GENNADY V. ROSHCHUPKIN

IMAGING GENETICS

Methodological approaches to overcoming
high dimensional barriers

IMAGING GENETICS
Methodological approaches to overcoming high dimensional barriers

Gennady Roshchupkin

ISBN:	978-94-9301-450-3
Cover:	Natalia Mishchenko

Layout:	Ilse Modder, www.ilsemodder.nl
Printing:	Gildeprint - Enschede, www.gildeprint.nl

IMAGING GENETICS
Methodological approaches to
overcoming high dimensional barriers

BEELDEN RELATEREN AAN GENETISCHE DATA
Methodologische benaderingen om hoog-dimensionale barrières te overwinnen

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof. dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
dinsdag 25 september 2018 om 13.30 uur

door

Gennady Vasilievich Roshchupkin
geboren te Sergiev Posad, Rusland

PROMOTIECOMMISSIE

Promotoren	Prof.dr. W.J. Niessen
	Prof.dr. M.A. Ikram
Overige leden	Prof.dr.ir.B.P.F. Lelieveldt
	Prof.dr. A.G. Uitterlinden
	Prof.dr. M.W. Vernooij

"Per aspera ad astra"

TABLE OF CONTENTS

1
Introduction

MANUSCRIPTS IN THIS THESIS

CHAPTER 2

S. J. van der Lee *, **G.V. Roshchupkin***, et al. (2017). "Gray matter heritability in family-based and population-based studies using voxel-based morphometry." <u>Human Brain Mapping</u>

G.V. Roshchupkin, B. A. Gutman, et al. (2016). "Heritability of the shape of subcortical brain structures in the general population." <u>Nature Communications</u>

CHAPTER 3

G.V. Roshchupkin, H. H. Adams, et al. (2016). "Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology." <u>Neurobiology of Aging</u>

H. I. Zonneveld*, **G.V. Roshchupkin*** et al "The neural substrate of cognition in general population." <u>Submitted</u>

CHAPTER 4

G.V. Roshchupkin, H. Adams, et al. (2016). "HASE: Framework for efficient high-dimensional association analyses." <u>Scientific Reports</u>

G.V. Roshchupkin et al "Sum Ranking, simple but powerful method for detecting pleiotropic loci." <u>In preparation</u>

CHAPTER 5

H.H. Adams*, A. Teumer*, D.P. Hibar*, **G.V. Roshchupkin*** et al "Genetic architecture of the human anterior commissure." <u>In preparation</u>

E. Hofer*, **G.V. Roshchupkin*,** et al. "GWAS of cortical thickness, surface area and volume in cortical regions of interest." <u>Submitted</u>

G.V. Roshchupkin et al „Genome-wide association studies of 1.5 million features of brain morphometry." <u>In preparation</u>

INTRODUCTION

Evidence that heredity, and later on genetics, shapes human brain structure and function has been accumulating for nearly two centuries[1,2]. However, the way in which this relation could be studied was quite limited for a long time. Many of the initial links that were observed between genetics and brain structure arose from clinical observations of patients with a family history of diseases or observations during surgical interventions [3]. In the last decades, through several revolutionary steps in both the field of imaging and genetics, our possibilities to study the relation between genetics and brain structure and function have dramatically increased. First, the development of magnetic resonance imaging (MRI) technology made it possible to observe human brain anatomy, function and pathology in vivo, and to study its changes during the lifespan. Subsequently, genotyping and other omics data technologies[4] were developed and became sufficiently cheap to allow to collect large datasets with sufficient statistical power to analyze genetic influence on the brain.

For a long time neuroimaging and genetics have developed independently, with most of brain related findings being driven by one of this field. A multitude of case-control genome-wide association studies (GWAS) have been conducted to screen the whole genome for causal loci of complex brain diseases[5-7], without specifically using imaging information. At the same time, in the imaging domain, researchers compared MRI brain scans from healthy persons and patients in an attempt to reveal spatiotemporal patterns of brain changes related to disease[8-10]. This has led to a situation where we have many genetic variants which have been linked to diseases, and knowledge which brain regions are affected by disease, but almost no clue if and how genetic variants, linked with disease, associate with brain structures.

Owing to the increasing availability of studies with both genetic and imaging data[11], a new research domain has developed in the last decade, i.e. the field of imaging genetics. Within the neuro domain, this has become a very active and rich research field. By combining genetic and neuroimaging data we do not only have new opportunities to significantly improve our understanding of brain genetic architecture but we also have the potential to discover genetic disease factors, by directly linking variants to brain structures. For example, if a gene is found that influences hippocampus volume or shape, it is quite possible that this gene plays a role in illnesses associated with changes in the hippocampus (e.g., Alzheimer disease).

Recently, due to large-scale international efforts by researchers organized in different consortiums, such as CHARGE and ENIGMA, several large-scale GWAS studies of brain structures have been performed[12,13] (the combined dataset includes more than 30.000 scans collected from all over the world). These studies have found new loci with potential links to neurode-

generation, Alzheimer's disease, microcephaly, and schizophrenia. While this is a fantastic result, this is only the tip of the iceberg of how the field of imaging genetics can contribute to our understanding of brain genetics.

Like any emerging field, imaging genetics has its own challenges and issues. Despite the evidence that most of the neuropsychiatric disorders are to varying degrees genetically determined[14], unraveling which genetic variants are responsible for such effect remains difficult. One way to address this challenge is to use so-called endophenotypes in the analysis. An endophenotype is a phenotype which is assumed to have a simpler genetic architecture or reduced genetic complexity than other phenotypes, and which has a higher penetrance for genetic variants because they are biologically closer to the genes. The advantages of endophenotypes is not only in increasing statistical power in gene discovery studies, but also, especially for neuroimaging, in improved interpretability and insight into underlying degenerative and developmental conditions.

Determining which the optimal imaging endophenotypes are is non-trivial. Currently, in neuroimaging studies mainly aggregate measures such as volume, mean cortical thickness or surface area are mostly used. These are limited in the sense that they do not capture the complexity of brain structures. For example, the hippocampus is made up of several subfields, each with partially independent functional roles, as well as thalamus consist of multiple nuclei. The same is known for the cerebral cortex, it has been shown that genetic influence and genetic correlation between different regions does not necessarily overlap with biologically defined boundaries[15].

Using more-refined brain mapping techniques allows localizing genetic effects to specific tissue layers and sub-regions as opposed to using aggregate measures. Studying the relation between genetics and brain anatomy, at the higher resolution, would likely benefit researches aimed to advance our understanding of the brain and of the underlying neurobiology of complex disorders.

Neuroimaging methods are now able to provide hundreds or even thousands of different types of high-dimensional (HD) phenotypes, and it is therefore critical to select the most appropriate for analysis. Not all of them have genetic architecture or provide additional information beyond the standard measures.

Therefore, the overall aim of this thesis is to address a number of challenges in imaging genetics field. First, to demonstrate the utility of high-dimensional neuroimaging phenotypes for genetics analysis. Second, to develop methods and software which are able to handle compu-

tationally and data size wise the HD imaging genetics analysis. Third, to apply such technique on multi-center level and discover the genetic architecture of HD phenotypes.

In **chapter 2,** we explore two different HD endophenotypes and their relevance for genetic analysis. First, in **chapter 2.1,** using voxel-based morphometry (VBM), we study gray matter density on the voxel level (the highest level of possible resolution derived from MRI). We investigate to what extent gray matter density at the voxel level is heritable (how much of variability can be explained by genetics) and analyze the consistency of results between different populations. In **chapter 2.2** we propose subcortical shapes as a new endophenotype for genetic analysis and study their genetic architecture in relation to volumetric measures.

Chapter 3 focuses on the application of these endophenotypes in population studies. **In chapter 3.1** we describe the relation between brain gray matter voxels and genetic variants of Alzheimer Disease. We also introduce a novel technique which integrates 3D brain maps of genes expression with spatially localized associations from VBM results. **In chapter 3.2,** we demonstrate how HD phenotypes can help in mapping the relation between brain and cognitive functions.

Given the enormous potential of high-dimensional phenotypes and the increasing availability of large biobank initiatives with hundreds of thousands participants with both MRI and omics data, it becomes crucial to have methodology to be able to effectively analyze such large datasets. In **chapter 4.1** I developed a framework (HASE) which dramatically reduces computational time for HD GWAS analysis from years to only hours and implemented a new meta-analytical approach, which requires just several gigabytes to be exchanged between collaborators, compare to terabytes in conventional methods.

At the same time, multi-center and consortia efforts make GWAS summary data more easily accessible for many different diseases and phenotypes. Therefore approaches to mine biologically meaningful information from such summary statistic, without access to the raw data, become indispensable. In **chapter 4.2,** I propose a new method for pleiotropy detection, the phenomena when one gene affects many outcomes. This method is very powerful for single loci pleiotropy detection and only requires GWAS summary data.

In **chapter 5.2, 5.3,** I apply the HASE framework for GWAS analyses of neuroimaging phenotypes. In **chapter 5.2** we performed GWAS analyses of cortical thickness, surface area and volume in the whole cortex as well as in 34 regions of interest in 22.000 individuals from 18 cohorts within the CHARGE consortium and UK Biobank. In **chapter 5.3** we collected VBM data from several cohorts (more 18.000 subjects) and were able to perform HD vox-

el-wise genome-wide association study (1.5 million GWASes) just within 20 hours.

Finally, **Chapter 6** provides a broader discussion of the main findings and results of this thesis, along with my opinion about the future perspective of the imaging genetics field and directions of further research.

1. Jacyna, S. The most important of all the organs: Darwin on the brain. *Brain* **132**, 3481-3487, doi:10.1093/brain/awp283 (2009).

2. Cowan, W. M., Harter, D. H. & Kandel, E. R. The emergence of modern neuroscience: Some implications for neurology and psychiatry. *Annu Rev Neurosci* **23**, 343-391, doi:DOI 10.1146/annurev.neuro.23.1.343 (2000).

3. Tanzi, R. E. Genetic linkage studies of human neurodegenerative disorders. *Curr Opin Neurobiol* **1**, 455-461 (1991).

4. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nat Rev Genet* **19**, 299-310, doi:10.1038/nrg.2018.4 (2018).

5. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* **45**, 1452-1458, doi:10.1038/ng.2802 (2013).

6. Nalls, M. a. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics* **056**, 1-7, doi:10.1038/ng.3043 (2014).

7. Smoller, J. W. *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-1379, doi:10.1016/S0140-6736(12)62129-1.Identification (2013).

8. Ellison-Wright, I. & Bullmore, E. Meta-analysis of diffusion tensor imaging studies in schizophrenia. *Schizophrenia research* **108**, 3-10, doi:10.1016/j.schres.2008.11.021 (2009).

9. Ikram, M. A. *et al.* Brain tissue volumes in relation to cognitive function and risk of dementia. *Neurobiology of aging* **31**, 378-386, doi:10.1016/j.neurobiolaging.2008.04.008 (2010).

10. Zhang, T. & Davatzikos, C. Optimally-Discriminative Voxel-Based Morphometry significantly increases the ability to detect group differences in schizophrenia, mild cognitive impairment, and Alzheimer's disease. *NeuroImage* **79**, 94-110, doi:10.1016/j.neuroimage.2013.04.063 (2013).

11. Thompson, P. M. *et al.* The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging and behavior* **8**, 153-182, doi:10.1007/s11682-013-9269-5 (2014).

12. Adams, H. H. H. *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nature Neuroscience* **id**, doi:10.1038/nn.4398 (2016).

13. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **8**, doi:10.1038/nature14101 (2015).

14. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279, doi:10.1093/bioinformatics/btw613 (2017).

15. Chen, C. H. *et al.* Hierarchical genetic organization of human cortical surface area. *Science* **335**, 1634-1636, doi:10.1126/science.1215330 (2012)
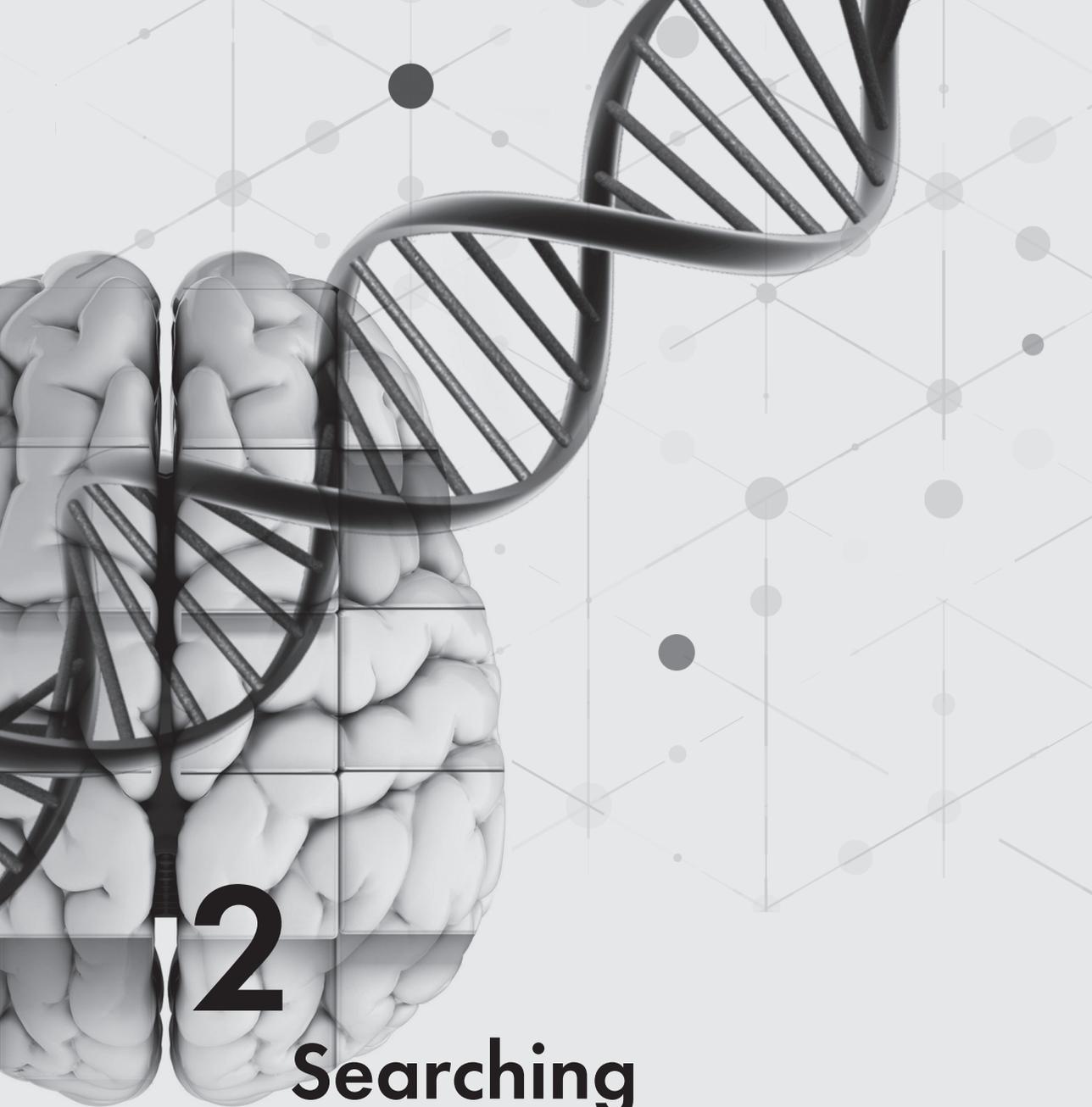
2

Searching for the best high-dimensional endophenotypes

CHAPTER 2.1

Grey matter heritability
in family-based and
population-based studies
using voxel-based morphometry

ABSTRACT

Background

The combination of genetics and imaging has improved our understanding of the brain through studies of aggregate measures obtained from high-resolution structural imaging. Voxel-wise analyses have the potential to provide more detailed information of genetic influences on the brain. Here we report a large-scale study of the heritability of grey matter at voxel resolution ($1 \times 1 \times 1$mm).

Methods

Validated voxel-based morphometry (VBM) protocols were applied to process magnetic resonance imaging data of 3239 unrelated subjects from a population-based study and 491 subjects from two familybased studies. Genome-wide genetic data was used to estimate voxel-wise gray matter heritability of the unrelated subjects and pedigree-structure was used to estimate heritability in families. We subsequently associated two genetic variants, known to be linked with subcortical brain volume, with most heritable voxels to determine **i**f this would enhance their association signals.

Results

Voxels significantly heritable in both estimates mapped to subcortical structures, but also voxels in the language areas of the left hemisphere were found significantly heritable. When comparing regional patterns of heritability, familybased estimates were higher than population-based estimates. However, regional consistency of the heritability measures across study designs was high (Pearson's correlation coefficient=0.73, $p=2.6 \times 10^{-13}$). We further show enhancement of the association signal of two previously discovered genetic loci with subcortical volume by using only the most heritable voxels.

Conclusion

Grey matter voxel-wise heritability can be reliably estimated with different methods. Combining heritability estimates from multiple studies is feasible to construct reliable heritability maps of grey matter voxels.

INTRODUCTION

The human brain shows large inter-individual variation, which could be explained by genetic and environmental influences. Studying these influences is essential in better understanding brain structure and function. The degree to which genetics explains phenotypic variation, in other words heritability, depends on many factors: the actual genetic contribution to the trait, environmental effects, measurement error, study design and sample characteristics.[1-3] Recently an overview was published of fifty years of worldwide heritability research in twins encompassing thousands of traits, showing heritability studies are highly informative on how large the genetic contribution to a trait is.[4] Heritability studies could aid future genetic research to focus on particular regions of interest in the brain. For example, large scale genetic studies of brain structures with the highest heritability typically yield the most findings.[5] When studying the multitude of measures of voxel based magnetic resonance imaging (MRI), limiting genetic studies to the most heritable traits could be feasible in light of multiple testing. Recent studies have focused on heritability of detailed MRI measures at a voxel level.[6-12] Different study designs showed comparably high estimates for white matter tract heritability in twin pairs,[9,12] sib-pairs,[7] and extended pedigrees (heritability = 50-90%).[10] The heritability of grey matter was studied by voxel-based morphometry (VBM) previously,[13-15] but the studies were relatively small and relatively large voxels were studied. Moreover, heritability of grey matter VBM has not been estimated in population-based studies.

Here, we perform a large multi-site study to estimate the voxel-wise heritability of grey matter. We calculate pedigree-based heritability in two family-based studies and heritability based on genome-wide genetic data in a large population-based study of unrelated subjects. Using these approaches, we created two grey matter heritability maps and described which regions contain significantly heritable voxels in both designs. We furthermore estimated overall regional consistency of the heritability measures across study designs and explored if usage of our heritability maps could potentially enhance association signals of two genetic variations, previously discovered by genome-wide association studies.[5,16,17]

METHODS

STUDY SUBJECTS AND IMAGING PROTOCOL

Rotterdam Study – The Rotterdam Study is a population-based cohort study among inhabitants of a district of Rotterdam (Ommoord), The Netherlands, and aims to examine the determinants of disease and health in the elderly with a focus on neurogeriatric, cardiovascular, bone, and eye disease.[18] In 1990 to 1993, 7983 persons participated and were re-examined every 3 to 4 years (RS-I). In 2000 to 2001 the cohort was expanded by 3011 persons who had not yet been part of the Rotterdam Study (RS-II). All participants had DNA extracted from blood at their first visit. In 2006-2008 a second expansion (RS-III) of 3,932 persons aged 45 and over was realized. Genotyping was performed at the Human Genotyping Facility, Genetic Laboratory Department of Internal Medicine, Erasmus MC, Rotterdam. Genotypes were imputed to the 1000 genomes phase I version 3 reference panel, using standard methods and software.[19] From 2005 onwards MRI is part of the core protocol of the Rotterdam study.[20] For this study a total of 4071 unique study participants had both MRI and genetic data and were available for analysis. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Wet Bevolkingsonderzoek: ERGO (Population Studies Act: Rotterdam Study). All participants provided written informed consent to participate in the study and to obtain information from their treating physicians.

