critical Juncture in the Course of Psychopathology and Mental Health.”

Development and Psychopathology 16 (4): 799–806.

Schumann, C. M. 2004. “The Amygdala Is Enlarged in Children But Not Adolescents with

Autism; the Hippocampus Is Enlarged at All Ages.” *Journal of Neuroscience*.

<https://doi.org/10.1523/jneurosci.1297-04.2004>.

Schumann, Cynthia Mills, Cynthia Carter Barnes, Catherine Lord, and Eric Courchesne. 2009.

“Amygdala Enlargement in Toddlers with Autism Related to Severity of Social and Communication Impairments.” *Biological Psychiatry* 66 (10): 942–49.

Senni M, Paulus WJ, Gavazzi A, Fraser AG, Díez J, Solomon SD, Smiseth OA, Guazzi M, Lam

CS, Maggioni AP, Tschöpe C, Metra M, Hummel SL, Edelmann F, Ambrosio G, Stewart

Coats AJ, Filippatos GS, Gheorghiade M, Anker SD, Levy D, Pfeffer MA, Stough WG,

Pieske BM. 2014. “New Strategies for Heart Failure with Preserved Ejection Fraction: The

Importance of Targeted Therapies for Heart Failure Phenotypes.” *European Heart Journal*

35 (40): 2797–2815.

Seymour, Karen E., Kerri L. Kim, Grace K. Cushman, Megan E. Puzia, Alexandra B.

Weissman, Thania Galvan, and Daniel P. Dickstein. 2015. “Affective Processing Bias in

Youth with Primary Bipolar Disorder or Primary Attention-Deficit/hyperactivity Disorder.”

European Child & Adolescent Psychiatry. <https://doi.org/10.1007/s00787-015-0686-4>.

Sgouros, S., J. H. Goldin, A. D. Hockley, M. J. Wake, and K. Natarajan. 1999. “Intracranial

Volume Change in Childhood.” *Journal of Neurosurgery* 91 (4): 610–16.

Shaw, P., D. Greenstein, J. Lerch, L. Clasen, R. Lenroot, N. Gogtay, A. Evans, J. Rapoport, and

J. Giedd. 2006. “Intellectual Ability and Cortical Development in Children and

Adolescents.” *Nature* 440 (7084): 676–79.

Song, Ming, Zhengyi Yang, Jing Sui, and Tianzi Jiang. 2017. “Biological Subtypes Bridge

Diagnoses and Biomarkers: A Novel Research Track for Mental Disorders.” *Neuroscience*

Bulletin 33 (3): 351–53.

Song, Sheng-Kwei, Shu-Wei Sun, Won-Kyu Ju, Shiow-Juan Lin, Anne H. Cross, and Arthur H.

Neufeld. 2003. “Diffusion Tensor Imaging Detects and Differentiates Axon and Myelin

- Degeneration in Mouse Optic Nerve after Retinal Ischemia.” *NeuroImage* 20 (3): 1714–22.
- Song, Sheng-Kwei, Shu-Wei Sun, Michael J. Ramsbottom, Chen Chang, John Russell, and Anne H. Cross. 2002. “Dysmyelination Revealed through MRI as Increased Radial (but Unchanged Axial) Diffusion of Water.” *NeuroImage* 17 (3): 1429–36.
- Sowell, Elizabeth R., Doris A. Trauner, Anthony Gamst, and Terry L. Jernigan. 2002. “Development of Cortical and Subcortical Brain Structures in Childhood and Adolescence: A Structural MRI Study.” *Developmental Medicine and Child Neurology* 44 (1): 4–16.
- Squeglia, Lindsay M., Joanna Jacobus, Scott F. Sorg, Terry L. Jernigan, and Susan F. Tapert. 2013. “Early Adolescent Cortical Thinning Is Related to Better Neuropsychological Performance.” *Journal of the International Neuropsychological Society: JINS* 19 (9): 962–70.
- Storsve, Andreas B., Anders M. Fjell, Christian K. Tamnes, Lars T. Westlye, Knut Overbye, Hilde W. Aasland, and Kristine B. Walhovd. 2014. “Differential Longitudinal Changes in Cortical Thickness, Surface Area and Volume across the Adult Life Span: Regions of Accelerating and Decelerating Change.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34 (25): 8488–98.
- Tesli M, Kauppi K, Bettella F, Brandt CL, Kaufmann T, Espeseth T, Mattingdal M, Agartz I, Melle I, Djurovic S, Westlye LT, Andreassen OA. 2015. “Altered Brain Activation during Emotional Face Processing in Relation to Both Diagnosis and Polygenic Risk of Bipolar Disorder.” *PloS One* 10 (7): e0134202.
- Tesli, M., T. Espeseth, F. Bettella, M. Mattingdal, M. Aas, I. Melle, S. Djurovic, and O. A. Andreassen. 2014. “Polygenic Risk Score and the Psychosis Continuum Model.” *Acta Psychiatrica Scandinavica* 130 (4): 311–17.
- Ulay, Halime Tuna, and Aygün Ertuğrul. 2009. “[Neuroimaging findings in autism: a brief

