KU LEUVEN

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

Academic year 2022

Preface

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

The Author
1 January 2010

Contents

List of Figures

List of Tables

Abbreviations

ANOVA analysis of variance

CCA canonical correlation analysis

DA directional asymmetry

DNA deoxyribonucleic acid

DOF degrees of freedom

FA fluctuating asymmetry

GWAS genome wide association studies

HSC hierarchical spectral clustering

LD linkage disequilibrium

MRI magnetic resonance imaging

mvGWAS multivariate genome-wide association study

RSS residual sum of squares

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 magnetic resonance imaging (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 (HSC), 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

1.1 The notion of brain symmetry

Cerebral bilateral symmetry is a universal quality of organisms belonging to the Bilateria lineage [5][1]. For the mammals group, the brain is anatomically divided into a left and right hemisphere. Asymmetric cell division during brain development, initially observed among neuroblasts, causes shape deviations between the two hemispheres (Figure 1.1). This property gives rise to partial functional disassociation, called brain lateralization , with subsets of tasks requiring differential activation of each hemisphere. Lateralization becomes visible when examining organisms' behavior, with the most studied trait in humans being handedness and language [20]. Along with the purely genetic reasons, environment also plays a significant role in affecting cerebral bilateral asymmetry. Of primary interest in this work is directional asymmetry (DA), the asymmetric component that arises by comparing a single individual's hemispheric 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' [14], that is observed in humans between the right and left hemisphere (Figure 1.2). Past studies have shown that abnormal DA may be an indication of certain diseases. The lack of it may imply schizophrenia predisposition [19]. Any significant abnormalities may be indicative of other psychiatric disorders, such as autism or developmental language disorder [9][13].

1.2 Data description

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

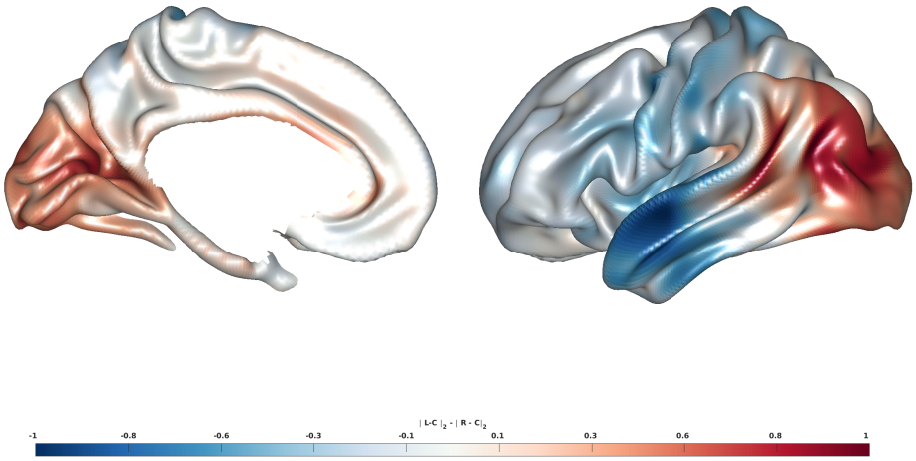

Figure 1.1: Illustration of brain asymmetry. Normalized differences of the distances of landmarks from the center of mass of each averaged hemisphere midthickness surface, after a scaling and alignment process, across the studied population.

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 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.