Erasmus Rucphen Family (ERF) – The ERF study is a family-based cohort study in a genetically isolated population from a community in the South-West of the Netherlands (Rucphen municipality) including 3,000 participants. Participants are all descendants of a limited number of founders living in the 19th century, and all of Caucasian European descent. Extensive genealogical data is available for this population. The study population is described in detail elsewhere.[21] In a follow-up analysis, non-demented hypertensive (systolic blood pressure ≥ 160, diastolic blood pressure ≥ 100 or use of antihypertensive medication) subjects aged 55-75 years were included for a new battery of tests, including MRI scanning.[22] These 122 participants from the ERF were related to each other in one large pedigree. This large pedigree was split into multiple small pedigrees for heritability calculations (pedcut version 1.19 http://mga.bionet.nsc.ru/soft/). Participants related to each other in 27 families with in total 880 relatives. The average size of the pedigrees was 32.6 relatives (range 20-44) with on average 4.5 participants with MRI per family. All participants gave informed consent to participate in the study and to obtain information from their treating physicians. The study was approved by the medical ethics committee at Erasmus MC University Medical Center, Rotterdam, The Netherlands.

MRI scanning for ERF and the Rotterdam Study was done on the same 1.5 T MRI unit (GE

Healthcare, Milwaukee, USA, Signa Excite software version 11×) fitted with a dedicated 8-channel head coil. The T1-weighted, proton density-weighted (PDw) and fluid-attenuated inversion recovery (FLAIR) sequences were used.[20] For the purpose of segmentation, the T1w scan is acquired in 3D at high in-plane resolution and with thin slices (voxel size < 1 mm3).[20]

Austrian Stroke Prevention Study (ASPS) – The ASPS study is a single-center, prospective follow-up study on the effects of vascular risk factors on brain structure and function in the normal elderly population of the city of Graz, Austria. The procedure of recruitment and diagnostic work-up of study participants has been described previously.[23,24] Between 2006 and 2013 the study was extended for the Austrian Stroke Prevention Family Study (ASPS-Fam).[25] Study participants of the ASPS and their first grade relatives were invited to enter ASPS-Fam. Inclusion criteria were no history of previous stroke or dementia and a normal neurological examination. In total 176 families connecting a total of 719 relatives, among which 369 were study participants with brain-MRI. The average size of the pedigrees was 4 (range 1-10) relatives with on average 2.4 participants with MRI per family. The diagnostic work-up was identical to the original study. The study protocol was approved by the ethics committee of the Medical University of Graz, Austria, and written and informed consent was obtained from all subjects. MRI scanning of the ASPS-Fam was done on a 3.0 T Tim Trio (Siemens, Erlangen). T1-MPRAGE 1×1×1mm was used for image processing.[25]

IMAGE PROCESSING

Prior to analysis, a number of pre-processing steps were performed. For multispectral image analysis, the different scans were spatially registered using rigid registration [20]. Subsequently, the brain was extracted from the scan. Hereto a manually segmented brain mask, which excludes cerebellum, eyes and skull, was non-rigidly registered to the T1-weighted image using Elastix.[20] Finally, scans were corrected for intensity non-uniformity using the N3 method; non-uniformity correction was carried out within the brain mask.[20] All T1-weighted images were segmented into supra-tentorial grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). For the Rotterdam Study and ERF, we used a previously described kNearest-Neighbor (kNN) algorithm, which was trained on six manually labeled atlases.[26] For the ASPS-Fam study a Quantib BV tissue segmentation tool was applied (www.quantib.org). Quantib® software implements the same algorithm, which we then used for tissue segmentation in the Rotterdam Study and ERF. There are thus no methodological differences between the methods, both of them based on kNN-based segmentation training on manually labeled subjects for segmenting GM, WM and CSF.

Voxel-based morphometry (VBM) was performed by the same optimized VBM protocol in all three studies[27] and previously described.[28] FSL software[29] was used for VBM data processing.

First, all GM density maps were non-linearly registered to the standard GM probability template. For this study we chose the MNI152 GM template (Montreal Neurological Institute) with a $1\times1\times1$ mm voxel resolution.[30] The MNI152 standard-space T1-weighted average structural template image is derived from 152 structural images, which have been warped and averaged into the common MNI152 coordinate system after high-dimensional nonlinear registration. A spatial modulation procedure was used to avoid differences in absolute grey matter volume due to the registration. This involved multiplying voxel density values by the Jacobian determinants estimated during spatial normalization. To decrease signal to noise ratio, all images were smoothed using a 3 mm (FWHM 8 mm) isotropic Gaussian kernel. Thus all results are in MNI space. Brain regions were segmented using atlas-based segmentation based on the Hammer atlas.[31] The modulation step in the VBM pipeline preserves the volume of a particular tissue within a voxel. The multiplication of the voxel values in the segmented images by the Jacobian determinants derived from the spatial normalization step allows us to calculate volumes by aggregating voxels. In total we estimated heritability for 1,405,508 grey matter voxels in all three studies.

REPRODUCIBILITY VBM MEASURES

We investigated the test-retest reliability of the VBM measures in a subset of 83 persons who were scanned twice within 1-9 weeks. We quantified the reproducibility by calculating the intraclass correlation (ICC) of the gray matter density measures for every voxel (Online viewer, **Supplementary Figure 1**).[32]

HERITABILITY ANALYSIS

Population-based heritability estimates were calculated using Genome-wide Complex Trait Analysis (GCTA v1.24)[33] (http://cnsgenomics.com/software/gcta/) in the population-based Rotterdam Study. GCTA implements REML (restricted maximum likelihood) analysis, this method compares genotypic similarity between individuals to their phenotypic similarity. Formula's underlying the GCTA method to determine heritability estimates are described elsewhere[3] and thoroughly explained in a commentary by the authors.[34] The 1000 Genomes imputed genotypes (Imputation quality (Rsq) > 0.5 and minor allele frequency (MAF) > 0.01) were used to create a genetic relationship matrix (GRM) in GCTA.[35] The power of GCTA analysis is determined by pair-wise genetic relationships in the studied population.[3,34] Therefore the three cohorts of the Rotterdam study were combined and analyzed as one in the voxel-wise heritability analysis. Pairwise genetic relatedness between all individuals (N=4,071) was calculated and for pairs with more than 0.02 genotype similarity[35] one person was removed ($N_{removed}$ = 832). REML analysis was then performed in the remaining 3,239 unrelated subjects using the GRM correcting for age and sex. All grey matter heritability was estimated once**.**

Family-based heritability was estimated using maximum-likelihood variance components methods implemented in the SOLAR (version 6.6.2) software.[36] Formulas for the calculation of heritability estimates are described in detail elsewhere.[36] Briefly, the algorithms in SOLAR employ maximum likelihood variance decomposition methods. The covariance matrix Ω for a pedigree of individuals is given by:

$$\Omega = 2 \cdot \Phi \cdot \sigma_g^2 + I \cdot \sigma_e^2$$

where σ_g^2 is the genetic variance due to the additive genetic factors, Φ is the kinship matrix representing the pair-wise kinship coefficients among all individuals, σ_e^2 is the variance due to individual-specific environmental effects, and I is an identity matrix (under the assumption that all environmental effects are uncorrelated among family members). Narrow sense heritability is defined as the fraction of phenotypic variance σ_p^2 attributable to additive genetic factors:

$$\overline{h^2 = \frac{\sigma_g^2}{\sigma_P^2}}$$

The variance parameters are estimated by comparing the observed phenotypic covariance matrix with the covariance matrix predicted by kinship (Almasy and Blangero, 1998). Significance of heritability is tested by comparing the likelihood of the model in which σ_g^2 is constrained to zero with that of a model in which σ_g^2 is estimated. Twice the difference between the $\log_e$ likelihoods of these models yields a test statistic, which is asymptotically distributed as a ½:½ mixture of a χ^2 variable with 1 degree-of-freedom and a point mass at zero. If the algorithm converges SOLAR outputs the heritability value, the significance value (p), and the standard error for each voxel.[8,36]

ERF study and ASPS-Fam were not jointly analysed because ERF subjects were scanned on a 1.5T MRI and ASPS-Fam subjects on a 3.0T MRI. Instead inverse variance meta-analysis using heritability and heritability standard errors was performed in METAL [37] to boost power and improve stability of heritability estimates [6]. Heritability estimates were calculated in both studies with age and sex as covariates. Variance component methods implemented in SOLAR are vulnerable for inflation if phenotypes have a leptokurtic to distribution. Therefore we applied inverse normal transformations in SOLAR to all voxels, but some voxels still violated the distribution of too high residual kurtosis (kurtosis >0.9) and were therefore excluded.[38] Non converging heritability estimates of 0 without standard errors were also excluded from the meta-analysis. In the family-based studies some voxels had valid p-values and a heritability of 1, but missing standard errors. These voxels were located in the middle of voxel-clusters with high heritability (online viewer reference) (close to 1). Therefore standard errors for such voxels were imputed to retain these voxels for meta-analysis. This resulted in imputation of

the standard error for 6.4% of voxels in the ERF study and a negligible percentage of voxels in ASPS-Fam (<0.001%).

ENHANCEMENT OF ASSOCIATION SIGNAL

We explored whether voxel heritability information could enhance the association of genetic variants with brain structures. The genetic variants most significantly associated with hippocampal volume (rs77956314 on 12q24.22, near the gene *HRK*) and putamen volume (rs945270 on 14q22.3, downstream of the gene *KTN1*) were selected from a recently published genome-wide association study on subcortical structures [5]. To select the most heritable voxels in the hippocampus and putamen, we ordered them using three approaches. First, we ranked the voxels from low to high family-based heritability estimates. Second, we ranked them from low to high population-based heritability estimates. In the third approach we summed the ranks obtained from both the family- and population-based estimates and used the sum of the ranks to prioritize the voxels. Using these three approaches we excluded the voxels in a step-wise manner by removing the 5% least heritable voxels. For each step we computed the volume by summing the values of the remaining voxels. As a voxel represents grey matter density in 1 mm^3, the sum of voxels gives the volume of grey matter. We determined the association of the two genetic variants in an additive model with the volumes in linear regression analyses (adjusted for age, sex, and the first three principal components) and compared this to association of the volume derived from all voxels mapped to the structure (i.e. the total VBM-volume of the hippocampus or putamen). The p-value of the association of the genetic variants with the subsets of voxels divided by the p-value of the association of the genetic variants with the total VBM-volume was calculated to measure change in the strength of the association. Genetic effects were calculated in the three cohorts of the Rotterdam study separately (RS-I = 844, RS-II = 1003, RS-III = 2190) and were combined using an inverse variance weighted meta-analysis in METAL.[37]

STATISTICAL ANALYSIS

Descriptive statistics were compared using one-way ANOVA and chi-squared tests. To correct for multiple comparisons we applied FDR p-value thresholds[39] for both population and family heritability separately to declare which voxels are significantly heritable.

RESULTS

POPULATION CHARACTERISTICS

Characteristics of the study population are shown in **Table 1.** The spread of the age of subjects in the ERF study (age range 55-76) was smaller compared to ASPS-Fam (38-86) and the Rotterdam Study (46-98) due to the fact that inclusion criteria for scanning was restricted to midlife **(Table 1)**. However, the average age at the time of MRI-scanning of the cohorts was very similar, ranging from 64.3 ($\pm$ 4.5) years in the ERF study, 64.9 ($\pm$ 10.7) years in ASPS and 64.9 ($\pm$ 10.7) in the Rotterdam Study (p = 0.86). The percentage of women was 52.5% in ERF, 60.4% in ASPS-Fam and 55.3% in the Rotterdam study, these differences were non-significant (p = 0.13) (**Table 1**).

HERITABILITY ESTIMATES

In total 454,184 (33.3% of all voxels) were FDR-significant in the family-based estimates. Mean heritability of significant voxels was 0.44 $\pm$ 0.12 SD (all voxels 0.29 $\pm$ 0.17 SD), with heritability estimates ranging from 0.23 to 1. In total 68,616 (4.9% of all voxels) were FDR-significant in the population-based estimates. Mean heritability of the significant voxels was 0.34 $\pm$ 0.04 SD (all voxels 0.11 $\pm$ 0.10), with heritability estimates ranging from 0.25 to 0.56. We found heritability of 44,349 voxels (3.2% of all voxels) to be FDR significant in the family- as well as the population-based heritability estimates**.** These significantly heritable voxels were clustered, mostly within subcortical brain structures (**Figure 1**). **Table 2** shows the percentage of voxels that were significantly heritable of the total of voxels in a structure in both estimates, as well as the average regional heritability, considering all voxel-wise heritability estimates. Highest percentage of significantly heritable in both estimates voxels were located in the caudate nucleus (right 72.4% and left 68.6%) followed by the putamen (right 57.5% and left 32.6%). Other subcortical structures with a large percentage of significantly heritable voxels were; left pallidum (32.2%), left nucleus accumbens (29.7%), right pallidum (28.5%), left amygdala (21.4%), left hippocampus (17.9%), left thalamus (14.4%), right amygdala (12.8%) and the right insula (11.4%). Apart from the subcortical structures, parts of the right lateral occipitotemporal gyrus (gyrus fusiformis) (10.4%), left straight gyrus (gyrus rectus) (10.4%), left subcallosal area (8.0%) and the left lingual gyrus (7.9%) harbored a proportion significantly heritable voxels (**Figure 1 and Table 2**).
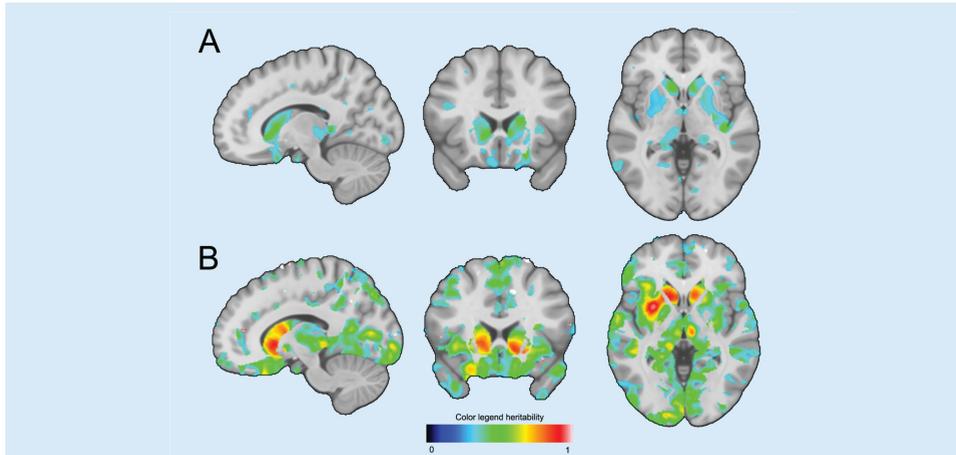
FIGURE 1: Example of FDR-Significant voxels in both population-based (**A**) and family-based (**B**) estimates. Significant voxels cluster in subcortical structures, such as the caudate nucleus. All results can be interactive accessed (www.ima-gene.nl/heritability) and downloaded from the website.

TABLE 1: Descriptive statistics

	ERF	**ASPS-Fam**	**RS**	**p**
Country	Netherlands	Austrian	Netherlands	
Study type	Family-based	Family-based	Population-based	
Field strength	1.5T	3.0T	1.5T	
Sequence	T1-weighted	T1-MPRAGE	T1-weighted	
TR/TE (ms)	13.8/2.8	1900/2.19	13.8/2.8	
TI (ms)	400	900	400	
Flip angle (degrees)	20	9	20	
Voxel size	1x1x1mm	1x1x1mm	1x1x1mm	
Minimum-maximum age	56-76	38-86	46-98	
Age ($\pm$ SD)	64.3 $\pm$ 4.5	64.9 $\pm$ 10.7	64.7 $\pm$ 10.7	0.86[‡]
Minimum-maximum age	56-76	38-86	46-98	
Percentage females	52.5%	60.4%	55.3%	0.13[†]
N participants with MRI-scan	122	369	3239*	
Total relatives in pedigrees	880	718	-	

Descriptive statistics of the included studies. ‡ p calculated with one-way ANOVA, † p chi-squared test. *The total number of participants with brain magnetic imaging and genetics data in the Rotterdam study was 4071, but for pairs with more than 0.02 genotype similarity one person was removed ($N_{removed}$ = 832). MRI = Magnetic Resonance Imaging, ERF = Erasmus Rucphen Family study, ASPS-Fam = Austrian Stroke Prevention Family Study, RS = Rotterdam Study, SD = standard deviation. T1-MPRAGE = T1 weighted 3D sequences with magnetization preparation, TR = repetition time, TE = echo time, TI inversion time.