- review].” *Türk psikiyatri dergisi = Turkish journal of psychiatry* 20 (2): 164–74.
- Usher, Juliana, Stefan Leucht, Peter Falkai, and Harald Scherk. 2010. “Correlation between Amygdala Volume and Age in Bipolar Disorder — A Systematic Review and Meta-Analysis of Structural MRI Studies.” *Psychiatry Research: Neuroimaging*.
<https://doi.org/10.1016/j.psychresns.2009.09.004>.
- Van Rheenen TE, Lewandowski KE, Tan EJ, Ospina LH, Ongur D, Neill E, Gurvich C, Pantelis C, Malhotra AK, Rossell SL, Burdick KE. 2017. “Characterizing Cognitive Heterogeneity on the Schizophrenia–bipolar Disorder Spectrum.” *Psychological Medicine*.
<https://doi.org/10.1017/s0033291717000307>.
- Vapnik, Vladimir N. 2000. “The Nature of Statistical Learning Theory.”
<https://doi.org/10.1007/978-1-4757-3264-1>.
- Vértes, Petra E., and Edward T. Bullmore. 2015. “Annual Research Review: Growth Connectomics--the Organization and Reorganization of Brain Networks during Normal and Abnormal Development.” *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 56 (3): 299–320.
- Vijayakumar, Nandita, Sarah Whittle, Murat Yücel, Meg Dennison, Julian Simmons, and Nicholas B. Allen. 2014. “Thinning of the Lateral Prefrontal Cortex during Adolescence Predicts Emotion Regulation in Females.” *Social Cognitive and Affective Neuroscience* 9 (11): 1845–54.
- Voevodskaya, Olga, Andrew Simmons, Richard Nordenskjöld, Joel Kullberg, Håkan Ahlström, Lars Lind, Lars-Olof Wahlund, Elna-Marie Larsson, Eric Westman, and Alzheimer’s Disease Neuroimaging Initiative. 2014. “The Effects of Intracranial Volume Adjustment Approaches on Multiple Regional MRI Volumes in Healthy Aging and Alzheimer’s Disease.” *Frontiers in Aging Neuroscience* 6 (October): 264.

- Wallace, Gregory L., Briana Robustelli, Nathan Dankner, Lauren Kenworthy, Jay N. Giedd, and Alex Martin. 2013. "Increased Gyrfication, but Comparable Surface Area in Adolescents with Autism Spectrum Disorders." *Brain*. <https://doi.org/10.1093/brain/awt106>.
- Wang, Lubin, Longfei Su, Hui Shen, and Dewen Hu. 2012. "Decoding Lifespan Changes of the Human Brain Using Resting-State Functional Connectivity MRI." *PloS One* 7 (8): e44530.
- Wierenga, Lara M., Marieke Langen, Bob Oranje, and Sarah Durston. 2014. "Unique Developmental Trajectories of Cortical Thickness and Surface Area." *NeuroImage* 87 (February): 120–26.
- Wolfers, T., Doan, N. T., Kaufmann, T., Alnæs, D., Moberget, T., Agartz, I., Buitelaar, J. K., Ueland, T., Melle, I., Franke, B., Andreassen, O. A., Beckmann, C. F., Westlye, L. T., & Marquand, A. F. 2018. "Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models." *JAMA Psychiatry* 75 (11): 1146–55.
- Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. 2011. "The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity." *Journal of Neurophysiology* 106 (3): 1125–65.
- Zhang, W., Xiao, Y., Sun, H., Rodrigo Patino, L., Tallman, M. J., Weber, W. A., Adler, C. M., Klein, C., Strawn, J. R., Nery, F. G., Gong, Q., Sweeney, J. A., Lui, S., & DelBello, M. P. . 2018. "Discrete Patterns of Cortical Thickness in Youth with Bipolar Disorder Differentially Predict Treatment Response to Quetiapine but Not Lithium." *Neuropsychopharmacology*. <https://doi.org/10.1038/s41386-018-0120-y>.

Vita

Jonika Tannous was born in Byblos, Lebanon, the daughter of Gilberte Tannous and Doumit Tannous. After completing her work at Dulles High School, Sugar Land, Texas in 2010, she entered Rice University in Houston, Texas. She received the degree of Bachelor of Arts with a double major in Cognitive Science and Psychology from Rice in May, 2014. In August of 2015 she entered The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences.

Permanent address:

214 Greenwood St

Houston, Texas 77011