1.3 Breaking the complexity into parts

The present work evaluates the brain asymmetry genetic landscape in a coarse-to-fine segmentation, through the application of hierarchical spectral clustering (HSC)[17], discussed in section 4.2. The technique has been used in a number of different related studies [4][15], yielding results that are in accordance with the underlying anatomic 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-

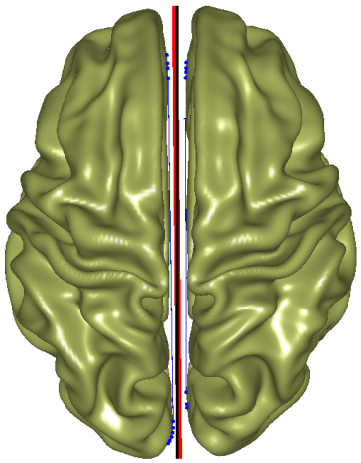

FIGURE 1.2: Illustration of the Yakovlevian torque. Displayed by a red line rotated counterclockwise 0.51 degrees in relation to the perfectly vertical black line, as calculated by using the average angle of the longest edges (in blue) of the convex hulls of the horizontal plane projection of each hemisphere midthickness surface, after a scaling and alignment process, across the studied population.

based clustering is governed by the least quantity of assumptions, regarding the shape or form of the cluster [25]. 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 [23].

1.4 Searching for the origin

The genomic studies are performed under the framework of single nucleotide polymorphism (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 [8]. 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 [3]. Which approach performs best mainly lies

on the dataset properties and the nature of the scientific question. Factors such as low sample size [21], genes pleiotropic effects [7] or within-study variation [10] 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. Diverse experiments, analyzed in chapter 5, have 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 [15], the cerebral and facial shapes.

1.5 Novelties based on related literature

Due to the biological importance of cerebral bilateral asymmetry, it is a subject that has been rigorously studied from multiple viewpoints.

1.5.1 Evolution

From an evolutionary stand, it is extremely rare for the right conditions to occur, in order for any soft tissue specimen to be preserved, across a considerable amount of time. The only known way is through mineralization [18]. This fact renders a mammal's ancestor brain almost impossible to retrieve. Nevertheless, endocranial imprints have been used as a proxy to describe the relationships between hominids and their ancestors [2][16]. The reason behind this phenotypic delegation is purely practical. The brain size and shape follow the container volume restrictions. The brain sulci (i.e. grooves) and gyri (i.e. bumps) in humans are the result of the tremendous expansion of the cerebral cortex surface area during fetal development, folding and wrinkling in order to fit the skull [6]. Although such studies support the theory of propagating asymmetry among studied individuals, with the most evident signs of DA in human skulls, little information about the surface shape can be retrieved, as only the convex hull shape of the brain can be delineated from such process. Through the association of brain asymmetry with DNA, a universal code among organisms, it becomes possible to deploy tools used by evolutionary geneticists, to identify the phylogenetic tree of this complex trait, locating conserved regions among organisms and their predicted divergence in time, under a pleiotropic model [12].

1.5.2 Clinical studies

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

Bilateral asymmetry is mainly described using three components in literature [11][24]; directional asymmetry (DA), the focus of this study, corresponds to the hemispheric side effect; antisymmetry, which is related to the effect where sidedness is random in a population (i.e. left-right randomly switches to right-left), is not observed in the human cerebral cortex, in contrast to other internal organs positions, or organisms [16]; fluctuating asymmetry (FA), encompasses any random developmental and environmental effects, that cannot be explained with the existing knowledge. The observed deviations can be statistically linearly modeled as products of two effects, the hemisphere side studied and the individual specimen analyzed, as well as their interaction [11]. Formally, based on [22] assuming the presence of replications of the observation per individual, to account for technical error, a mixed linear model representing the aforementioned dependencies is defined as:

$$Y_{ijk} = \mu + \beta + I_i + S_{ij} + E_{ijk} \tag{4.1}$$

where Y_{ijk} is the phenotype of the i-th individual, from the j-th side, under the k-th replication, μ and β are the fixed intercept and fixed side effect respectively, $I_i \sim \mathcal{N}(0, \sigma^2_{ind})$ is the random individual effect, $S_{ij} \sim \mathcal{N}(0, \sigma^2_{FA})$ is the random side and individual specific effect, matched to FA, and $E_{ijk} \sim \mathcal{N}(0, \sigma^2_{ME})$ is the measurement error. Replications are necessary in such a study, in order to differentiate FA effect from the measurement error. Given this definition, a way to measure the statistical significance is through an F-test applied on 2-way analysis of variance (ANOVA), to relate the residual sum of squares (RSS) ratios of effects to observable error terms, and of fluctuating effect to the measurement error. Extra care needs to be given on the determination of the degrees of freedom (DOF) of each term, given the preprocessing applied to bring the hemispheres surfaces into Kendall shape space. Those are extracted from the rigorous work in [11]. Given that the analysis is performed on a pair of symmetric objects, and not on a single symmetric object, this