When comparing regional heritability, estimates calculated in families was always higher than the population-based estimates (p<0.001) **(Figure 2A)** and the difference in heritability between family-based estimates and population-based estimates was relatively stable (mean difference of regional heritability = 0.21 ± 0.08) **(Table 2)**. Therefore, the regional heritability pattern of the family-based estimates significantly predicted the regional pattern of heritability in the population-based study (Pearson's correlation coefficient = 0.73, p = 2.6 × 10^{-13}) **(Figure 2B)**.
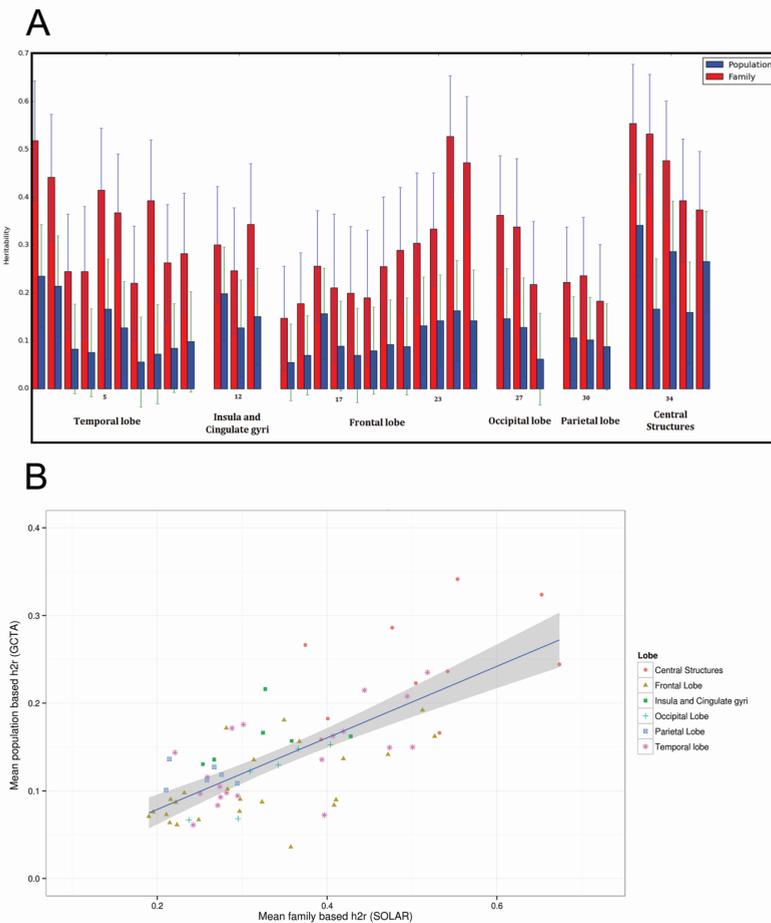
FIGURE 2: A Barplot showing regional brain heritability. Structures that are in both the left as well as the right hemisphere were averaged for this figure. It can clearly be seen that the heritability from family-based studies is higher than heritability form the unrelated population (P<0.001). **B** Scatter plot of the average regional heritability of all brain structures. The correlation of the family-based and population-based estimates was high (Pearson's correlation coefficient = 0.73, p = 2.6 × 10^{-13}). Data points per structure correspond to family and population heritability in table 2.
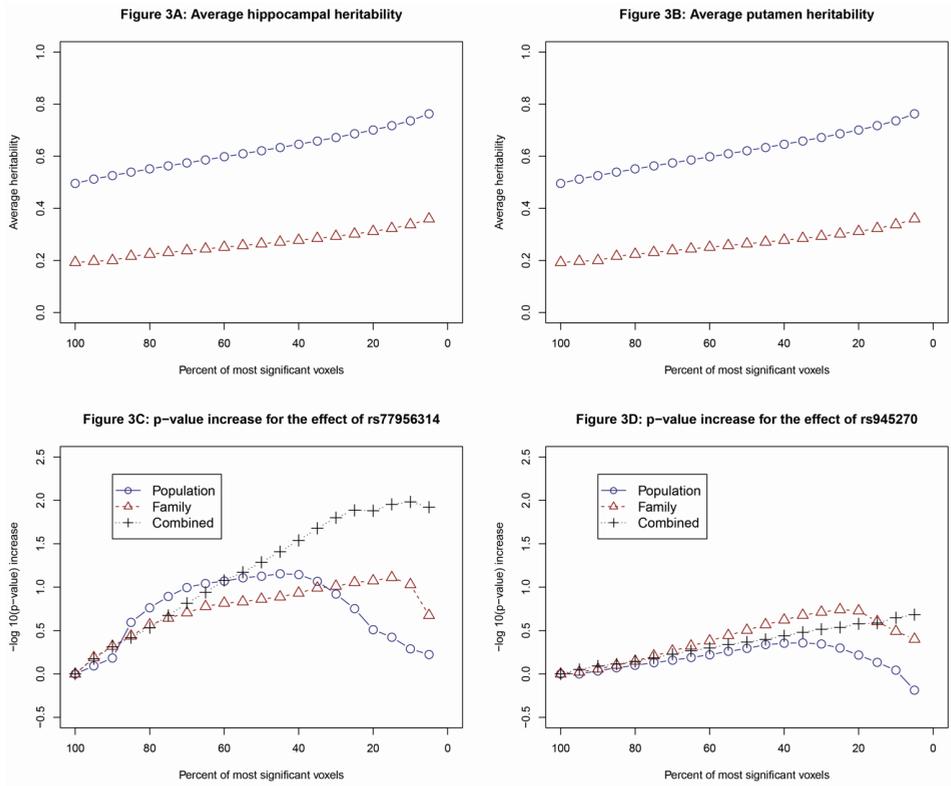
FIGURE 3: Enhancement of the association signal of variants with the most heritable voxels of the hippocampus and putamen. **A,B**: Average heritability (y-axis) of the voxels in hippocampus (A) and putamen (B) given a percentage of the most heritable voxels in that region (x-axis) in steps of 5%. **C,D:** The -log(p-value) increase comparing the p-value of association with subsets of the most heritable voxels and all voxels in the region. The –log(p-value) increase for association of hippocampal with rs77956314 (HRK gene) and putamen voxels with rs945270 (KTN1 gene) is shown. Associations were corrected for age, sex, and the first three principal components.

TABLE 2: Regional heritability estimates

Lobe	Structure	N-voxels	Population significant (%)	Family significant (%)	Significant both (%)	Population (±SD)	Family (±SD)	Δh2
Temporal lobe	Amygdala left	2328	21.4	79.6	21.4	0.21 (±0.09)	0.44 (±0.11)	0.23
Temporal lobe	Amygdala right	2443	13	88.2	12.8	0.21 (±0.09)	0.49 (±0.10)	0.29
Temporal lobe	Anterior temporal lobe, lateral part left	5169	0.1	21.9	0.1	0.08 (±0.08)	0.27 (±0.16)	0.19
Temporal lobe	Anterior temporal lobe, lateral part right	5535	1.2	34.6	0.4	0.10 (±0.08)	0.27 (±0.14)	0.17
Temporal lobe	Anterior temporal lobe, medial part left	10427	3.3	33.9	0.9	0.09 (±0.09)	0.27 (±0.15)	0.18
Temporal lobe	Anterior temporal lobe, medial part right	11157	7.4	34.1	3	0.17 (±0.08)	0.29 (±0.10)	0.12
Temporal lobe	Gyri parahippocampalis et ambiens left	7411	10.7	66.5	6.8	0.17 (±0.10)	0.42 (±0.17)	0.25
Temporal lobe	Gyri parahippocampalis et ambiens right	7548	4.5	87.9	4.3	0.15 (±0.10)	0.50 (±0.12)	0.35
Temporal lobe	Hippocampus left	3482	18.1	93.7	17.9	0.24 (±0.07)	0.52 (±0.15)	0.28
Temporal lobe	Hippocampus right	3725	8.1	83.1	7.7	0.15 (±0.11)	0.47 (±0.16)	0.32
Temporal lobe	Lateral occipitotemporal gyrus (gyrus fusiformis) left	6603	0	69.5	0	0.07 (±0.07)	0.40 (±0.14)	0.32
Temporal lobe	Lateral occipitotemporal gyrus (gyrus fusiformis) right	6631	10.9	68.9	10.4	0.16 (±0.09)	0.41 (±0.15)	0.24
Temporal lobe	Medial and inferior temporal gyri left	23978	0	26.7	0	0.06 (±0.06)	0.24 (±0.14)	0.18
Temporal lobe	Medial and inferior temporal gyri right	24547	1.9	33.8	1.3	0.12 (±0.08)	0.26 (±0.17)	0.14
Temporal lobe	Posterior temporal lobe left	64214	3.5	28.5	2.3	0.10 (±0.10)	0.25 (±0.15)	0.15
Temporal lobe	Posterior temporal lobe right	64529	2.6	36.9	0.6	0.09 (±0.09)	0.29 (±0.16)	0.2
Temporal lobe	Superior temporal gyrus, anterior part left	6618	4.9	49.1	2.3	0.18 (±0.08)	0.30 (±0.09)	0.13
Temporal lobe	Superior temporal gyrus, anterior part right	6874	3	32.8	0.4	0.10 (±0.08)	0.28 (±0.12)	0.18
Temporal lobe	Superior temporal gyrus, central part left	17666	8.1	66.3	5.7	0.14 (±0.10)	0.39 (±0.12)	0.26
Temporal lobe	Superior temporal gyrus, central part right	18329	11.8	20.7	1.5	0.14 (±0.11)	0.22 (±0.13)	0.08
Parietal Lobe	Postcentral gyrus left	39052	4	35	2.3	0.11 (±0.09)	0.29 (±0.17)	0.19
Parietal Lobe	Postcentral gyrus right	34055	4.7	30	1.7	0.13 (±0.10)	0.27 (±0.15)	0.14
Parietal Lobe	Remainder of parietal lobe left (including supramarginal and angular gyrus)	56090	8.3	21	2.4	0.14 (±0.11)	0.21 (±0.15)	0.08
Parietal Lobe	Remainder of parietal lobe right (including supramarginal and angular gyrus)	55452	4	18	1.1	0.10 (±0.10)	0.21 (±0.15)	0.11
Parietal Lobe	Superior parietal gyrus left	53683	3.6	28.9	2.3	0.11 (±0.09)	0.26 (±0.16)	0.15
Parietal Lobe	Superior parietal gyrus right	59456	2.8	32.1	1.6	0.12 (±0.09)	0.28 (±0.16)	0.16
Occipital Lobe	Cuneus left	15193	3.2	40.6	2.5	0.12 (±0.09)	0.31 (±0.17)	0.19
Occipital Lobe	Cuneus right	14886	6.6	51.6	5.8	0.13 (±0.10)	0.34 (±0.15)	0.21
Occipital Lobe	Lateral remainder of occipital lobe left	61105	0.6	35.3	0.3	0.07 (±0.07)	0.30 (±0.16)	0.23
Occipital Lobe	Lateral remainder of occipital lobe right	61419	0.4	21.9	0.3	0.07 (±0.08)	0.24 (±0.15)	0.17
Occipital Lobe	Lingual gyrus left	18574	8.3	65.7	7.9	0.15 (±0.10)	0.40 (±0.14)	0.25

TABLE 2 CONTINUED: Regional heritability estimates

Lobe	Structure	N-voxels	Population significant (%)	Family significant (%)	Significant both (%)	Population (±SD)	Family (±SD)	Δ h2
Occipital Lobe	Lingual gyrus right	19412	10.5	60.2	6.1	0.15 (±0.11)	0.37 (±0.14)	0.22
Insula and Cingulate gyri	Cingulate gyrus, anterior (supragenual) part left	12284	6.3	27.6	1.3	0.14 (±0.10)	0.27 (±0.09)	0.13
Insula and Cingulate gyri	Cingulate gyrus, anterior (supragenual) part right	13770	1.1	23.4	0.1	0.13 (±0.08)	0.25 (±0.11)	0.12
Insula and Cingulate gyri	Cingulate gyrus, posterior part left	12016	13.2	40.1	6.9	0.17 (±0.12)	0.32 (±0.14)	0.16
Insula and Cingulate gyri	Cingulate gyrus, posterior part right	12643	6.7	51.9	5.6	0.16 (±0.09)	0.36 (±0.14)	0.2
Insula and Cingulate gyri	Insula left	20040	8.5	62.5	4.5	0.16 (±0.10)	0.43 (±0.16)	0.27
Insula and Cingulate gyri	Insula right	20796	20.6	46.3	11.4	0.22 (±0.10)	0.33 (±0.15)	0.11
Frontal Lobe	Anterior orbital gyrus left	9240	0.1	25	0.1	0.07 (±0.07)	0.25 (±0.14)	0.18
Frontal Lobe	Anterior orbital gyrus right	8652	0	11.4	0	0.10 (±0.08)	0.23 (±0.20)	0.13
Frontal Lobe	Inferior frontal gyrus left	26296	1.3	33.6	0.3	0.08 (±0.08)	0.30 (±0.19)	0.22
Frontal Lobe	Inferior frontal gyrus right	24054	0.7	11.3	0.2	0.07 (±0.08)	0.21 (±0.16)	0.14
Frontal Lobe	Lateral orbital gyrus left	5264	0	59.3	0	0.04 (±0.05)	0.36 (±0.11)	0.32
Frontal Lobe	Lateral orbital gyrus right	5348	0.3	42	0.1	0.09 (±0.10)	0.30 (±0.15)	0.21
Frontal Lobe	Medial orbital gyrus left	8408	3.4	40.2	3.3	0.09 (±0.10)	0.32 (±0.15)	0.24
Frontal Lobe	Medial orbital gyrus right	8486	4.6	27.2	3.9	0.10 (±0.10)	0.28 (±0.21)	0.18
Frontal Lobe	Middle frontal gyrus left	53570	0.9	12.4	0.2	0.06 (±0.08)	0.21 (±0.17)	0.15
Frontal Lobe	Middle frontal gyrus right	56150	1.4	9.5	0.2	0.07 (±0.08)	0.19 (±0.15)	0.12
Frontal Lobe	Posterior orbital gyrus left	7098	0	58.8	0	0.09 (±0.07)	0.41 (±0.20)	0.32
Frontal Lobe	Posterior orbital gyrus right	7469	8.5	38.6	5.5	0.14 (±0.11)	0.31 (±0.20)	0.18
Frontal Lobe	Precentral gyrus left	44289	0.3	18.1	0	0.06 (±0.07)	0.22 (±0.14)	0.16
Frontal Lobe	Precentral gyrus right	45058	1.3	24.1	0.8	0.09 (±0.09)	0.22 (±0.16)	0.13
Frontal Lobe	Pre-subgenual anterior cingulate gyrus left	1888	0	84.3	0	0.08 (±0.05)	0.41 (±0.08)	0.32
Frontal Lobe	Pre-subgenual anterior cingulate gyrus right	1166	2.6	90	0.3	0.14 (±0.07)	0.47 (±0.10)	0.33
Frontal Lobe	Straight gyrus (gyrus rectus) left	5397	16.6	69.7	10.4	0.18 (±0.11)	0.35 (±0.10)	0.17
Frontal Lobe	Straight gyrus (gyrus rectus) right	6467	6.7	41.5	3.5	0.17 (±0.11)	0.28 (±0.12)	0.11
Frontal Lobe	Subcallosal area left	527	8	87.7	8	0.19 (±0.07)	0.51 (±0.11)	0.32
Frontal Lobe	Subcallosal area right	496	4.2	92.5	4.2	0.16 (±0.09)	0.53 (±0.08)	0.36
Frontal Lobe	Subgenual anterior cingulate gyrus left	1932	1.1	73.3	1.1	0.14 (±0.08)	0.42 (±0.10)	0.28
Frontal Lobe	Subgenual anterior cingulate gyrus right	2044	1.9	60.5	1.9	0.16 (±0.09)	0.37 (±0.13)	0.21
Frontal Lobe	Superior frontal gyrus left	69674	0.9	13.4	0.3	0.08 (±0.08)	0.20 (±0.16)	0.12
Frontal Lobe	Superior frontal gyrus right	73393	1.3	14.2	0.4	0.09 (±0.08)	0.22 (±0.19)	0.13
Central Structures	Caudate nucleus left	6042	68.6	96.4	68.6	0.32 (±0.10)	0.65 (±0.13)	0.33

TABLE 2 CONTINUED: Regional heritability estimates

Lobe	Structure	N-voxels	Population significant (%)	Family significant (%)	Significant both (%)	Population (±SD)	Family (±SD)	Δ h2
Central Structures	Caudate nucleus right	6172	75.6	91.7	72.4	0.34 (±0.09)	0.55 (±0.15)	0.21
Central Structures	Nucleus accumbens left	580	29.7	92.2	29.7	0.24 (±0.10)	0.54 (±0.15)	0.31
Central Structures	Nucleus accumbens right	443	6.1	94.6	6.1	0.17 (±0.08)	0.53 (±0.09)	0.37
Central Structures	Pallidum (globus pallidus) left	1893	32.2	89.2	32.2	0.22 (±0.09)	0.50 (±0.11)	0.28
Central Structures	Pallidum (globus pallidus) right	1812	39.8	70.3	28.5	0.27 (±0.07)	0.37 (±0.11)	0.11
Central Structures	Putamen left	7819	32.6	96.2	32.6	0.24 (±0.07)	0.67 (±0.14)	0.43
Central Structures	Putamen right	7325	59.2	90.8	57.5	0.29 (±0.07)	0.48 (±0.14)	0.19
Central Structures	Thalamus left	11142	15.2	76.2	14.4	0.18 (±0.10)	0.40 (±0.11)	0.22
Central Structures	Thalamus right	10774	6.8	68.9	6.6	0.16 (±0.09)	0.39 (±0.13)	0.23

The average voxel-wise heritability in 72 brain structures from family-based and population-based estimates. N = total number of tested voxels, Significant Population (%) = percentage of N which was significant at False discovery Rate threshold q=0.05 in the population-based estimates. Significant family (%) = percentage of N which was significant at False discovery Rate threshold q=0.05 in family-based estimates. Significant both (%) = percentage of N which was significant at False discovery Rate threshold q=0.05 in both family-based as population-based estimates. Family = mean family-based heritability estimates per brain structure, Population = mean population-based heritability estimates per brain structure. Δ h2 = difference in regional heritability (Family-Population).

ENHANCEMENT OF ASSOCIATION SIGNAL

We explored if applying our heritability map could enhance the statistical association signal of previously discovered genome-wide significant loci. As expected the T-allele of rs77956314 (*HRK*) associated with a smaller total volume of the hippocampus ($p = 5.1$ x 10^{-7}) and the C-allele of rs945270 (*KTN1*) significantly associated with larger total volume of the putamen ($p = 4.3$ x 10^{-3}). When excluding the less heritable voxels the average heritability in the re-maining voxels increased **(Figure 3A and 3B)**. With rising average heritability we observed a gradual decrease in p-values (Figure 3C), and consequently a more significant association of HRK with the more heritable part of the hippocampus**.** The maximum enrichment of associati-on was reached when the 10% most significantly heritable voxels when combining heritability information from family-based and population-based studies was used. This increase corres-ponds to a 95.9 times more significant association, as the *p*-value decreased from $p=5.1$ x 10^{-7} to $p = 5.4$ x 10^{-9}. Using only the family-based estimates the association was 12.9 times more significant. A less substantial decrease in p-value was observed for the association of KTN1 with the more heritable part of the putamen **(Figure 3D).** The p-value decreased when restricting to voxels that belong to the 25% most heritable voxels from the only the family-ba-sed study. This corresponds to a 5.5 times more significant association (*p*-value decrease from $p = 4.3$ x 10^{-3} to $p = 7.9$ x 10^{-4}).

DISCUSSION

In this study we presented grey matter voxel heritability maps at resolution of $1\times1\times1$ mm from population- and family-based studies. First we found that clusters of voxels that are significantly heritable in family-based heritability estimates as well as in an unrelated population-based study are predominantly located in subcortical regions. Second, when comparing the overall regional patterns of voxel-wise heritability the family-based estimates were always higher compared to population-based estimates and predicted the population-based heritability estimates. Lastly, we showed that the heritability estimates from our studies could be used to enhance the association signal of two genetic variants with subcortical volumes.

Voxels with significant heritability formed clusters within mainly the subcortical structures. This is in line with the findings of previous studies that the volumes of subcortical structure are among the most heritable in the brain.[40] There are multiple explanations for this consistent finding. First, subcortical structures probably are under tight genetic control as they exert vital functions within the brain. The percentage of significantly heritable voxels was relatively low in the frontal and parietal lobes. Although intra-individual measurability was high throughout the brain (Supplementary Figure 1), intra-individual differences in cortical folding patterns could explain the lower heritability in frontal and parietal regions. These might give a reliable measurability of the voxels, while it makes comparisons of voxel values between individuals less meaningful, thus yielding a lower heritability compared with the subcortical structures. Finally, environmental effects could have a larger effect on cortical grey matter compared to subcortical structures. As the effects of non-genetic factors (e.g. lifestyle factors) accumulate over an individual's lifetime, the heritability of total brain volume and brain structures volume was found to reduce in adulthood up until old age[41] in line with the accumulation of environmental influences over age. Their reported maximum age was 70 years. We studied relatively old participants (~65 years), therefore study participants might have reduced estimated heritability because of their older age.

Apart from the subcortical structures, we found three cortical regions in the left hemisphere, the dominant hemisphere in over 95% of individuals, involved in speech production and word processing to have more than 5% significant voxels; the subcallosal area (also called Broca area), central part of the superior temporal gyrus (contains Wernicke's area) and the lingual gyrus. Moreover, their right counterparts contained less significant voxels compared to the left side. Language skills[42] and brain networks [43] are thought to be under tight genetic control and the left hemisphere language areas have been found more heritable than the right hemisphere before.[14] Regions with significant heritability could in theory be connected by white matter connections, which in turn then also are under high genetic control, suggesting a common

genetic architecture. In a recent report evidence for this theory was found.[44] Cortical thickness in some regions with high heritability, were connected by heritable white matter connections. These connections and the cortical regions were anatomically distant but showed significant genetically correlation.[44]

We found a relatively stable difference in the regional patterns of the total additive genetic heritability. The heritability calculated from familial relations was always higher than the total additive variance explained by all autosomal variants calculated in unrelated subjects. This known difference between family and population-based heritability estimates has been extensively described.[45,46] The difference can in part be explained by overestimation of heritability in families due to sharing of environmental factors within the family. These factors are interpreted as genetic effects and cause the overestimation of heritability in twin and nuclear family studies.[47] Subjects in multi-generational families share less environmental factors. Therefore multi-generational families, as ASPS-Fam and especially the ERF study, are more likely to yield an unbiased estimate of heritability. However, we assumed that all environmental factors affecting brain voxel volume are uncorrelated among family members (unique environmental effects) therefore some unassessed common environmental effects might be causing the higher heritability in our family-based estimates. At the same time an underestimation of the heritability calculated from genetic data in unrelated populations could occur because of an incomplete coverage of the causal variants and exclusion of rare variants. We used imputed data to increase coverage of the causal variants. Imputed data provide a much denser coverage of the genome than only genotyped variants, but we did exclude rare variants (MAF<0.01) which may in part be responsible for some missing heritability.

