

Global-To-Local Segmentation and Genotypic Analysis Of Brain Shape Asymmetry

Vasileios Lemonidis

Thesis submitted for the degree of
Master of Science in Bioinformatics

Thesis supervisors:

Prof. Peter Claes
Prof. Isabelle Cleynen

Mentors:

MSc. Meng Yuan
MSc. Seppe Goovaerts

© Copyright KU Leuven

Without written permission of the thesis supervisors and the author it is forbidden to reproduce or adapt in any form or by any means any part of this publication. Requests for obtaining the right to reproduce or utilize parts of this publication should be addressed to Faculteit Ingenieurswetenschappen, Kasteelpark Arenberg 1 bus 2200, B-3001 Heverlee, +32-16-321350.

A written permission of the thesis supervisors is also required to use the methods, products, schematics and programmes described in this work for industrial or commercial use, and for submitting this publication in scientific contests.

Preface

The text of the preface. A few paragraphs should follow.

The Author
1 January 2010

Contents

Preface	i
List of Figures	iii
List of Tables	iii
List of Abbreviations	v
Abstract	vi
1 Introduction	1
2 Dataset Description	5
2.1 Phenotype	5
2.2 Genome	5
3 Data Preprocessing	7
4 Asymmetry Phenotypic Analysis	9
4.1 Statistical Analysis of Asymmetry Components	9
4.2 Phenotypic Partitioning	9
5 Asymmetry Genetic Analysis	11
5.1 Post-Processing	11
5.2 Meta-Analysis	11
6 Discussion	13
Bibliography	15

List of Figures

1.1	Illustration of the Yakovlevian torque. Displayed by a red line rotated counterclockwise 0.51 degrees in relation to the perfectly vertical black line, as calculated by using the average angle of the longest edges (in blue) of the convex hulls of the horizontal plane projection of each averaged hemisphere, after a scaling and alignment process, across the studied population.	2
-----	---	---

List of Tables

Abbreviations

CCA canonical correlation analysis

DA directional asymmetry

GWAS genome wide association studies

HSC hierarchical spectral clustering

LD linkage disequilibrium

MRI magnetic resonance imaging

mvGWAS multivariate genome-wide association study

SNP single nucleotide polymorphism

Abstract

Overall purpose of this thesis is to complement the existing bibliography on the detection and examination of the genetic associations of brain shape asymmetry. Asymmetry components are computed based on the brain MRI dataset provided by UK Biobank database. A data-driven approach is followed, where the brain surface is partitioned in an unsupervised manner, through Hierarchical Spectral Clustering, a technique that allows for a coarse-to-fine segmentation. Aggregated asymmetry measurements are retrieved from the segments, whose genetic correlation is examined through a multivariate genome-wide association study (mvGWAS) statistical analysis. Recognized significant single nucleotide polymorphisms (SNPs) are then analyzed individually or in groups, through comparison with existing results and databases. The genetic overlap with neurodevelopmental disorders and traits, that have been reported to exhibit phenotypic associations with brain structure asymmetry, such as Autism, Alzheimer's Disease or intelligence, are examined. Functional annotations of variants associated with the genes where significant SNPs were detected are constructed, offering an insight into the functional reasoning behind the brain shape asymmetry existence. Further comparisons with other past human phenotypic characteristics studies are lastly applied.

Chapter 1

Introduction

Cerebral symmetry is a universal quality of organisms belonging to the Bilateria lineage [4][1]. For the mammals group, the brain is anatomically divided into a left and right hemisphere. Asymmetrical cell division during brain development, initially observed among neuroblasts, causes shape deviations between the two hemispheres, leading to what is called, brain lateralization. This property gives rise to partial functional disassociation, with subsets of tasks requiring differential activation of each hemisphere. Along with the purely genetic reasons, environment also plays a significant role in affecting cerebral bilateral asymmetry. The observed deviations can be statistically formulated as products of two factors, the hemisphere side studied and the individual specimen analyzed, as well as their interaction [9]. Of primary interest in this work is directional asymmetry (DA), the asymmetric component that arises by comparing a single individual's hemispherical surfaces landmarks differences, computed through a process of alignment, reflection and subtraction of the landmarks pairs, as discussed in detail in chapter 4. DA captures information about anatomic characteristics, such as the overall counterclockwise torque, named 'Yakovlevian torque' [11], that is observed in humans between the right and left hemisphere (Figure 1.1). Past studies have shown that abnormal DA may be an indication of certain diseases. The lack of it may imply schizophrenia predisposition [14]. Any significant abnormalities may be indicative of other psychiatric disorders, such as autism or developmental language disorder [7][10].