configuration is named matching asymmetry analysis. In order to avoid further de facto assumptions, regarding the distribution of TO BE CONTINUED

4.2 Phenotypic partitioning

Hierarchical spectral clustering (HSC) is an unsupervised method of iterative partitioning, that makes use of the distance matrix eigenvectors [17]. It results into a binary tree structure (i.e. 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. TO BE CONTINUED

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 oldquestions from new flatworm clades. *Development genes and evolution*, 223(3):149–157, may 2013.

[2] A. Balzeau, E. Gilissen, and D. Grimaud-Hervé. Shared Pattern of Endocranial Shape Asymmetries among Great Apes, Anatomically Modern Humans, and Fossil Hominins. *PLoS ONE*, 7(1):e29581, jan 2012.

[3] 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.

[4] 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.

[5] M. C. Corballis. The evolution and genetics of cerebral asymmetry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1519):867, 2009.

[6] M. F. Bear, B. W. Connors, and M. A. Paradiso. Special Features of the Human CNS. In *Exploring the brain*, chapter 7, pages 205–207. Wolters Kluwer, 4th edition, 2016.

[7] 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.

[8] T. E. Galesloot, K. Van Steen, L. A. Kiemeney, L. L. Janss, and S. H. Vermeulen. A comparison of multivariate genome-wide association methods. *PLoS ONE*, 9(4), apr 2014.

[9] 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.

[10] D. Jackson, R. Riley, and I. R. White. Multivariate meta-analysis: Potential and promise. *Statistics in Medicine*, 30(20):2481, sep 2011.

[11] 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.

[12] E. M. Koch and S. R. Sunyaev. Maintenance of Complex Trait Variation: Classic Theory and Modern Data. *Frontiers in Genetics*, 12:2198, nov 2021.

[13] 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.

[14] 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.

[15] 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.

[16] S. Neubauer, P. Gunz, N. A. Scott, J. J. Hublin, and P. Mitteroecker. Evolution of brain lateralization: A shared hominid pattern of endocranial asymmetry is much more variable in humans than in great apes. *Science Advances*, 6(7):9935–9949, 2020.

[17] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, 2002.

[18] M. A. Purnell, P. J. Donoghue, S. E. Gabbott, M. E. McNamara, D. J. Murdock, and R. S. Sansom. Experimental analysis of soft-tissue fossilization: opening the black box. *Palaeontology*, 61(3):317–323, may 2018.

[19] M. Ribolsi, Z. J. Daskalakis, A. Siracusano, and G. Koch. Abnormal Asymmetry of Brain Connectivity in Schizophrenia , 2014.

[20] J. Schmitz, O. Güntürkün, and S. Ocklenburg. Building an asymmetrical brain: The molecular perspective. *Frontiers in Psychology*, 10(APR):982, 2019.

[21] 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.

[22] S. Van Dongen, G. Molenberghs, and E. Matthysen. The statistical analysis of fluctuating asymmetry: REML estimation of a mixed regression model. *Journal of Evolutionary Biology*, 12(1):94–102, jan 1999.

[23] 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.

[24] G. Vingerhoets, R. Gerrits, and H. Verhelst. Atypical Brain Asymmetry in Human Situs Inversus: Gut Feeling or Real Evidence? *Symmetry 2021, Vol. 13, Page 695*, 13(4):695, apr 2021.

[25] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing 2007 17:4*, 17(4):395–416, aug 2007.