The overall regional patterns of heritability from families strongly predicted the population-based heritability. This suggests that the regional pattern of variance explained by additive genetic effects is similar across populations, despite different ways to measure heritability, study design and scanner types. On the website (http://www.imagene.nl/heritability) both the population-based estimates and the family-based estimates can be viewed separately and can be downloaded. Combining current maps with results from other studies will further increase accuracy of the heritability estimates.

HERITABILITY IN GENETIC STUDIES

Within the putamen and hippocampus we observed highly heritable clusters of grey matter voxels alternating with parts of the subcortical structures that were less heritable. Differences in heritability within structures might be due to technical limitations (e.g. voxels that are difficult to measure) or due to genetic or functional correlations. We hypothesized that studying the genetics of only highly heritable voxels could enhance signals in imaging genetics, either

through reducing signal to noise ratio or through studying a more genetically homogeneous trait. We picked two genetic variants with a proven and strongly replicated biological effect, identified through genome-wide association studies, on the subcortical structure volume (hippocampus, putamen) to explore if enhancement was possible [5]. We show enhancement of the statistical signal of almost hundred-fold for the association of *HRK* (rs77956314) with hippocampal volume and a five-fold increase for the association of *KTN1* (rs945270) with putamen volume. Based on Figure 3 we can deduct that for future genetic studies in both examples a maximum power for association analyses was observed using voxels with a heritability over ~0.3 from the population-based heritability estimates and a heritability over ~0.7 from family-based heritability estimates. Despite these encouraging results there are limitations of our analysis. First, we only tested two genetic variants in two subcortical structures. While we expect that the increased signal of genetic variants with more heritable voxels will not be limited to the two variants tested in current study, future studies applying this method should be performed to determine whether this truly is the case. Second, we calculated heritability estimates and genetic association of *HRK* and *KTN1* variants with voxels in the same subjects of the Rotterdam Study. As voxels with a large (technical) measurement error have lower heritability and therefore were excluded first in our analysis, the decreased measurement error of the more heritable voxels could result in the more significant association of genetic variants. In other words, the enhancement of signal is a reflection of a higher signal to noise ratio. Also a higher test re-test reliability of the highly heritable voxels, reduce signal to noise ratio. Third, we used the same data for the calculation of population-based heritability estimates and genetic testing, resulting in a possible inflation of the increase in signal due to non-independence.[48] However, when only the family-based heritability estimates were used to select the voxels for genetic associations (**Figure 3C,D**) the analyses were independent. In these analyses, we still observed an increase in the signal – and the enhancement was actually even stronger for the putamen – arguing against inflation due to non-independence. However, for the hippocampus the best enhancement was achieved using the combined sample when restricting to less than 55% the most significant voxels. While this could be due to non-independence, this is contradicted by the fact that the population-only results (i.e., fully dependent) are in fact worse at this and lower percentages. An explanation other than non-independence could be that the combined sample provides more accurate heritability estimates and therefore results in a better enhancement. Last, highly heritable voxels which are in close proximity of each other could share their genetic background. However finding a cluster of heritable voxels does not directly prove genetic correlation.

STRENGTHS AND LIMITATIONS

Major strengths of this study are the large sample size of the population based study and unified imaging processing. Subjects from ERF and the Rotterdam Study subjects were scan-

ned using the same 1.5T scanner, identical MRI protocols and images were processed with exactly the same software. The ASPS-Fam was scanned on a 3T scanner, but segmented using similar protocols and VBM processing was performed in the same way as ERF and the Rotterdam Study. Important to note is that softwares used for tissue segmentation are different, but both implement the same kNN algorithm.[26] The ERF and the Rotterdam Study both are both from the Netherlands, a genetically homogeneous country.[49] The ASPS-Fam study is from Austria, Austrians likely have slightly different genetic architecture than the Dutch. Maximum likelihood iterative optimization was used to estimation heritability. The iterations are prone to convergence failures when sample sizes are small. The percentage of voxels that did not converge was 9% in ASPS-Fam ($N_{participants}$ = 369) and 36% in ERF ($N_{participants}$ = 122). The methods used for population-based estimation of heritability always output an estimate. It has been shown that not converging occurs frequently in small datasets in SOLAR producing conservative estimates.[47,50] We further note that using only VBM to assess heritability of brain morphology is a limitation of the current study. Cortical thickness, surface area and other MRI measures, including tensor-based (i.e. deformation) morphometry (TBM)[51,52] and shape analysis are all potentially interesting for future heritability and genetic studies. The differences between measures have been attributed both to biology[53,54] and methodology.[55,56] Most probably, these measures reflect a different genetic architecture[53] and should therefore be studied separately.

FUTURE PERSPECTIVES

Genetic association with several voxels within an anatomical structure is biologically relevant as it shows an important genetic contribution to a sub region of the structure. Apart from the biological relevance, this sub region of voxels could have clinical significance. For example, it was shown previously that subfields of the anatomically defined hippocampus contributed differently to schizophrenia[57] and β-Amyloid load.[58] If only highly heritability brain voxels are studied in future voxel-wise genome-wide association studies we do not expect statistical signals to be uniformly enhanced. However, for the tested genetic variant that was identified for putamen volume, we did find statistical enhancement. High heritability estimates capture a variety of sources that can affect power to detect associations, including lower signal to noise ratios and higher genetic homogeneity (i.e. genetic correlation). Using these benefits to increase statistical signal is desirable, irrespective of the underlying cause. Ideally we envision selecting groups of voxels for genetic studies based on high heritability and measured high genetic correlation. Genetic correlation can be calculated for any of the commonly used MRI-measures, but it would still require genetic testing of sufficiently powered (large) studies. A promising future direction would be to enable the calculation of genetic correlations, genetic association (millions of voxels times millions of genetic variants) and meta-analyses of these associations. Programs which make the calculation of genetic correlation and genetic

association computationally possible in sufficiently powered studies (i.e. meta-analyses) are essential to the field. Currently these programs tailored to large scale genetic studies are developed and genetic studies started.[59] The results of these studies will be able to prove to which extend clusters of heritable voxels have a common genetic architecture.

CONCLUSIONS

Heritability estimates can be reliably estimated using different methods and on different cohorts and combining heritability estimates from multiple studies leads to the construction of a reliable heritability map of grey matter. These maps can be used to prioritize highly heritable regions in future genetic imaging studies.

ACKNOWLEDGEMENTS

REFERENCES

1. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* 2008; **9**(4): 255-66.

2. Visscher PM, Medland SE, Ferreira MA, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet* 2006; **2**(3): e41.

3. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42**(7): 565-9.

4. Polderman TJ, Benyamin B, de Leeuw CA, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015; **47**(7): 702-9.

5. Hibar DP, Stein JL, Renteria ME, et al. Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**(7546): 224-9.

6. Jahanshad N, Kochunov PV, Sprooten E, et al. Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA-DTI working group. *Neuroimage* 2013; **81**: 455-69.

7. Jahanshad N, Lee AD, Barysheva M, et al. Genetic influences on brain asymmetry: a DTI study of 374 twins and siblings. *Neuroimage* 2010; **52**(2): 455-69.

8. Kochunov P, Jahanshad N, Marcus D, et al. Heritability of fractional anisotropy in human white matter: a comparison of Human Connectome Project and ENIGMA-DTI data. *Neuroimage* 2015; **111**: 300-11.

9. Kochunov P, Glahn DC, Lancaster JL, et al. Genetics of microstructure of cerebral white matter using diffusion tensor imaging. *Neuroimage* 2010; **53**(3): 1109-16.

10. Ganjgahi H, Winkler AM, Glahn DC, Blangero J, Kochunov P, Nichols TE. Fast and powerful heritability inference for family-based neuroimaging studies. *Neuroimage* 2015; **115**: 256-68.

11. Kochunov P, Fu M, Nugent K, et al. Heritability of complex white matter diffusion traits assessed in a population isolate. *Hum Brain Mapp* 2016; **37**(2): 525-35.

12. Brouwer RM, Mandl RC, Peper JS, et al. Heritability of DTI and MTR in nine-year-old children. *Neuroimage* 2010; **53**(3): 1085-92.

13. Hulshoff Pol HE, Schnack HG, Posthuma D, et al. Genetic contributions to human brain morphology and intelligence. *J Neurosci* 2006; **26**(40): 10235-42.

14. Thompson PM, Cannon TD, Narr KL, et al. Genetic influences on brain structure. *Nat Neurosci* 2001; **4**(12): 1253-8.

15. Peper JS, Schnack HG, Brouwer RM, et al. Heritability of regional and global brain structure at the onset of puberty: a magnetic resonance imaging study in 9-year-old twin pairs. *Hum Brain Mapp* 2009; **30**(7): 2184-96.

16. Bis JC, DeCarli C, Smith AV, et al. Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* 2012; **44**(5): 545-51.

17. Stein JL, Medland SE, Vasquez AA, et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 2012; **44**(5): 552-61.

18. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015; **30**(8): 661-708.

19. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008; **40**(2): 161-9.

20. Ikram MA, van der Lugt A, Niessen WJ, et al. The Rotterdam Scan Study: design update 2016 and main findings. *Eur J Epidemiol* 2015; **30**(12): 1299-315.

21. Sayed-Tabatabaei FA, van Rijn MJ, Schut AF, et al. Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. *Stroke* 2005; **36**(11): 2351-6.

22. Ibrahim-Verbaas CA, Zorkoltseva IV, Amin N, et al. Linkage analysis for plasma amyloid beta levels in persons with hypertension implicates Abeta-40 levels to presenilin 2. *Hum Genet* 2012; **131**(12): 1869-76.

23. Schmidt R, Lechner H, Fazekas F, et al. Assessment of cerebrovascular risk profiles in healthy persons: definition of research goals and the Austrian Stroke Prevention Study (ASPS). *Neuroepidemiology* 1994; **13**(6): 308-13.

24. Schmidt R, Fazekas F, Kapeller P, Schmidt H, Hartung HP. MRI white matter hyperintensities: three-year follow-up of the Austrian Stroke Prevention Study. *Neurology* 1999; **53**(1): 132-9.

25. Seiler S, Pirpamer L, Hofer E, et al. Magnetization transfer ratio relates to cognitive impairment in normal elderly. *Front Aging Neurosci* 2014; **6**: 263.

26. Vrooman HA, Cocosco CA, van der Lijn F, et al. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage* 2007; **37**(1): 71-81.

27. Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* 2001; **14**(1 Pt 1): 21-36.

28. Roshchupkin GV, Adams HH, van der Lee SJ, et al. Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology. *Neurobiol Aging* 2016; **48**: 204-11.

29. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 2004; **23 Suppl 1**: S208-19.

30. Fonov V, Evans AC, Botteron K, et al. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* 2011; **54**(1): 313-27.

31. Hammers A, Allom R, Koepp MJ, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp* 2003; **19**(4): 224-47.

32. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**(2): 420-8.

33. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**(1): 76-82.

34. Visscher PM, Yang J, Goddard ME. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). *Twin Res Hum Genet* 2010; **13**(6): 517-24.

35. Adams HH, Verlinden VJ, Callisaya ML, et al. Heritability and Genome-Wide Association Analyses of Human Gait Suggest Contribution of Common Variants. *J Gerontol A Biol Sci Med Sci* 2016; **71**(6): 740-6.

36. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**(5): 1198-211.

37. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190-1.

38. Blangero J, Williams JT, Almasy L. Variance component methods for detecting complex trait loci. *Adv Genet* 2001;

42: 151-81.

39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 1995; **57**(1): 289-300.

40. Blokland GA, de Zubicaray GI, McMahon KL, Wright MJ. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Res Hum Genet* 2012; **15**(3): 351-71.

41. Batouli SA, Trollor JN, Wen W, Sachdev PS. The heritability of volumes of brain structures and its relationship to age: a review of twin and family studies. *Ageing Res Rev* 2014; **13**: 1-9.

42. Gayan J, Olson RK. Reading disability: evidence for a genetic etiology. *Eur Child Adolesc Psychiatry* 1999; **8 Suppl 3**: 52-5.

43. Budisavljevic S, Dell'Acqua F, Rijsdijk FV, et al. Age-Related Differences and Heritability of the Perisylvian Language Networks. *J Neurosci* 2015; **35**(37): 12625-34.

44. Shen KK, Dore V, Rose S, et al. Heritability and genetic correlation between the cerebral cortex and associated white matter connections. *Hum Brain Mapp* 2016; **37**(6): 2331-47.

45. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 2012; **109**(4): 1193-8.

46. Zuk O, Schaffner SF, Samocha K, et al. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A* 2014; **111**(4): E455-64.

47. Koran ME, Thornton-Wells TA, Jahanshad N, et al. Impact of family structure and common environment on heritability estimation for neuroimaging genetics studies using Sequential Oligogenic Linkage Analysis Routines. *J Med Imaging (Bellingham)* 2014; **1**(1): 014005.

48. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 2009; **12**(5): 535-40.

49. Boomsma DI, Wijmenga C, Slagboom EP, et al. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 2014; **22**(2): 221-7.

50. Blangero J, Diego VP, Dyer TD, et al. A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees. *Adv Genet* 2013; **81**: 1-31.

51. Brun CC, Lepore N, Pennec X, et al. Mapping the regional influence of genetics on brain structure variability--a tensor-based morphometry study. *Neuroimage* 2009; **48**(1): 37-49.

52. Yoon U, Perusse D, Lee JM, Evans AC. Genetic and environmental influences on structural variability of the brain in pediatric twin: deformation based morphometry. *Neurosci Lett* 2011; **493**(1-2): 8-13.

53. Winkler AM, Kochunov P, Blangero J, et al. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage* 2010; **53**(3): 1135-46.

54. Voets NL, Hough MG, Douaud G, et al. Evidence for abnormalities of cortical development in adolescent-onset schizophrenia. *Neuroimage* 2008; **43**(4): 665-75.

55. Hutton C, Draganski B, Ashburner J, Weiskopf N. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *Neuroimage* 2009; **48**(2): 371-80.

56. Blankstein U, Chen JYW, Mincic AM, McGrath PA, Davis KD. The complex minds of teenagers: Neuroanatomy of personality differs between sexes. *Neuropsychologia* 2009; **47**(2): 599-603.

57. Kuhn S, Musso F, Mobascher A, Warbrick T, Winterer G, Gallinat J. Hippocampal subfields predict positive symptoms in schizophrenia: first evidence from brain morphometry. *Transl Psychiatry* 2012; **2**: e127.

58. Schroeder C, Park MT, Germann J, et al. Hippocampal shape alterations are associated with regional Abeta load in cognitively normal elderly individuals. *Eur J Neurosci* 2016.

59. Roshchupkin GV, Adams HHH, Vernooij MW, et al. HASE: Framework for efficient high-dimensional association analyses. *Sci Rep* 2016; **6**(36076).

SUPPLEMENTARY FILES
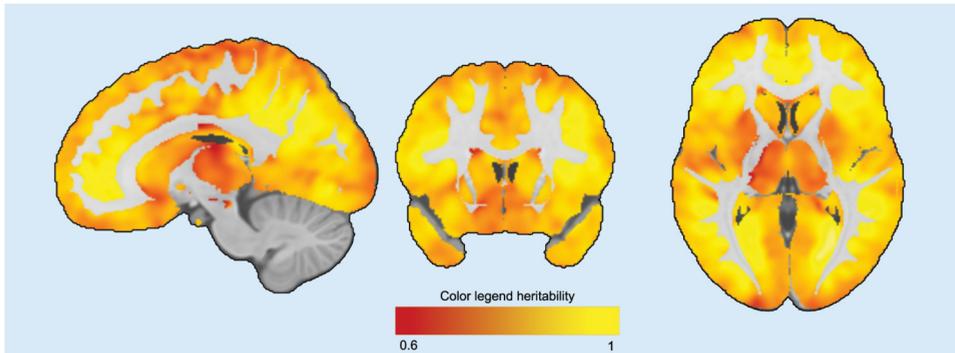
SUPPLEMENTARY FIGURE 1: Example of the intraclass correlation (ICC) in 83 individuals scanned twice within several weeks. In general voxels have a high ICC. All results can be interactive accessed (www.imagene.nl/heritability) and downloaded from the website.

CHAPTER 2.2

Heritability of the shape
of subcortical brain
structures in the general population

ABSTRACT

The volumes of subcortical brain structures are highly heritable, but genetic underpinnings of their shape remain relatively obscure. Here we determine the relative contribution of genetic factors to individual variation in the shape of 7 bilateral subcortical structures: the nucleus accumbens, amygdala, caudate, hippocampus, pallidum, putamen and thalamus. In 3,686 unrelated individuals aged between 45 and 98 years, brain magnetic resonance imaging and genotyping was performed. The maximal heritability of shape varied from 32.7% to 53.3% across the subcortical structures. Genetic contributions to shape extend beyond influences on intracranial volume and the gross volume of the respective structure. The regional variance in heritability was related to the reliability of the measurements, but could not be accounted for by technical factors only. These findings could be replicated in an independent sample of 1040 twins. Differences in genetic contributions within a single region reveal the value of refined brain maps to appreciate the genetic complexity of brain structures.

Subcortical brain regions are important for a multitude of biological processes, including cognitive and motor functions.[1,2] There is substantial structural variation in these regions, both within the normal range[3] and in the context of various neuropsychiatric diseases.[4,5] Factors driving individual variation could provide insight into brain development, healthy aging, and pathological states, but these remain largely unknown. Variation in subcortical brain structures is affected by environmental factors, such as education, diet and stress, but a considerable proportion of the variation is determined by genes.[6,7] A recent twin study of gross subcortical volumes found heritability estimates ranging between 0.44 and 0.88,[8] which were especially high for the caudate and thalamus.

Even so, aggregate measures such as volume do not capture the complexity of subcortical structures. The hippocampus, for example, is made up of several subfields, each with partially independent functional roles. More recently, image processing methods have been developed to characterize brain structure beyond purely volumetric measures, and yielding a range of shape descriptors.[9-13] The high-dimensionality allows the detection of more localized differences in brain structure, and shape can provide relevant biological information in addition to aggregate measures.[14-17] Several genetic variants that influence the volume of subcortical structures have been identified,[18-20] but their effect could be localized to certain sub-regions using shape analyses.[19,20] However, the extent to which genes contribute to the variability in shape of subcortical structures has yet to be determined.

Here, we quantify genetic influences on shape variability of 14 subcortical brain structures in 3,686 unrelated individuals from the population-based Rotterdam Study. We compare the heritability of vertex-wise shape measures to gross volumes as well as other aggregate measures of shape obtained through dimension-reduction techniques. We show that the shape of subcortical structures is under genetic control, and investigate the relation of the resulting profiles with the gross volume and measures of reproducibility.