In this study, targeted on humans, a cross section between the dependent cerebral asymmetry and the independent genetic factors is performed, in an effort to discover affiliated genetic regions and provide a novel understanding of the related genes cooperation. Genome wide association studies (GWAS) have shed light on the correlation between phenotypic traits and genetic content. With the advent of technology capable to collect and process genomes from different individuals in relatively high speed, vast databases have been constructed. One of the main players in the data collection has been UK Biobank; a large-scale database from a randomized consortium of 500,000 individuals, whose genome has been collected, from whom 43,000 subjects had also participated in brain magnetic resonance imaging (MRI) collection process, as of December 2020. In this thesis, we exploit this newly acquired

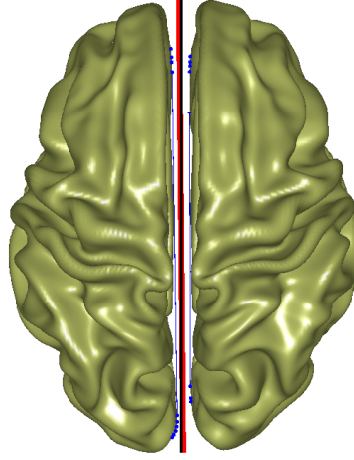


FIGURE 1.1: Illustration of the Yakovlevian torque. Displayed by a red line rotated counterclockwise 0.51 degrees in relation to the perfectly vertical black line, as calculated by using the average angle of the longest edges (in blue) of the convex hulls of the horizontal plane projection of each averaged hemisphere, after a scaling and alignment process, across the studied population.

dataset to identify the key loci that are related to the human brain surface symmetry. Only healthy self-proclaimed white European individuals were considered. The surface analyzed is exclusively the one coined as midthickness, derived as the one half way between the pial and white matter, of the cerebral cortex. More details on the dataset description and the involved data preprocessing can be found in chapters 2 and 3.

The present work evaluates the brain asymmetry genetic landscape in a coarse-to-fine segmentation, through the application of hierarchical spectral clustering (HSC)[13], discussed in section 4.2. The technique has been used in a number of different related studies [3][12], yielding results that are in accordance with the underlying anatomical features. The main reason behind this partitioning is the intrinsic complexity of the studied phenotype, eliciting expected differences in the genomic profiles of each cerebral cortex region. The basic assumption made is that topologically close landmarks share similar genetic background. In general though, this type of distance-based clustering is governed by the least quantity of assumptions, regarding the shape or form of the cluster [17]. The partitions' genetic juxtaposition is valuable for identifying which regions share similar significant genetic loci, highlighting the corresponding genes contribution, or showcasing the specialization of certain regions that share little to no similarities with their neighbors. Identifying the latter provides a way of mapping the developmental activation of each locus, bringing forth the opportunity to augment the results of related developmental studies [16].

The genomic studies are performed under the framework of SNP-by-SNP canonical correlation analysis (CCA). The goal is to incorporate multi-allelic SNPs and, more importantly, multivariate phenotype, in a single hypothesis test per SNP, that is whether the phenotype is significantly correlated with each analyzed SNP. In general, there is an abundance of strategies on how to perform multivariate GWAS, ranging from direct methods, that approximate the inputs relation either in an unbiased manner or making certain educated guesses, to more complex techniques, that increase statistical power by transforming the inputs, at the expense of explanatory ability [6]. There are also methods that are based on the meta-analysis of outcomes from univariate studies, commonly used to juxtapose experiments from separate sources, for which the original data is absent or the exact replication of the study is arduous [2]. Which approach performs best mainly lies on the dataset properties and the nature of the scientific question. Factors such as low sample size [15], genes pleiotropic effects [5] or within-study variation [8] tend to handicap the statistical modeling and increase the type I and II errors of the corresponding hypothesis tests. In this study, CCA was primarily chosen due to the high capacity in efficiently reducing the inputs dimensionality while preserving most information regarding their correlation. A variety of experiments, analyzed in chapter 5, has been applied to identify the method that gives high fidelity results, consistent with relevant literature. The analysis outcome requires further processing, as explained in section 5.1, to account for the main weakness of this method, that it does not consider the SNP-to-SNP effect, tackled using as proxy the notion of linkage disequilibrium (LD), and subsequently to topologically and functionally enhance the filtered findings. Once this additional step has been performed, a cross-traits analysis is applied, described in section 5.2, where the DA genetic signature is compared with the signatures of phenotypic traits analyzed in a similar study, the cerebral and facial shape [12].

Chapter 2

Dataset Description

2.1 General

A large dataset of 19,654 individuals was used as the main, discovery dataset, while a smaller one, coming from a different batch, of 16,342 individuals was used as a replication dataset during GWAS.

2.2 Phenotype

2.3 Genome

Chapter 3

Data Preprocessing

Chapter 4

Asymmetry Phenotypic Analysis

4.1 Statistical Analysis of Asymmetry Components

4.2 Phenotypic Partitioning

Hierarchical spectral clustering is an unsupervised method of iterative partitioning, that makes use of the distance matrix eigenvectors [13]. It results into a binary tree structure, where each parent shape is partitioned into two children. In the current study, a level-4 partitioning is performed, resulting into 31 partitions. Subsequently they are transformed to the corresponding principal components that explain 80% of the variance, for reasons of further dimensionality reduction.