METHODS

STUDY POPULATION

This work was performed in the Rotterdam Study,[21] a population-based cohort study in the Netherlands including a total of 14,926 participants (aged 45 years or over at enrollment). The overall aim of the study is to investigate causes and determinants of chronic diseases in elderly people, the participants were not selected for the presence of diseases or risk factors. Since 2005, all participants underwent brain magnetic resonance imaging (MRI) to examine the causes and consequences of age-related brain changes.[22] Between 2005 and 2013, a total of 5,691 unique persons were scanned. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Wet Bevolkingsonderzoek: ERGO (Population Studies Act: Rotterdam Study). All participants provided written informed consent to participate in the study and to obtain information from their treating physicians.

Replication was performed in 1040 healthy young adult twins from the Queensland Twin IMaging (QTIM) project [de Zubicaray et al. 2008]. All participants of the imaging sample were Caucasian and right-handed for throwing and writing (Annett's Handedness Questionnaire). The genetic analyses were conducted in the 350 complete twin pairs (n = 700): 148 monozygotic (100 male), 120 dizygotic (39 male), and 82 opposite-sex pairs. Self-reported data was used to screen participants for contraindications for imaging as well as any significant medical, psychiatric or neurological conditions, history of substance abuse and current use of psychoactive medication. The study was approved by the Human Research Ethics Committees of the Queensland Institute of Medical Research, the University of Queensland, and Uniting Health Care, Wesley Hospital. Informed consent was obtained from each participant and parent or guardian for participants under 18 years of age.

GENOTYPING AND IMPUTATION

Genotyping in the Rotterdam Study was performed using the Illumina 550K and 550K duo arrays.[21] Subsequently, we removed samples with call rate below 97.5%, gender mismatch, excess autosomal heterozygosity, duplicates or family relations and ancestry outliers, and variants with call rate below 95.0%, failing missingness test, Hardy–Weinberg equilibrium p-value $< 10^{-6}$, and minor allele frequency $< 1\%$. Genotypes were imputed using MACH/minimac software[23] to the 1000 Genomes phase I version 3 reference panel (all population). For QTIM, genotyping of nine markers was used to determine the zygosity of same-sex twins, which was later confirmed for >92% of the sample with the Illumina 610K SNP array.

IMAGE ACQUISITION

For the Rotterdam Study, MRI scanning was done on a 1.5-T MRI unit with a dedicated eight-channel head coil (GE Healthcare). The MRI protocol consisted of several high-resolution axial sequences, including a T1-weighted sequence (slice thickness 0.8 mm), which was used for further image processing. In addition, 85 persons were rescanned within days to weeks after the first scan to estimate the reproducibility of imaging-derived measures. A detailed description of the MRI protocol was presented by Ikram *et al.*[22]

The twin pairs of QTIM were scanned on a 4T Bruker Medspec (Bruker, Germany) whole body MRI system paired with a transverse electromagnetic (TEM) head coil. Structural T1-weighted 3D images were acquired (TR=1500ms, TE=3.35ms, TI=700ms, 240mm FOV, 0.9mm slice thickness, 256 or 240 slices depending on acquisition orientation (86% coronal (256 slices), 14% sagittal (240 slices)).

IMAGE PROCESSING

The T1-weighted MRI scans were processed using FreeSurfer[24] (version 5.1) to obtain segmentations and volumetric summaries of the following 7 subcortical structures for each hemisphere: nucleus accumbens, amygdala, caudate, hippocampus, pallidum, putamen, and thalamus (**Figure 1A**).

Next, segmentations were processed using a previously described shape analysis pipeline.[9,10] Briefly, a mesh model was created for the boundary of each structure. Subcortical shapes were registered using the "Medial Demons" framework, which matches shape curvatures and medial features to a pre-computed template.[25,26] To do this, a medial model of each individual surface model is fit following Gutman *et al.*[27], and medial as well as intrinsic features of the shape drive registration to a template parametrically on the sphere. To minimize metric distortion, the registration was performed in the fast spherical demons framework.[10] The templates and mean medial curves were previously constructed and are distributed as part of the ENIGMA-Shape package (http://enigma.ini.usc.edu/ongoing/enigma-shape-analysis/). The resulting meshes for the 14 structures consist of a total of 27,120 vertices (**Figure 1A**). For these vertices, two measures were used to quantify shape: the radial distance and the natural logarithm of the Jacobian determinant. The radial distance represents the distance of the vertex from the medial curve of the structure (**Figure 1B**). The Jacobian determinant captures the deformation required to map the subject-specific vertex to a template and indicates surface dilation due to sub-regional volume change (**Figure 1C**). Detailed information is provided in the **Supplementary Material**.

Finally, we performed 28 principal component analyses: for each of the 14 subcortical struc-

tures and for both types of shape measures (radial distance and Jacobian determinant), we computed the full set of components. This yielded the same number of principal components as the original number of vertices that were used to describe shape (**Figure 1A**). The components were sorted in descending order of the eigenvalues, which corresponds to the amount of explained variance of shape.

HERITABILITY ESTIMATION

We used Massively Expedited Genome-wide Heritability Analysis (MEGHA)[28] to estimate heritability in our sample of unrelated individuals. This method allows fast and accurate estimates of heritability across thousands of phenotypes based on genome-wide genotype data of common genetic variants from unrelated individuals. As previously described,[28] a genetic relationship matrix was constructed using the 1000 Genomes imputed genotypes, filtered on imputation quality ($R^2 < 0.5$) and allele frequency ($MAF < 0.01$). We calculated pairwise genetic relatedness between all individuals. We removed one person for pairs with more than 0.025 genotype similarity, resulting in a final study population of 3,686 subjects.

Twin-based heritability was estimated using maximum-likelihood variance components methods implemented in the SOLAR software (version 6.6.2). [29] To test the hypothesis that no variance can be explained genetically, log likelihoods of models with no genetic components were compared to those with genetic and environmental components. As twice the log likelihood is distributed as a mixture of chi-squared distributions, the hypothesis test and p-value can be derived parametrically.[29]

To correct for multiple comparisons across all vertices and all structures, we used the standard False Discovery Rate (FDR) threshold at q=0.05 to localize regions of significant heritability within each of the subcortical structures.[30]
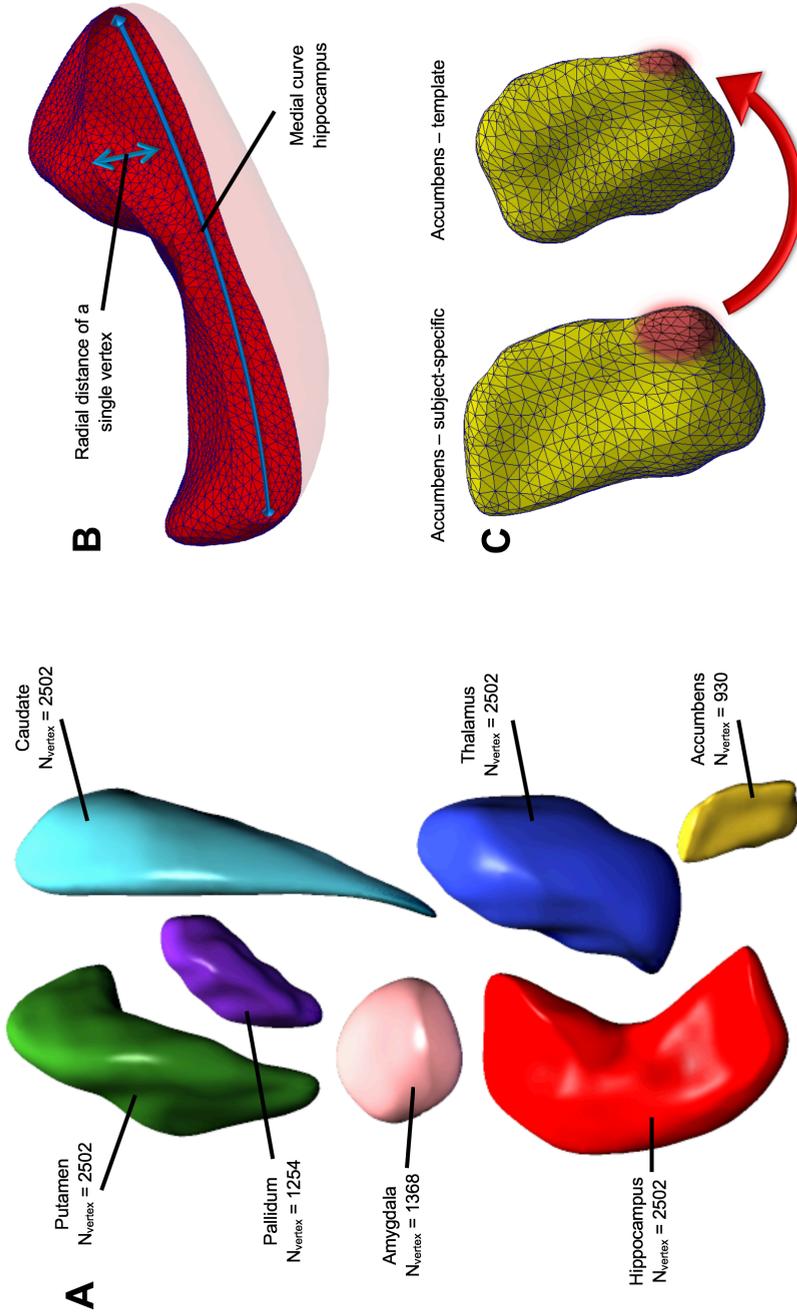
FIGURE 1: Subcortical brain structures and the derivation of shape measures.

Overview of the subcortical brain structures studied in this manuscript and the derivation of the shape measures. Panel A shows the seven structures with corresponding number of vertices: accumbens, amygdala, caudate, hippocampus, pallidum, putamen, and thalamus. Panels B and C illustrate the two vertex-wise measures of shape: the radial distance is defined as the distance of a vertex to the medial curve of the structure, e.g. the hippocampus in Panel B. The Jacobian determinant captures the deformation that is needed to map a subject-specific shape to a template, which is shown with an example of the accumbens in Panel C.

RESULTS

STUDY POPULATION

The characteristics of the study population are shown in **Table 1**. The mean age of the Rotterdam study population was 65.9 ± 10.9 years, and 55.0% were women. For the 14 subcortical structures, the mean volumes were between 0.49 and 6.25 mL. For the QTIM study, mean age was 22.9 ± 2.8 years, and 61.6% were women. Mean subcortical volumes were higher than in the Rotterdam study across the board, ranging from 0.79 and 7.82 mL.

TABLE 1: Characteristics of the study population.

Characteristic	Rotterdam Study (N = 3,686)	QTIM (N = 1,040)
Age, mean (SD), years	65.9 (10.9)	22.9 (2.8)
Female sex, n (%)	2,029 (55.0%)	641 (61.6%)
Intracranial volume, mean (SD), cm^3	1478.6 (161.3)	1484 (157.1)
Left hemisphere, mean (SD), cm^3		
Accumbens	0.56 (0.10)	0.83 (0.15)
Amygdala	1.31 (0.21)	1.84 (0.25)
Caudate	3.40 (0.56)	3.76 (0.50)
Hippocampus	3.84 (0.62)	4.32 (0.46)
Pallidum	1.47 (0.24)	1.61 (0.25)
Putamen	4.62 (0.68)	6.60 (0.72)
Thalamus	6.25 (0.79)	7.82 (0.89)
Right hemisphere, mean (SD), cm^3		
Accumbens	0.49 (0.09)	0.79 (0.11)
Amygdala	1.39 (0.22)	1.88 (0.25)
Caudate	3.51 (0.58)	3.92 (0.53)
Hippocampus	3.85 (0.59)	4.32 (0.46)
Pallidum	1.41 (0.25)	1.53 (0.18)
Putamen	4.45 (0.65)	6.00 (0.65)
Thalamus	6.25 (0.79)	7.43 (0.88)

Abbreviation: SD = standard deviation.

HERITABILITY OF SUBCORTICAL STRUCTURES: VOLUME AND SHAPE

The structure of subcortical brain regions was quantified by calculating their gross volume as well as two measures of their shape. Age- and sex-adjusted heritability estimates for the gross volume of each of the subcortical structures were between 1.6% and 43.4% (**Table 2**). For the two vertex-wise shape measures, the maximal heritability estimates per structure ranged

from 32.7% to 53.3% (**Table 2**). Both the radial distance (**Figure 2A-C**) and the Jacobian determinant (**Figure 2D-F**) showed clusters of high heritability under various models. Further adjustment for intracranial volume did not influence results (**Figure 2**), and estimates were highly correlated between both models (**Supplementary Figure 1**). The addition of the structure-specific gross volume to the model, however, did affect the heritability distribution across the structures (**Figure 2**), particularly for the shape measures that are highly correlated with the gross volume (**Supplementary Figure 2**).

TABLE 2: Heritability estimates of various structural measures of subcortical brain regions.

Region	Gross volume		Radial distance		Jacobian determinant		PCA radial distance		PCA Jacobian determinant	
	h^2	p	h^{2*}	p	h^{2*}	p	h^{2*}	p	h^{2*}	p
Left hemisphere										
Amygdala	8.1	0.18	47.7	1.72×10^{-6}	35.4	2.85×10^{-4}	29.9	4.40×10^{-4}	27.9	9.30×10^{-4}
Accumbens	11.6	0.099	34.0	4.71×10^{-4}	33.7	5.11×10^{-4}	28.7	7.04×10^{-4}	42.0	1.45×10^{-6}
Caudate	33.7	8.6×10^{-5}	49.9	6.33×10^{-7}	52.9	1.40×10^{-7}	42.4	1.20×10^{-6}	35.1	4.73×10^{-5}
Hippocampus	10.8	0.12	32.7	7.32×10^{-4}	29.2	2.23×10^{-3}	28.9	6.59×10^{-4}	29.6	5.03×10^{-4}
Pallidum	32.2	1.7×10^{-4}	39.6	5.75×10^{-5}	44.1	8.65×10^{-6}	30.8	2.96×10^{-4}	27.0	1.33×10^{-3}
Putamen	43.4	6.8×10^{-7}	49.4	7.43×10^{-7}	52.7	1.45×10^{-7}	34.1	7.16×10^{-5}	40.7	2.92×10^{-6}
Thalamus	34.1	7.4×10^{-5}	53.3	1.05×10^{-7}	45.3	5.07×10^{-6}	30.2	3.78×10^{-4}	29.4	5.26×10^{-4}
Right hemisphere										
Amygdala	20.4	0.012	33.5	5.45×10^{-4}	31.5	1.08×10^{-3}	30.5	3.45×10^{-4}	27.7	1.03×10^{-3}
Accumbens	1.6	0.43	33.1	6.30×10^{-4}	35.1	3.13×10^{-4}	34.5	5.99×10^{-5}	31.7	2.10×10^{-4}
Caudate	34.7	5.4×10^{-5}	46.7	2.86×10^{-6}	47.5	1.95×10^{-6}	29.9	4.45×10^{-4}	33.8	8.75×10^{-5}
Hippocampus	8.0	0.19	33.7	5.26×10^{-4}	17.7	4.23×10^{-2}	30.8	3.00×10^{-4}	28.9	6.44×10^{-4}
Pallidum	36.6	2.3×10^{-5}	46.4	3.12×10^{-6}	44.5	7.22×10^{-6}	41.4	1.97×10^{-6}	29.2	5.77×10^{-4}
Putamen	37.1	1.8×10^{-5}	42.6	1.70×10^{-5}	37.5	1.32×10^{-4}	32.7	1.36×10^{-4}	33.4	1.01×10^{-4}
Thalamus	30.8	3.0×10^{-4}	46.2	3.50×10^{-6}	50.4	4.50×10^{-7}	37.1	1.78×10^{-5}	31.8	2.02×10^{-4}

Estimate indicates highest heritability among all vertices or principal components.

Abbreviations: h^2 = heritability estimate in %, PCA = principal component analysis.

REPRODUCIBILITY OF SUBCORTICAL SHAPE

Next, we investigated the relation between our heritability estimates and the reproducibility of subcortical shape. In a subset of 83 persons who were scanned twice within 1-9 weeks, we quantified the reproducibility by calculating intraclass correlation coefficients for the vertex-wise shape measures (**Supplementary Figure 3**). There was considerable overlap between heritability and reproducibility (**Figure 3A-B**), and both were correlated within hemisphere (**Figure 3C-D**). Poorly reproducible shape measures were generally not heritable, whereas high reproducibility included the full range of heritability estimates (**Figure 3C-D**).
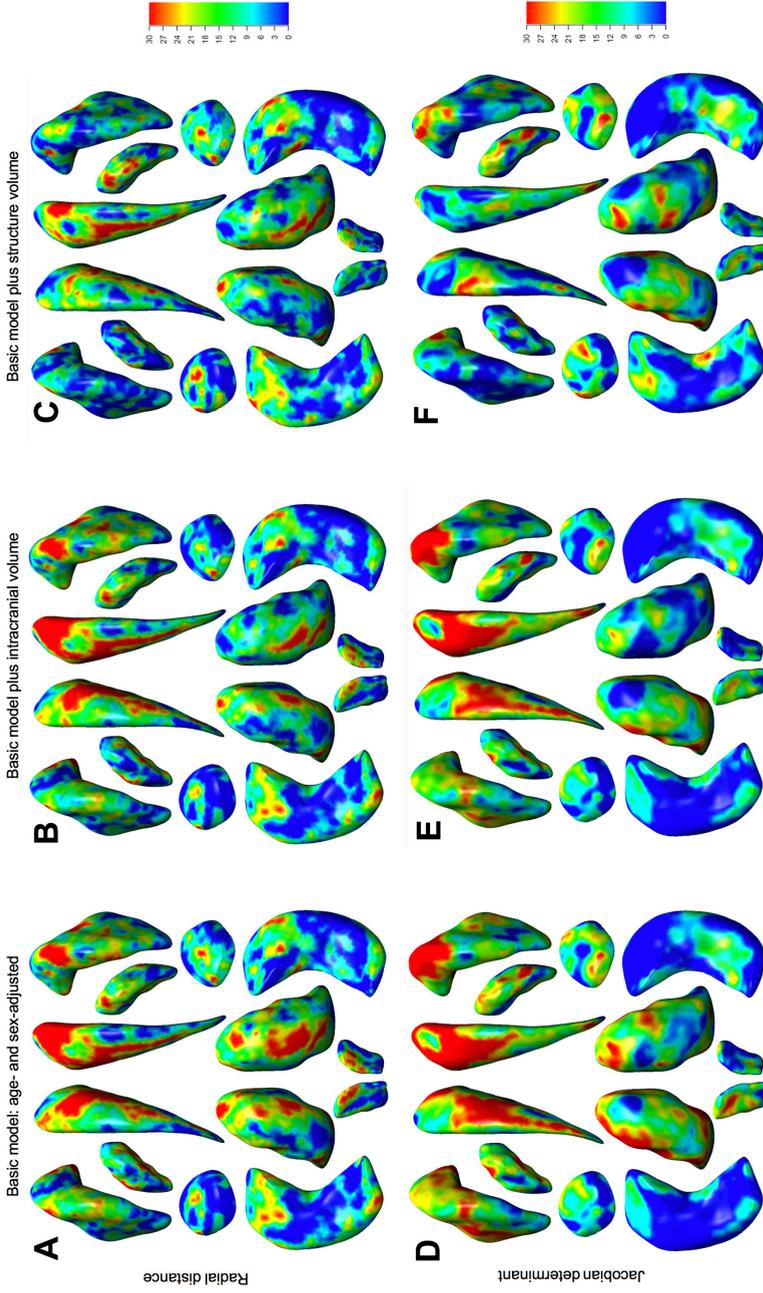
FIGURE 2: Heritability maps of shape measures of subcortical brain regions under various models.
Maps show the heritability of 7 bilateral subcortical structures for the shape measures of radial distance (Panels A-C) and the Jacobian determinant (Panels D-F). Heritability estimates were obtained using three different statistical models: a basic model with age and sex (Panels A and D), and additionally adjusting for either intracranial volume (Panels B and E) or the volume of the specific structure (Panels C and F).
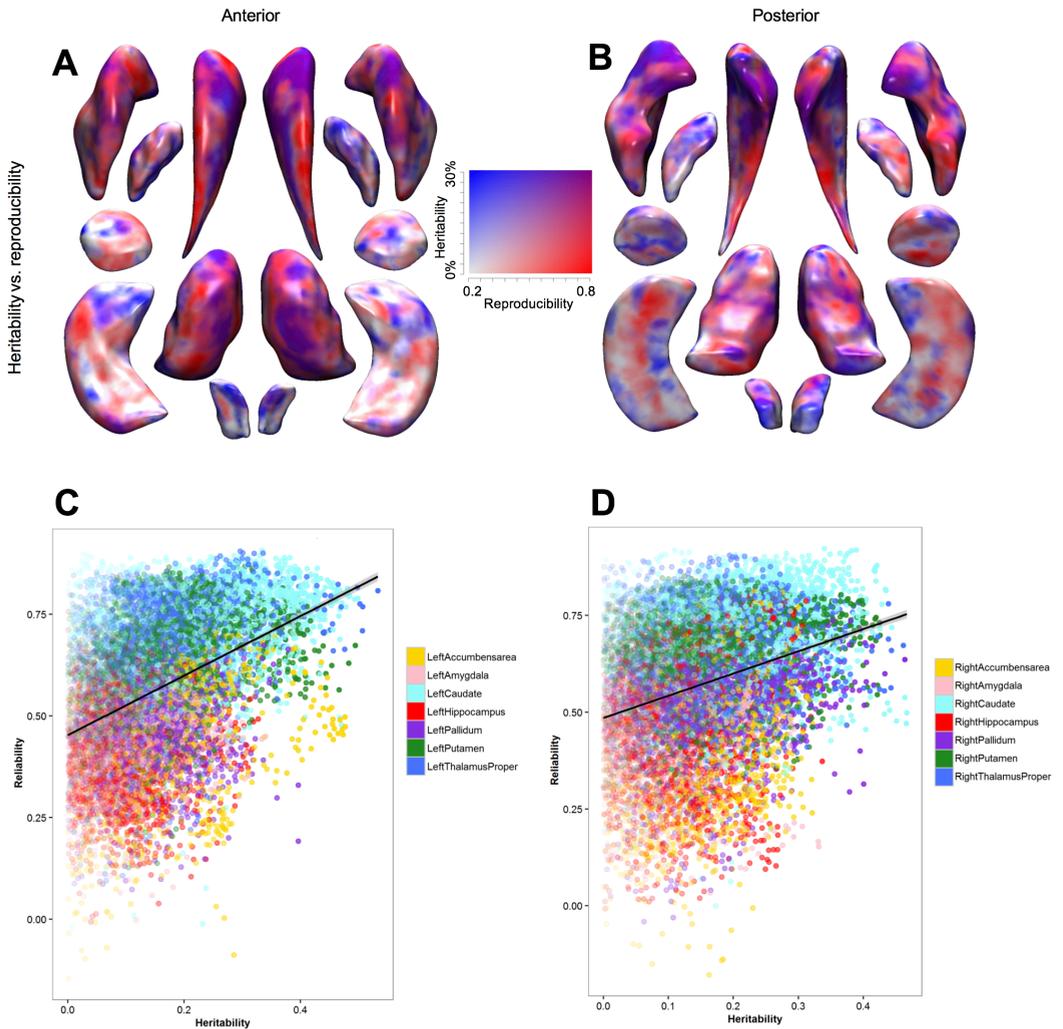
FIGURE 3: Concordance between the heritability of subcortical shape and reproducibility of the measures.

Figure showing the concordance between the heritability of the shape (radial distance) ofsubcortical structures and the reproducibililty of these measures. Maps illustrate heritability (high is red) and reproducibility (high is blue) and their overlap (purple) from the anterior (Panel A) and posterior (Panel B) direction. Scatter plots between heritability and reprodcubility of the left (Panel C) and right (Panel D) hemisphere for the 7 subcortical structures. Colors indicate the different structures (see figure legends).

HERITABILITY OF SHAPE MEASURES THROUGH DATA REDUCTION

Finally, we explored whether high-dimensional shape data could be reduced to a smaller set of variables with a larger genetic contribution. We performed principal component analyses on the two vertex-wise shape measures for each structure and computed the heritability of the resulting components. Except for the Jacobian determinant of both hippocampi, the maximal heritability was lower than for the vertex-wise measures (**Table 2**). Similarly, the components were in general less heritable than the vertex-wise measures (**Figure 4**). Furthermore, the order of the components based on the eigenvalues did not correlate well with the order based on the heritability (ρ ranges from -0.038 to 0.096; **Supplementary Table 1**).

REPLICATION OF HERITABILITY OF SUBCORTICAL STRUCTURES IN TWINS

The maximum heritability estimates for the two vertex-wise shape measures per structure ranged from 48.9% to 78.3%. Both the radial distance (**Supplementary Figure 4A-C**) and the Jacobian determinant (**Supplementary Figure 4D-F**) showed clusters of high heritability under various models. Further adjustment for intracranial volume did not influence the results (**Supplementary Figure 4C, E**). The addition of the structure-specific gross volume to the model, however, did affect the heritability distribution across the structures (**Supplementary Figure 4C, F**). Comparing the results of the twin-based and population study, we found a considerable overlap and significant correlation (p-value = 3.03×10^{-306}) in estimated heritability (**Supplementary Figure 5**).
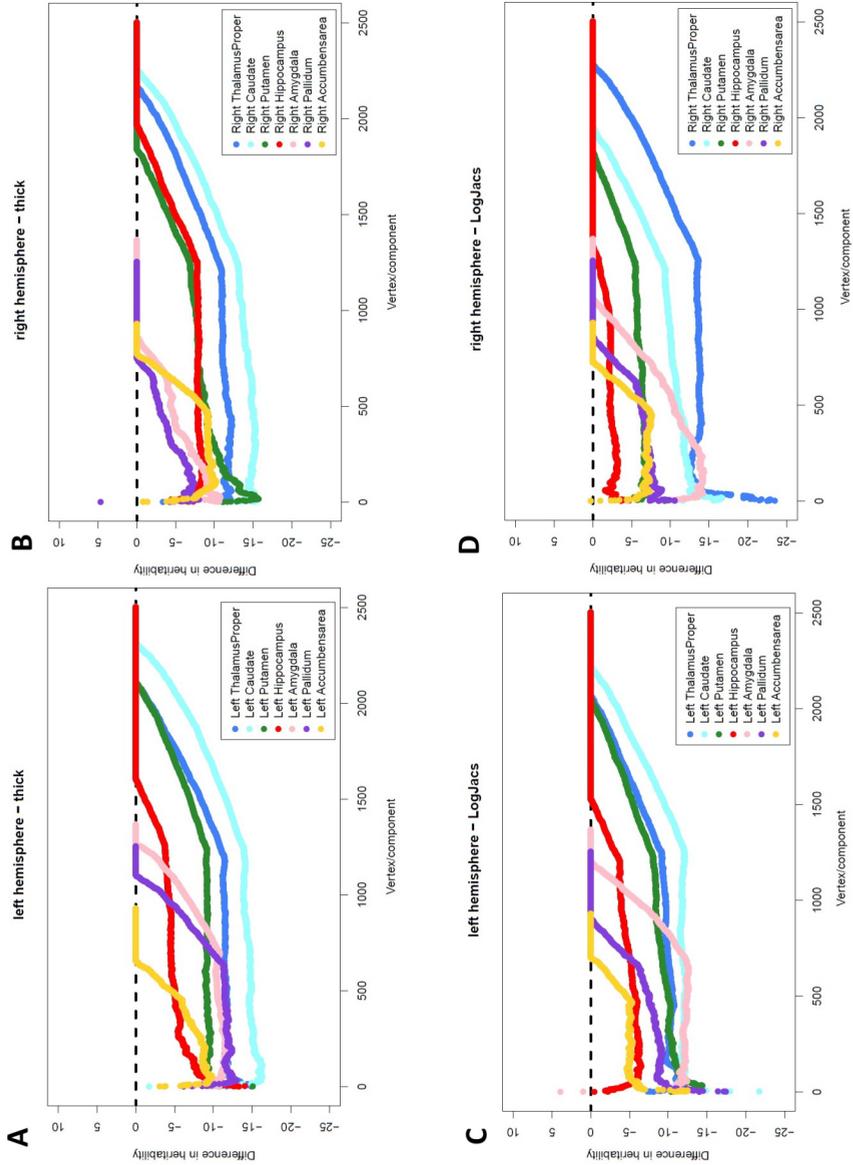
FIGURE 4: Difference in heritability between vertex-wise shape measures and PCA components.

Plots illustrate the difference between heritability estimates obtained from the vertex-wise shape measures and the heritability of the components obtained through principal component analysis for 7 bilateral subcortical structures. Separate panels are provided for the shape measures of radial distance (A-B) and the Jacobian determinant (C-D) and the left (A,C) and right (B,D) hemisphere. All vertex-wise shape measures and principal components were first sorted in descending order of heritability, and the vertex-wise measures were substracted from the corresponding component's heritability. Colors indicate the different structures (see figure legends).

DISCUSSION

Here we show that, in a general population of middle-aged and elderly individuals, the shapes of subcortical structures are under genetic control. The vertex-wise heritability is higher than for aggregate measures such as volume and principal components. Moreover, the heritability pattern underlines the importance of reproducibility in deriving shape measures, but also reveals that the extent of genetic influences is not uniformly distributed across subcortical structures. We confirmed our findings in an independent cohort of twins, suggesting that the genetic architecture of subcortical shapes is similar across populations, despite differences in the sample, the study design, scanner types, and methods to compute the heritability.

The higher vertex-wise heritability could reflect true biological differences in the degree of genetic contribution to the variability in shape. For the cerebral cortex, it has already been shown that different genes influence distinct parts of the brain and that the heritability also differs between regions.[31-33] Subcortical structures are also heterogeneous and consist of functionally diverging sub-regions, such as the nuclei of the pallidum or the head and tail of the caudate. Our results are in line with a recent study by Whelan *et al.* showing that hippocampal subfields differ in their heritability.[34] However, methodological reasons for this difference in heritability should also be considered. Particularly, a lower signal-to-noise ratio in some of the measures might have influenced the results, leading to low heritability estimates. Issues in the segmentation or registration steps will thus obscure true biological differences if these systematically affect certain sub-regions of a structure. We investigated whether this plays a role by overlapping our heritability maps with maps of the technical reproducibility. Indeed, shape measures that could be poorly reproduced were not heritable. However, while high reproducibility was required for detecting a substantial genetic component, it did not necessarily translate into a high heritability. For example, for the shape measures with a high reproducibility (intraclass correlation coefficients > 0.75), a wide range of heritability estimates was observed (0% to 53%). Thus, even when the signal-to-noise ratio was comparable, we still observed regional differences in the degree of genetic contribution. The highly heritable measures are interesting targets for more in-depth genetic studies.

Heritability estimates calculated in our analysis represent both upper and low bounds of narrow-sense heritability. Our results are consistent with the theory that twin-based heritability tends to be higher than population-based estimates. However, we did not find a high correlation between the results, which could be due to several factors. Our population study consisted of relatively older individuals, which may impact the heritability: the effects of non-genetic factors on subcortical structures (e.g., lifestyle factors) accumulate over an individual's lifetime and the overall contribution of genes might be reduced compared to younger individu-

als. Causal variants not captured on the genotyping array or through subsequent imputation also could lead to a different distribution of the heritability. Additionally, apart from array limitations, non-additive genetic factors are not taken into account when computing population based heritability. These factors should be taken into account when interpreting our results. An important question for future research on shape is which variables need to be controlled for in a regression analysis. Here, we aimed to provide an answer by studying two controversial adjustment variables: the total intracranial volume and the gross volume of the structure under study. For the heritability estimates of shape, adjustment for intracranial volume did not affect the results, suggesting that the genes regulating shape are not general brain growth genes, but rather more specific for a structure or its sub-regions. The volume adjustments did change some of the results, but more so for vertices whose shape measures correlate most with the gross volume of the structure. Likely, the genes underlying a structure's gross volume are largely driven by these vertices as they typically represent the widest parts of a structure (highest mean radial distance), where radial measures tend to be highly correlated with its volume. Our results are in agreement with previous work,[35] where the heritability of region-specific measures was reduced after adjustments for the total cortical surface area and thickness.

The detailed information provided by shape measures being their most attractive feature, the increase in dimensionality is potentially counterproductive, especially in the case of genetic homogeneity across a structure. We therefore also performed principal component analyses to demonstrate that the amount of variability explained by the components did not seem related to the heritability: near-zero correlations were found between the order of the components based on the eigenvalues and the heritability estimates. Although the principal component analysis captures most of the variation using fewer variables, methods, which are based on the genetic correlation, may lead to biologically more meaningful results.

While heritability provides an estimate of how much of the variance is determined by genetics, it does not point to specific genetic loci. The most commonly accepted method for gene discovery is to perform an unbiased screen of all genetic variants, i.e. genome-wide association study (GWAS) in order to identify specific genetic factors. However, such efforts require large-scale collaborations in the order of tens of thousands of individuals in order to identify a robust association[18-20,39]. Furthermore, additional multiple testing correction should be considered when performing GWAS of 54,000 shape measures. This could lead to a loss of power if the effects are homogeneous across a structure. However, if the effects are localized and mostly affect specific vertices, then a GWAS of shape measures may actually increase power since the effect sized will be larger compared to a GWAS of an aggregate volume.

Data reduction methods always rely on assumptions and are often aimed at resolving computational issues. However, with the advent of big data collection, methods have been developed to analyze such large datasets efficiently. Software packages designed for high-dimensional data include MEGHA,[36] for heritability analyses, BOLT-LMM,[37] for genetic correlation analyses, and HASE,[38] for genome-wide association studies. These improvements in software, and also hardware, now pave the way for full-scale analyses without reliance on data reduction methods.

In conclusion, our work demonstrates that the shape of subcortical brain structures is a relevant phenotype for genetic studies, complementary to aggregated measures. Fine-scale maps of genetic influences on the brain are likely to reveal a complex mosaic of genetic modules, with partially divergent sets of genes that drive them.

REFERENCES

1. Bressler, S. L. & Menon, V. Large-scale brain networks in cognition: emerging methods and principles. *Trends in cognitive sciences* **14**, 277-290 (2010).

2. Doyon, J. & Benali, H. Reorganization and plasticity in the adult brain during learning of motor skills. *Current opinion in neurobiology* **15**, 161-167 (2005).

3. Andreasen, N. C. *et al.* Intelligence and brain structure in normal individuals. *American Journal of Psychiatry* **150**, 130-130 (1993).

4. Tekin, S. & Cummings, J. L. Frontal–subcortical neuronal circuits and clinical neuropsychiatry: an update. *Journal of psychosomatic research* **53**, 647-654 (2002).

5. Verstraete, E., Veldink, J. H., den Berg, L. H. & den Heuvel, M. P. Structural brain network imaging shows expanding disconnection of the motor system in amyotrophic lateral sclerosis. *Human brain mapping* **35**, 1351-1361 (2014).

6. Blokland, G. A. M., de Zubicaray, G. I., McMahon, K. L. & Wright, M. J. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Research and Human Genetics* **15**, 351-371 (2012).

7. Peper, J. S. *et al.* Genetic influences on human brain structure: a review of brain imaging studies in twins. *Human brain mapping* **28**, 464-473 (2007).

8. den Braber, A. *et al.* Heritability of subcortical brain measures: a perspective for future genome-wide association studies. *Neuroimage* **83**, 98-102 (2013).

9. Gutman, B. A. *et al.* in *Information Processing in Medical Imaging*   205-218 (Springer International Publishing, 2015).

10. Gutman, B. A., Madsen, S. K., Toga, A. W. & Thompson, P. M. in *Multimodal Brain Image Analysis* Vol. 8159 *Lecture Notes in Computer Science* (eds Li Shen *et al.*) Ch. 24, 246-257 (Springer International Publishing, 2013).

11. Reuter, M., Wolter, F.-E. & Peinecke, N. Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids. *Comput. Aided Des.* **38**, 342-366, doi:10.1016/j.cad.2005.10.011 (2006).

12. Wang, Y. *et al.* Surface-based TBM boosts power to detect disease effects on the brain: an N=804 ADNI study. *Neuroimage* **56**, 1993-2010, doi:10.1016/j.neuroimage.2011.03.040 (2011).

13. Yonggang, S. *et al.* Metric Optimization for Surface Analysis in the Laplace-Beltrami Embedding Space. *Medical Imaging, IEEE Transactions on* **33**, 1447-1463, doi:10.1109/tmi.2014.2313812 (2014).

14. Cole, J. H. *et al.* Subregional hippocampal morphology and psychiatric outcome in adolescents who were born very preterm and at term. *PloS one* **10**, e0130094 (2015).

15. Wade, B. S. C. *et al.* in *SPIE Medical Imaging.* 94171S-94171S-94178 (International Society for Optics and Photonics).

16. McKeown, M. J. *et al.* Shape (but not volume) changes in the thalami in Parkinson disease. *BMC Neurol* **8** (2008).

17. Bron, E. E. *et al.* Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge. *Neuroimage* **111**, 562-579 (2015).

18. Bis, J. C. *et al.* Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* **44**,

545-551 (2012).

19. Stein, J. L. *et al.* Identification of common variants associated with human hippocampal and intracranial volumes. *Nature genetics* **44**, 552-561 (2012).

20. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224-229, doi:10.1038/nature14101 (2015).

21. Hofman, A. *et al.* The Rotterdam Study: 2014 objectives and design update. *Eur J Epidemiol* **28**, 889-926, doi:10.1007/s10654-013-9866-z (2013).

22. Ikram, M. A. *et al.* The Rotterdam Scan Study: design and update up to 2012. *Eur J Epidemiol* **26**, 811-824, doi:10.1007/s10654-011-9624-z (2011).

23. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics* **44**, 955-959 (2012).

24. Fischl, B. *et al.* Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23 Suppl 1**, S69-84 (2004).

25. Gutman, B. A. *et al.* Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features. *Neuroimage* **70**, 386-401, doi:http://dx.doi.org/10.1016/j.neuro-image.2012.12.052 (2013).

26. Gutman, B. A. *et al.* in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on.* 1402-1406 (IEEE).

27. Gutman, B. A., Wang, Y., Rajagopalan, P., Toga, A. W. & Thompson, P. M. in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on.* 716-719 (IEEE).

28. Adams, H. H. *et al.* Heritability and Genome-Wide Association Analyses of Human Gait Suggest Contribution of Common Variants. *J Gerontol A Biol Sci Med Sci*, doi:10.1093/gerona/glv081 (2015).

29. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics* **62**, 1198-1211 (1998).

30. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300 (1995).

31. Chen, C.-H. *et al.* Hierarchical genetic organization of human cortical surface area. *Science* **335**, 1634-1636 (2012).

32. Chen, C.-H. *et al.* Genetic influences on cortical regionalization in the human brain. *Neuron* **72**, 537-544 (2011).

33. Chen, C.-H. *et al.* Genetic topography of brain morphology. *Proceedings of the National Academy of Sciences* **110**, 17089-17094 (2013).

34. Whelan, C. D. *et al.* Heritability and reliability of automatically segmented human hippocampal formation subre-gions. *Neuroimage* **128**, 125-137 (2016).

35. Eyler, L. T. *et al.* A comparison of heritability maps of cortical surface area and thickness and the influence of ad-justment for whole brain measures: a magnetic resonance imaging twin study. *Twin Research and Human Genetics* **15**, 304-314 (2012).