Chapter 5

Asymmetry Genetic Analysis

5.1 Post-Processing

5.2 Meta-Analysis

Chapter 6

Discussion

Bibliography

- [1] X. Bailly, H. Reichert, and V. Hartenstein. THE URBILATERIAN BRAIN REVISITED: NOVEL INSIGHTS INTO OLD QUESTIONS FROM NEW FLATWORM CLADES. *Development genes and evolution*, 223(3):149–157, may 2013.
- [2] A. Cichonska, J. Rousu, P. Marttinen, A. J. Kangas, P. Soininen, T. Lehtimäki, O. T. Raitakari, M. R. Järvelin, V. Salomaa, M. Ala-Korpela, S. Ripatti, and M. Pirinen. metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. *Bioinformatics*, 32(13):1981–1989, jul 2016.
- [3] P. Claes, J. Roosenboom, J. D. White, T. Swigut, D. Sero, J. Li, M. K. Lee, A. Zaidi, B. C. Mattern, C. Liebowitz, L. Pearson, T. González, E. J. Leslie, J. C. Carlson, E. Orlova, P. Suetens, D. Vandermeulen, E. Feingold, M. L. Marazita, J. R. Shaffer, J. Wysocka, M. D. Shriver, and S. M. Weinberg. Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nature Genetics*, 50(3):414–423, 2018.
- [4] M. C. Corballis. The evolution and genetics of cerebral asymmetry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1519):867, 2009.
- [5] S. B. Fernandes, K. S. Zhang, T. M. Jamann, and A. E. Lipka. How Well Can Multivariate and Univariate GWAS Distinguish Between True and Spurious Pleiotropy? *Frontiers in Genetics*, 11:1747, jan 2021.
- [6] T. E. Galesloot, K. Van Steen, L. A. Kiemeny, L. L. Janss, and S. H. Vermeulen. A comparison of multivariate genome-wide association methods. *PLoS ONE*, 9(4), apr 2014.
- [7] M. R. Herbert, D. A. Ziegler, C. K. Deutsch, L. M. O’Brien, D. N. Kennedy, P. A. Filipek, A. I. Bakardjiev, J. Hodgson, M. Takeoka, N. Makris, and V. S. Caviness Jr. Brain asymmetries in autism and developmental language disorder: a nested whole-brain analysis. *Brain*, 128(1):213–226, jan 2005.
- [8] D. Jackson, R. Riley, and I. R. White. Multivariate meta-analysis: Potential and promise. *Statistics in Medicine*, 30(20):2481, sep 2011.

- [9] C. P. Klingenberg, M. Barluenga, and A. Meyer. Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. *Evolution*, 56(10):1909–1920, 2002.
- [10] X.-Z. Kong, M. C. Postema, T. Guadalupe, C. de Kovel, P. S. W. Boedhoe, M. Hoogman, S. R. Mathias, D. van Rooij, D. Schijven, D. C. Glahn, S. E. Medland, N. Jahanshad, S. I. Thomopoulos, J. A. Turner, J. Buitelaar, T. G. M. van Erp, B. Franke, S. E. Fisher, O. A. van den Heuvel, L. Schmaal, P. M. Thompson, and C. Francks. Mapping brain asymmetry in health and disease through the ENIGMA consortium. *Human Brain Mapping*, 43(1):167–181, jan 2022.
- [11] M. LeMay. Morphological cerebral asymmetries of modern man, fossil man, and nonhuman primate. *Annals of the New York Academy of Sciences*, 280:349–366, 1976.
- [12] S. Naqvi, Y. Sleyp, H. Hoskens, K. Indencleef, J. P. Spence, R. Bruffaerts, A. Radwan, R. J. Eller, S. Richmond, M. D. Shriver, J. R. Shaffer, S. M. Weinberg, S. Walsh, J. Thompson, J. K. Pritchard, S. Sunaert, H. Peeters, J. Wysocka, and P. Claes. Shared heritability of human face and brain shape. *Nature Genetics*, 53(6):830–839, 2021.
- [13] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, 2002.
- [14] M. Ribolsi, Z. J. Daskalakis, A. Siracusano, and G. Koch. Abnormal Asymmetry of Brain Connectivity in Schizophrenia , 2014.
- [15] J. Sheng, L. Wang, H. Cheng, Q. Zhang, R. Zhou, and Y. Shi. Strategies for multivariate analyses of imaging genetics study in Alzheimer’s disease. *Neuroscience Letters*, 762:136147, sep 2021.
- [16] N. Vijayakumar, N. B. Allen, G. Youssef, M. Dennison, M. Yücel, J. G. Simmons, and S. Whittle. Brain development during adolescence: A mixed-longitudinal investigation of cortical thickness, surface area, and volume. *Human Brain Mapping*, 37(6):2027–2038, jun 2016.
- [17] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing* 2007 17:4, 17(4):395–416, aug 2007.