36. Ge, T. *et al.* Massively expedited genome-wide heritability analysis (MEGHA). *Proceedings of the National Acade-my of Sciences* **112**, 2479-2484 (2015).

37. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature*

genetics **47**, 284-290 (2015).

38. Roshchupkin, G. *et al.* HASE: Framework for efficient high-dimensional association analyses. *bioRxiv*, 037382 (2016)

39. Adams H.H.H. et al. "Novel genetic loci underlying human intracranial volume identified through genome-wide association", *Nature Neuroscience* (2016), in press.
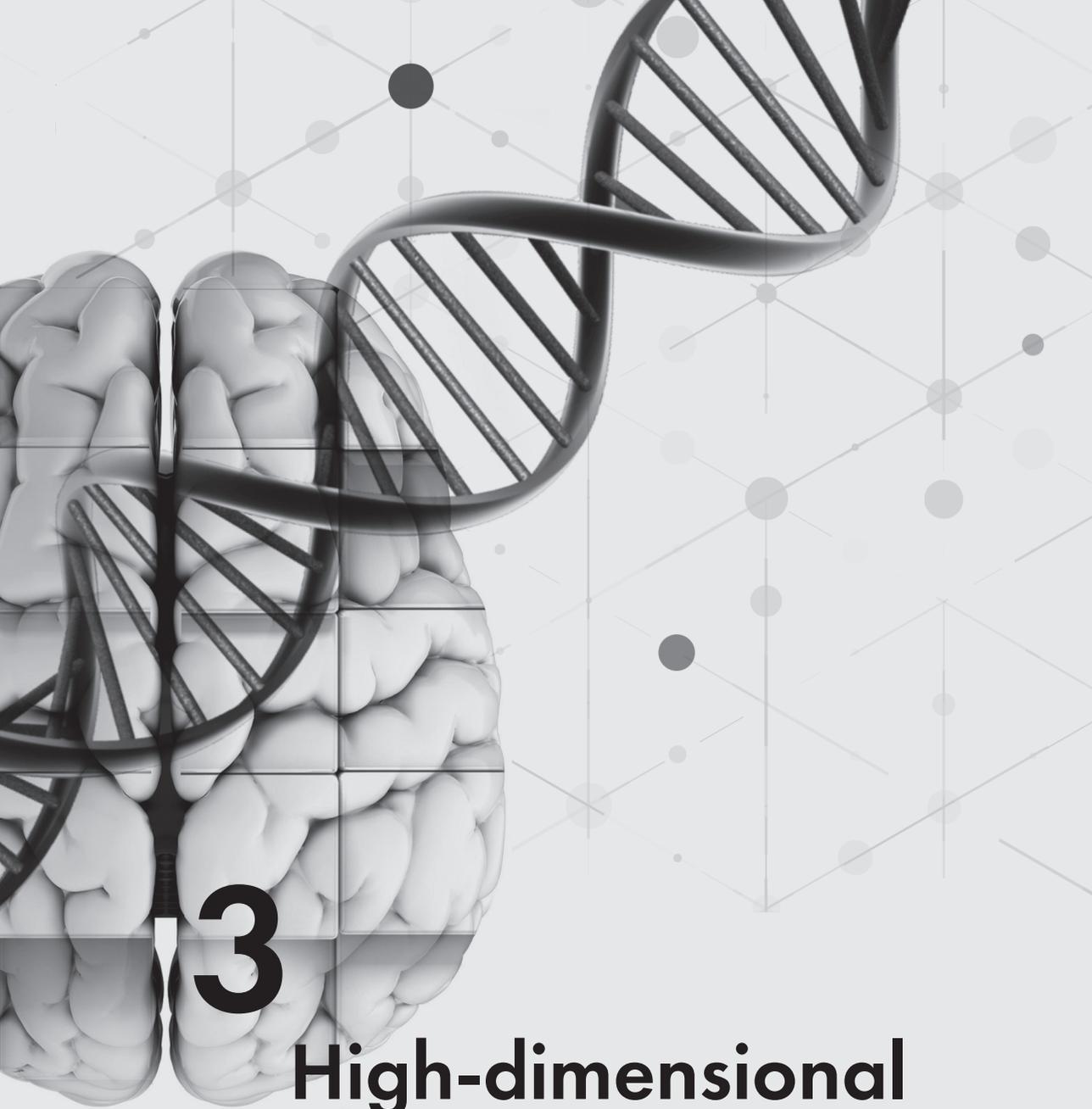
3

High-dimensional phenotypes as a fine-mapping instrument

CHAPTER 3.1

Fine-mapping the effects
of Alzheimer's disease risk
loci on brain morphology.

ABSTRACT

Background

The neural substrate of genetic risk variants for Alzheimer's disease (AD) remains unknown. We studied their effect on healthy brain morphology to provide insight into disease etiology in the pre-clinical phase.

Methods

We included 4071 non-demented, elderly participants of the population-based Rotterdam Study who underwent brain MRI and genotyping. We performed voxel-based morphometry (VBM) on all gray matter voxels for 19 previously identified, common AD risk variants. Whole-brain expression data from the Allen Human Brain Atlas was used to examine spatial overlap between VBM association results and expression of genes in AD risk loci regions.

Results

Brain regions most significantly associated with AD risk variants were the left postcentral gyrus with *ABCA7* (rs4147929, p = 4.45×10^{-6}), right superior frontal gyrus by *ZCWPW1* (rs1476679, p = 5.12×10^{-6}), and right postcentral gyrus by *APOE* (p = 6.91×10^{-6}). Though no individual voxel passed multiple testing correction, we found significant spatial overlap between the effects of AD risk loci on VBM and the expression of genes (*MEF2C, CLU, SLC24A4)* in the Allen Brain Atlas. Results are available online on www.imagene.nl/ADSNPs/. *Conclusion:* In this single largest imaging genetics dataset worldwide, we found that AD risk loci affect cortical gray matter in several brain regions known to be involved in AD, as well as regions that have not been implicated before.

Keywords

Alzheimer's disease; Dementia; genetics; magnetic resonance imaging; brain; voxel-based morphometry.

INTRODUCTION

Alzheimer's disease (AD) is a complex neurodegenerative disease and the most common cause of dementia. It has a long preclinical phase, during which there are no symptoms but structural brain changes can already be detected, such cortical atrophy and localized atrophy of the hippocampus (Thompson et al., 2001; Weiner et al., 2012).

In recent years, common genetic risk factors for AD have been discovered through large meta-analyses of genome-wide association studies (GWAS) (Lambert et al., 2013). However, the underlying neurobiological substrate leading to AD for the genes assigned to these risk loci remains to be uncovered. Identifying the brain structures affected by these genes can increase our understanding of AD and aid future functional studies. Previous studies have investigated some of the AD risk loci in relation to neuroimaging measures (Bis et al., 2012; Chauhan et al., 2015; Liu et al., 2014; Morgen et al., 2014). However, they were generally focused on candidate regions that are known to play a role in AD, such as the hippocampus (Bis et al., 2012; Chauhan et al., 2015) or did not investigate all known risk loci (Liu et al., 2014; Morgen et al., 2014). Unbiased approaches for analyzing brain images have great potential to give novel insights that would not have been considered a priori. Voxel-based morphometry (VBM) is a hypothesis-free technique for analyzing brain imaging data that characterizes regional tissue concentration differences across the whole brain, without the need to predefine regions of interest (Wright et al., 1995). Using VBM, we studied the association of 19 AD genetic risk loci with gray matter morphology at the voxel level in 4071 non-demented elderly from the Rotterdam study. This study provides insight into non-diseased brain morphology. Such knowledge is complementary and intertwined with better understanding disease etiology in the pre-clinical phase. Subsequently, we co-localized our results with publicly available genetic expression data. We thus identified genetic associations with known as well as novel regions affected in AD.

METHODS

STUDY POPULATION

The Rotterdam Study is an ongoing population-based cohort study in the Netherlands investigating diseases in the elderly and currently consists of 14,926 residents of Rotterdam who were aged 45 years or more at baseline (Hofman et al., 2011; Ikram and Lugt, 2011). The initial cohort was started in 1990 and expanded in 2000 and 2005. The whole population is subject to a set of multidisciplinary examinations every four years. MRI was implemented in 2005 and 5430 persons scanned until 2011 were eligible for this study. We excluded individuals with incomplete acquisitions, scans with artifacts hampering automated processing, participants with MRI-defined cortical infarcts, and subjects with dementia or stroke at the time of scanning. This resulted in a final study population of 4071 non-demented persons with information available on both genome-wide genotyping and MRI data. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands , implementing the Wet Bevolkingsonderzoek: ERGO (Population Studies Act: Rotterdam Study). All participants provided written informed consent to participate in the study and to obtain information from their treating physicians.

IMPUTATION OF GENOTYPES

The Illumina 550K and 550K duo arrays were used for genotyping. Samples with low call rate (<97.5%), with excess autosomal heterozygosity (>0.336) or with sex-mismatch were excluded, as were outliers identified by the identity-by-state clustering analysis (outliers were defined as being >3 standard deviation (SD) from population mean or having identity-by-state probabilities >97%). A set of genotyped input SNPs with call rate >98%, MAF >0.001 and Hardy–Weinberg equilibrium (HWE) P-value $> 10^{-6}$ was used for imputation. The Markov Chain Haplotyping (MACH) package version 1.0 software (Imputed to plus strand of NCBI build 37, 1000 Genomes phase I version 3) and minimac version 2012.8.6 were used for imputation. *APOE* status was genotyped separately, using a polymerase chain reaction, as described in (Verhaaren et al., 2013). *APOE*ɛ4 was coded as the number of ApoEɛ4 alleles.

MRI DATA

From August 2005 onwards, a dedicated 1.5 Tesla MRI scanner (GE Healthcare, Milwaukee, Wisconsin, USA) is operational in the Rotterdam Study research center in Ommoord. This scanner is operated by trained research technicians and all imaging data are collected according to standardized image acquisition protocols (Ikram and Lugt, 2011). Brain MRI scans included a high-resolution 3D T1-weighted fast RF spoiled gradient recalled acquisition in steady state with an inversion recovery pre-pulse (FASTSPGR-IR) sequence with thin slices (voxel size<1mm^3) (Ikram and Lugt, 2011).

IMAGE PROCESSING

Voxel based morphometry (VBM) was performed according to an optimized VBM protocol (Good et al., 2001). First, all T1-weighted images were segmented into supratentorial gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) using a previously described k-nearest neighbor (kNN) algorithm, which was trained on six manually labeled atlases (Vrooman et al., 2007)which has been successfully applied to classify brain tissue in MR data, requires training on manually labeled subjects. This manual labeling is a laborious and time-consuming procedure. In this work, a new fully automated brain tissue classification procedure is presented, in which kNN training is automated. This is achieved by non-rigidly registering the MR data with a tissue probability atlas to automatically select training samples, followed by a post-processing step to keep the most reliable samples. The accuracy of the new method was compared to rigid registration-based training and to conventional kNN-based segmentation using training on manually labeled subjects for segmenting gray matter (GM. FSL software (Smith et al., 2004) was used for VBM data processing. Then, all GM density maps were non-linearly registered to the standard GM probability template. For this study we chose the ICBM MNI152 GM template (Montreal Neurological Institute) with a 1x1x1 mm^3 voxel resolution. The MNI152 standard-space T1-weighted average structural template is derived from 152 structural images, which have been warped and averaged into the common MNI152 co-ordinate system after high-dimensional nonlinear registration.

A spatial modulation procedure was used to avoid differences in absolute GM volume due to the registration. This involved multiplying voxel density values by the Jacobian determinants estimated during spatial normalization. All images were smoothed using a 3mm (FWHM 8mm) isotropic Gaussian kernel.

STATISTICAL ANALYSIS

Linear regression models were fitted with voxel values of GM modulation density as the dependent variable and age, sex, and the number of reference alleles (risk alleles for Alzheimer's disease, **Supplementary Table 5**) as independent variables. In total 1,534,602 voxels were processed.

To perform a nonparametric permutation test, we randomly shuffled the genotype data between persons and performed the VBM association analysis with all 1,534,602 voxels in gray matter. This was repeated 10,000 times and for every permutation we saved the minimum p-value. Subsequently, we took the 5th percentile of this minimum p-value distribution to compute FWE p-value threshold, which was 3.0×10^{-7} (Churchill and Doerge, 1994). This was then divided by 19 to account for the number of independent SNPs, resulting in the final

threshold of 1.66 x 10^{-8}.

GENETIC RISK SCORE

Genetic risk scores (GRS) were constructed by multiplying the number of risk alleles by their reported odds ratio (after natural logarithm transformation) for the disease, and summing this weighted allele score of each variant up into a disease risk score for AD(Adams et al., 2015). We tested a GRS based on all 19 AD SNPs and second GRS excluding APOEε4.

APOEε4 STRATIFIED ANALYSIS

To investigate whether it is possible to enrich association signal of AD variants on brain morphology we split our sample into groups with increased chance for AD pathology by stratifying it for APOEε4 status. In total there were 1168 carrier and 2903 non-carrier in our data set.

THE ALLEN HUMAN BRAIN GENE EXPRESSION ANALYSIS.

The Allen Human Brain Atlas (http://human.brain-map.org) includes RNA microarray data collected from the postmortem brains of six donors, with no known neuropsychiatric or neuropathological history. Around 500 samples per subject, per hemisphere were tested for expression profiles of 29,191 genes represented by 58,692 probes. The expression profiles were normalized across samples and across different brains as described previously ("ALLEN Human Brain Atlas Normalization, Microarray Data," 2013).

In our analysis we used the three Caucasian donors. For each of these donors we extracted expression profiles of 216 genes, which are located within ± 500kb from AD risk loci and used the MNI coordinates to map the location of the samples. For each probe we derived z-score statistics, which represent deviation of gene expression in that sample relative to background expression. Next, using the VBM association results from all 19 tested AD SNPs, we formed clusters at the significance threshold of observed p-value<0.05 and identified all tissue samples localized inside these clusters or within 10 voxels from them.

We have performed 10.000 random VBM analyses to generate p-value maps of null associations. We formed clusters, based on a p-value threshold of < 0.05, and subsequently linked these to probes, exactly as described above. For three donors and all probes in the 216 genes (in total 667) we calculated the t-test statistic with a null hypothesis that expression of the gene within clusters is not significantly different from background expression. We saved the minimum p-values for every random VBM map. Subsequently, we took the 5th percentile of this minimum p-value distribution to compute the FWE p-value threshold. The obtained threshold was 1.7x10^{-5} . Then we performed the same t-test with the AD VBM maps. Thus, we compared expression of genes around AD risk loci in regions identified in the VBM analysis

with their background expression in the brain.

REGIONAL ANALYSIS

We used the Hammer atlas (Hammers et al., 2003)automated anatomical labeling of individual brain imaging datasets, and the statistical assessment of normal ranges for structure volumes and extents. No such manually constructed atlas is currently available for the frequently studied group of young adults. We studied 20 normal subjects (10 women, median age 31 years to segment the gray matter into 36 regions for both hemispheres and compare effects on specific brain regions. We summed all voxels values inside segmented region to estimate gray matter volume. For every risk locus and brain region we run the same regression model as for the VBM analysis.

VISUALIZATION

To provide easy access to study results for the research community, we developed an online freely available interactive visualization tool (www.imagene.nl/ADSNPs/).

RESULTS

ASSOCIATION OF AD RISK LOCI TO VOXEL-BASED MORPHOMETRY

The study population for VBM analysis consisted of 4071 non-demented persons with information available on both genome-wide genotyping and MRI data from the population-based Rotterdam Study. The mean age was 64.7 ($\pm$ 10.7) years and 2251 (55%) subjects were women.

We studied the association of 19 AD risk loci with 1,534,602 voxels of gray matter. None of the associations reached the multiple-testing correction threshold 1.66×10^{-8}. **Table 1** shows all associations between AD risk loci and gray matter voxel density with suggestive evidence for association p-values $< 1 \times 10^{-5}$. The strongest associations of gray matter voxel with AD risk loci were found in the left postcentral gyrus, right superior frontal gyrus, and right postcentral gyrus. In **Figure 1** we show the three-dimensional maps of the nominally significant (p-value<0.05) associations for the APOE risk loci. The negative clusters of APOE are located close to the medial temporal lobe, in particular around the hippocampus, whereas positive clusters are mainly in the occipital lobe. The GRSs association also did not reach the correction threshold. The strongest signal for risk score with APOE was found in the postcentral gyrus right (p-value$=8.02 \times 10^{6}$) and for the risk score without APOE in the lateral remainder of the occipital lobe right (pvalue$=1.47 \times 10^{-5}$). On **Supplementary Figure 1** are shown maps for all risk loci from **Table 1**. **Supplementary Table 2** provides the full list of the top three associated clusters of voxels for each risk locus and more detailed statistical information. All study results are available and can interactively be explored on the ImaGene website: www.imagene.nl/ADSNPs/.

TABLE 1: The most significant voxel-wise association signals with p-values$<10^{-5}$. Brain region labeling based on the Hammer Atlas segmentation. Effect direction indicates beta sign, and demonstrates risk loci associated with increasing gray matter tissue ($+$) or decreasing gray matter tissue (-).

Risk variant	Gene*	Minimum p-value	Effect direction	Brain Region
rs4147929	ABCA7	4.46×10^{-6}	-	postcentral gyrus left
rs1476679	ZCWPW1	5.12×10^{-6}	+	superior frontal gyrus right
rs429358/rs7412	APOEε4	6.91×10^{-6}	+	postcentral gyrus right
rs11771145	EPHA1	8.91×10^{-6}	-	precentral gyrus right
rs190982	MEF2C	9.55×10^{-6}	+	lateral remainder of occipital lobe right
Genetic Risk Score	All	8.02×10^{-6}	+	postcentral gyrus right
Genetic Risk Score	Without APOE	$1.47 \times 10^{-5**}$	+	lateral remainder of occipital lobe right

* Assigned risk gene according to Lambert et al [1]

** P-value is not less than 10^{-5}, shown to compare with GRS without exclusion APOE.

In APOEɛ4 stratified analysis none of the signals passed the threshold, however variant in MEF2C loci showed much more significant association compare to full sample size analysis (**Supplementary Table 5**). Additionally, the association signal for non-carrier was in general less significant (**Supplementary Table 6**).
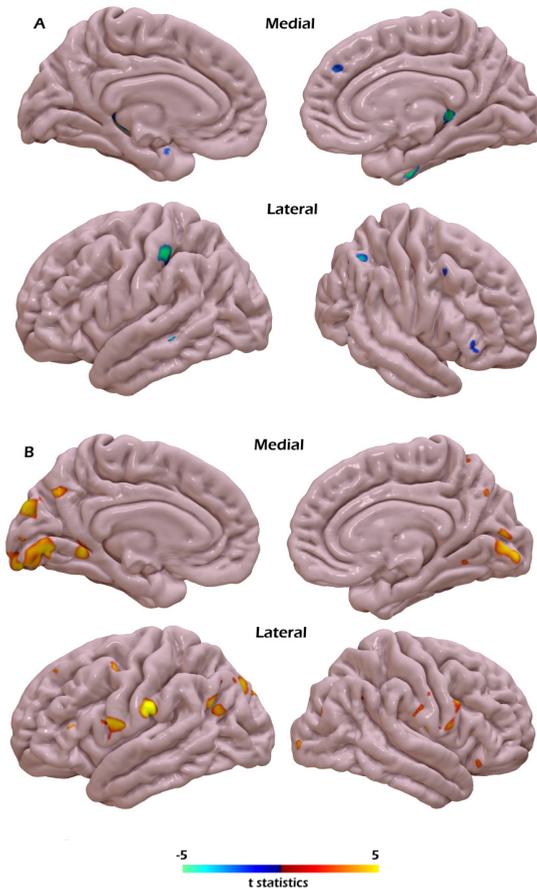
FIGURE 1: Projection of APOE risk loci association clusters from VBM to cortical surface. Colors reflect regression association: blue for negative (A), red for positive (B). Clusters formed based on nominal significant p-value threshold 0.05.

SPATIAL OVERLAP BETWEEN ASSOCIATION MAPS AND GENE EXPRESSION

To investigate whether the effect of AD risk loci on VBM overlaps with gene expression in the brain, we used the Allen Human Brain Atlas data. We overlapped brain regions identified through our VBM analysis with the maps of samples from three Allen Human Brain Atlas

donors (**Figure 3**). We compared expression within the identified voxel clusters with background expression. In total we tested the expression profiles of 216 protein-coding genes, located ± 500kb from the AD variants (**Supplementary Table 1**). We found that *MEF2C, CLU, SLC24A4* were significantly expressed (p-value<1.7x10^{-5}) in the identified voxel clusters compared to other genes at that particular locus. Interestingly, these were the genes that were previously assigned as the risk genes at each respective locus based on a review of the available literature (Lambert et al., 2013) (**Table 2**). Additionally, we found genes showing significantly different expression, which are located in the risk loci but were previously not proposed as the causal gene for AD. These are: *NGEF* (p-value=7.57x10^{-16}) for the region around rs35349669 and *GSTK1* (p-value=1.01x10^{-5}) for the region around rs11771145. **Supplementary Table 2** provides the full list of genes and more detailed statistical information.

TABLE 2: Results of spatial overlap between VBM risk loci association and gene expression profiles of 3 Caucasian donors from the Allen Human Brain Atlas. The table shows genes in risk loci regions, for which expression differs significantly (at corrected threshold 1.7x10^{-5}) from background expression in regions associated by VBM analysis.

Risk variant	Putative causal gene[a]	Genes showing significant overlap				
	Locus	Significant gene expression	Minimum p-value	Distance from risk loci, bp	Significant donors/ total number of donors	Significant probes/total number of probes
rs10498633	SLC24A4	SLC24A4	1.50x10^{-5}	138.027	1/3	1/2
rs190982	MEF2C	MEF2C	1.41x10^{-5}	44.275	1/3	1/3
rs9331896	CLU	CLU	4.43x10^{-7}	13.252	1/3	1/1
rs35349669	INPP5D	NGEF	7.57x10^{-16}	325.080	3/3	2/2
rs11771145	EPHA1	GSTK1	1.01x10^{-5}	169.576	2/3	3/4

[a] Assigned causal gene according to Lambert et al [1];

REGIONAL ANALYSIS

Figure 2 provides a heat map showing all AD risk loci and their effect on different brain regions sorted by lobe. None of the association signals passed Bonferroni correction, however several loci showed nominal significant association (p-value <0.05; cells with stars on **Figure 2**), among them variant in EPHA1 with less tissue in caudate and in insula, CELF1 with more tissue in accumbens and APOE with very strong positive effect in the occipital lobe. Variants in *APOE*, *FERMT2*, *PTK2B, CASS4* and *MS4A6A* showed the strongest effect on hippocampus and were associated with smaller gray matter volume. Risk variants in *EPHA1* and *SORL1* had the largest negative effect on deep gray matter structures: putamen, thalamus, and pallidum.
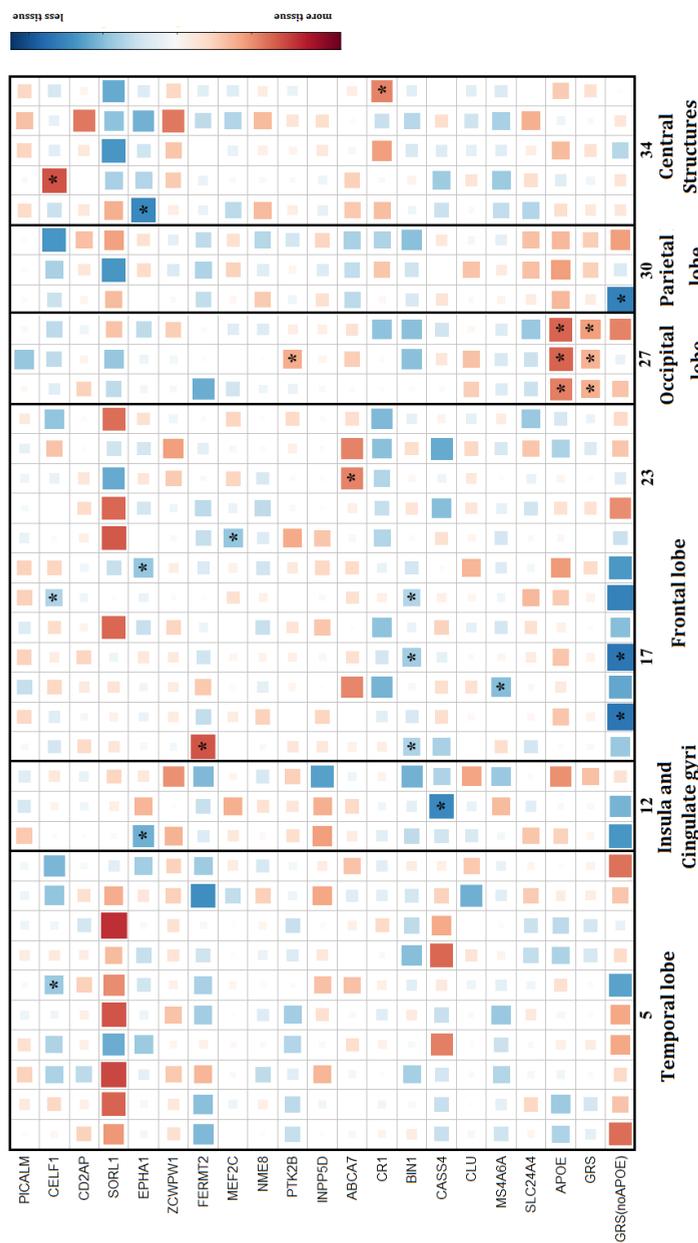
FIGURE 2: Heatmap of AD risk loci[1] association effects from ROI analysis. The 19 AD risk loci are on the y-axis and brain regions, grouped by lobe, on the x-axis. Brain region labeling was based on the Hammer Atlas. Blue indicates that risk loci were associated with less gray matter tissue; red indicates association with more gray matter tissue. Supplementary Table 3 provides a coded structure list. Regions with nominal significant association (p-value < 0.05) marked with *.

[1] Assigned risk gene according to Lambert et al [1]
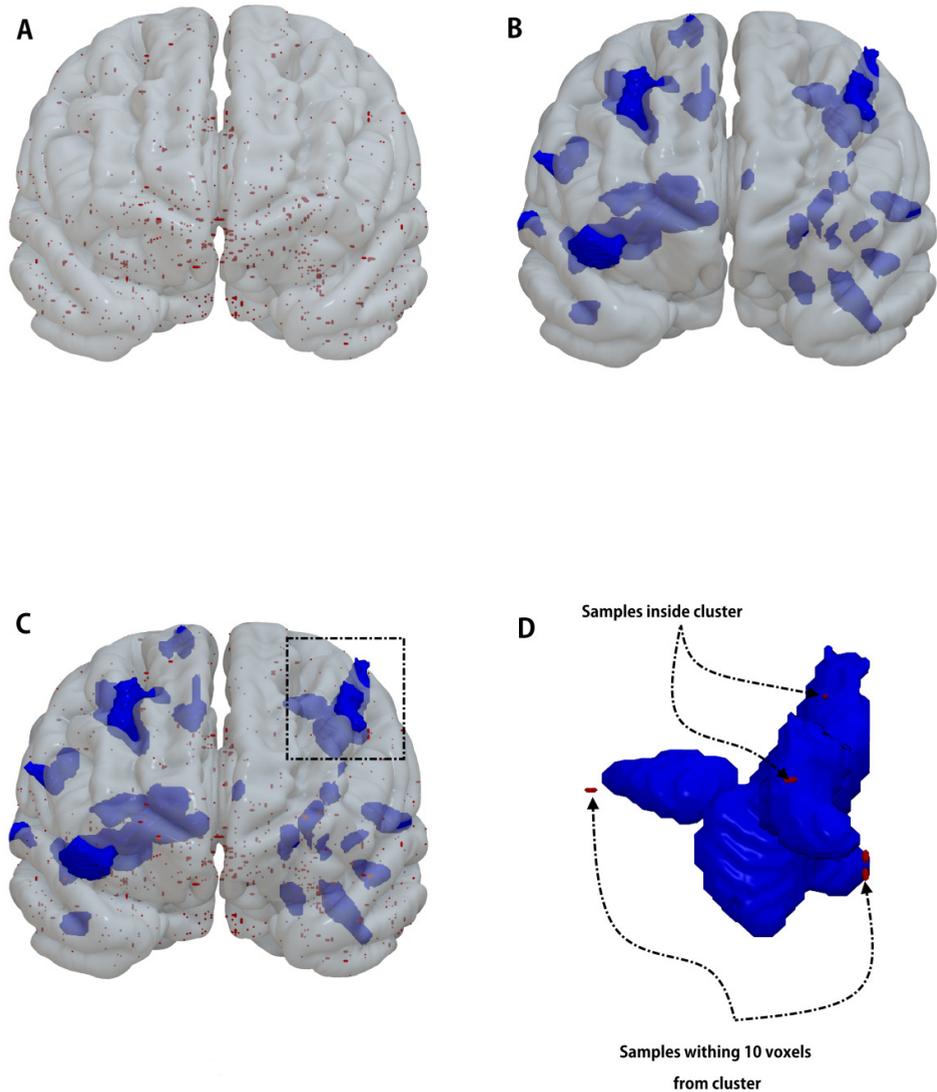
FIGURE 3: Example of spatial overlap between VBM association map for the MEF2C risk variant and MEF2C gene expression probes from Allen Human Brain Atlas. **(A)** –samples (red color) distribution from "donor9861" of Allen Human Brain Atlas; **(B)** – clusters of associated with MEF2C risk loci voxels (blue color) identified through VBM analysis formed using p-value threshold 0.05; **(C)** – Spatial overlap between Allen Brain probes and VBM clusters; **(D)** – example of VBM cluster and assigning sample location to them.

DISCUSSION

This study presents the association of 19 genome-wide significant AD risk loci (Lambert et al., 2013) with VBM of the gray matter, among 4071 middle aged and elderly subjects from the population-based Rotterdam Study. The unprecedented sample size has enabled this unbiased whole grey matter investigation of established risk variants and their effect on brain morphology. We found nominally significant associations with the left postcentral gyrus, the right superior frontal gyrus and the right postcentral gyrus. Furthermore, through comparing our VBM results to the Allen Brain atlases of human gene expression, we found significant spatial overlap for genes previously assigned to be the causal gene in these loci (*CLU, SLC24A4* and *MEF2C*). Additionally, we identified two genes, not previously suggested to be the causal gene in AD (*GSTK1* and *NGEF*), of which the expression in the brain significantly overlaps with our VBM results.

There currently exists no consensus for voxel-wise genetics studies regarding the significance threshold for avoiding false positive findings while not to being too conservative (Fritsch et al., 2015; Medland et al., 2014). A number of data processing and statistical analysis methods have been proposed in the literature to address this issue for neuroimaging analysis(Bullmore et al., 1999; Jenkinson et al., 2002; Smith and Nichols, n.d.). However, all these methods rely on a set of assumptions about the statistical structure of the data. Therefore, in our study we decided to use unbiased, but more conservative, non-parametric permutation methods to define the statistical threshold of significance. Although this is the largest genetic VBM study conducted to date, none of the voxels passed this conservative multiple testing correction. However, we have previously shown that AD risk loci are associated with cognitive functioning in the general population (Adams et al., 2015; Davies et al., 2015; de Bruijn et al., 2015; Verhaaren et al., 2013) as well as hippocampal volume in a larger sample (N= 9,232)(Bis et al., 2012). This showed that subclinical effects of AD risk loci exist and that effects on gray matter could be expected. Additionally, we constructed genetic risk scores, to explore the combined effect of all AD SNPs on brain morphology. The association signal of GRSs also did not pass correction threshold and the strongest signal for GRS with APOE was driven by APOE variant, while for GRS without APOE by MEF2C variant (Supplementary Figure 2).

Furthermore, it is reasonable to assume that the effects of the risk loci are not restricted to a single voxel, but rather to a cluster of voxels spanning a certain brain region. Therefore, we further explored the nominally significant associations we found by using the Allen Brain Human Atlas to analyze gene expression, and using Hammer brain atlas to estimate average effect on specific brain regions.

In Hammer regional analysis, we found that risk loci for Alzheimer's disease affect brain morphology in established regions such as the hippocampus (e.g. loci near *APOE, FERMT2, PTK2B*), putamen, thalamus (*SORL1, EPHA1*), as well as regions not often reported on including the insula (*EPHA1*) and occipital lobe (*APOE*). The heat map in **Figure 2** summarizes the association results over the whole brain.

Alzheimer's disease is a complex disorder with multiple variants from different pathways involved in its etiology(Jones, 2015; Mattson, 2004). Therefore, as previously shown(Chauhan et al., 2015), the effect of these variants on brain morphology could also differ and have different directions. **Figure 2** provides a detailed map of such heterogeneous effects. For example, large brain structures, such as the temporal lobe and central regions, are affected differently. Also, some risk loci have a different direction of effects, e.g. *FERMT2* is associated with less tissue and *SORL1* with more tissue in the temporal lobe. Of particular interest is that we found the positive association of APOE with the occipical lobe, which could possibly be explained by cerebral amyloid angiopathy (CAA). Indeed, CAA is linked to APOEε4 carriership (Esiri et al., 2015; Ringman et al., 2014)Alzheimer's disease (AD and has a predilection for the occipital neocortex (Nelson, 2013). Moreover, CAA is involved in Alzheimer's disease (Smith and Greenberg, 2009) and is characterized by ß-amyloid deposition in the media and adventitia of small and medium sized arteries. In healthy subjects, this may be observed as an increase in gray matter tissue density because of the influx of cells to clear the deposits. More research on the effects of AD risk loci on brain morphology is needed to further unravel the biological substrates involved in disease etiology.

Previous case-control studies showed ambiguous differential expression of putative causal genes for AD in the brain (Holton et al., 2013) or reported that the regional expression of each of the risk loci did not match the pattern of brain regional distribution in Alzheimer pathology (Karch et al., 2012). Most of AD variants are non-coding and for the follow up studies would be very important to explore the potential roles of these intronic and intergenic regions in the regulation of gene expression. Confirmed functional variants underlying validated GWAS hits are still sparse in the literature(Myers et al., 2007; Pandey and Manolio, 2010)when considering all the diseases and traits studied, but each of these is extremely valuable to the respective research and clinical environments. In our study, we found significant spatial overlap between VBM results in the Allen Human Brain atlas with some of the previously identified genes (*CLU, SLC24A4* and *MEF2C*). This could mean that genetic variability in these genes could act on gray matter density through differences in expression. This is also in line with the fact that most trait-specific GWAS signals are non-coding and probably act through modulation of gene expression(Ramasamy et al., 2014). Our results also suggest that VBM analysis combined with expression data could provide evidence for new candidate

genes in genetic loci, where the causal gene has not been strongly established by biological experiments(Steinberg et al., 2015). In AD loci, examples are *NGEF* for rs35349669 and *GSTK1* for rs11771145. Although the index variant rs35349669 is located within *INPP5D*, this gene is expressed at low levels in the brain (Lambert et al., 2013) and the linkage peak spans multiple genes with suggestive signals, including *NGEF* (Lambert et al., 2013). Neuronal Guanine Nucleotide Exchange Factor *(NGEF)*, among its related pathway is signaling by G protein–coupled receptors (GPCRs), which are involved at many stages of AD disease progression, and this class of receptors is a potential therapeutic target for AD (Thathiah and De Strooper, 2011).

Glutathione S-transferase Kappa 1 (GSTK1) is member of the superfamily of enzymes that function in cellular detoxification. Interestingly, a significant decrease of glutathione transferase activity in different brain regions in patients with Alzheimer disease was previously reported (Lovell et al., 1998), suggesting a possible link to Alzheimer through diabetes (Shield et al., 2010; Weyer et al., 2001).

Our study also has several limitations. The 19 AD risk loci do not include all genetics variants associated with AD and the index variants used may not be the causal variants. Another consideration is that the cross-sectional nature of our analyses precludes us from inferring causality from the associations. Although reverse causality is unlikely for genetic variants, it remains unclear whether our findings represent developmental or degenerative effects. The absence of significant association, as we mentioned before, could be due to strict permutation threshold or lack of power of our study sample size compare to GWAS analysis where these risk loci were discovered.

Additionally, in the experiment to determine spatial overlap between gene expression and regions identified through the VBM analysis, a number of considerations need to be taken into account. First, the threshold to form the clusters is a manual parameter and could be set to a different statistical threshold. However, with decreasing p-value threshold the number and size of the clusters goes down and, due to quite sparse distribution of Allen Brain samples, there are not enough clusters linked to samples to perform such analysis. Second, gene expression depends on the time of measurement and could be different over the lifespan and even during the day (Jaenisch and Bird, 2003). Second, the association between a risk locus and tissue density does not necessarily require the causative gene to be expressed in the same brain region, but could also be through a downstream effect of a functional pathway. Third, given the difficulties in obtaining brain tissue samples, these analyses are all based on relatively small samples.

CONCLUSION

Using a voxel-based morphometry study in over 4000 non-demented individuals, we provide a list of candidate brain regions that are potentially affected by AD risk loci and worthy of further study. Although detecting significant genetic effects on individual voxels will require even larger sample sizes, we show that data can be exploited by incorporating additional information in the analysis, such as gene expression data. All results of our study are available online on www.imagene.nl/ADSNPs/.

ACKNOWLEDGEMENTS

FINANCIAL DISCLOSURES

REFERENCES

1. Adams, H.H.H., de Bruijn, R.F. a. G., Hofman, A., Uiterlinden, A.G., van Duijn, C.M., Vernooij, M.W., Koudstaal, P.J., Ikram, M.A., 2015. Genetic risk of neurodegenerative diseases is associated with mild cognitive impairment and conversion to dementia. Alzheimer's Dement. 1–9. doi:10.1016/j.jalz.2014.12.008

2. ALLEN Human Brain Atlas Normalization, Microarray Data, 2013. 1–11.

3. Bis, J.C., DeCarli, C., Smith, A.V., van der Lijn, F., Crivello, F., Fornage, M., Debette, S., Shulman, J.M., Schmidt, H., Srikanth, V., Schuur, M., Yu, L., Choi, S.-H., Sigurdsson, S., Verhaaren, B.F.J., DeStefano, A.L., Lambert, J.-C., Jack, C.R., Struchalin, M., Stankovich, J., Ibrahim-Verbaas, C. a, Fleischman, D., Zijdenbos, A., den Heijer, T., Mazoyer, B., Coker, L.H., Enzinger, C., Danoy, P., Amin, N., Arfanakis, K., van Buchem, M. a, de Bruijn, R.F. a G., Beiser, A., Dufouil, C., Huang, J., Cavalieri, M., Thomson, R., Niessen, W.J., Chibnik, L.B., Gislason, G.K., Hofman, A., Pikula, A., Amouyel, P., Freeman, K.B., Phan, T.G., Oostra, B. a, Stein, J.L., Medland, S.E., Vasquez, A.A., Hibar, D.P., Wright, M.J., Franke, B., Martin, N.G., Thompson, P.M., Nalls, M. a, Uiterlinden, A.G., Au, R., Elbaz, A., Beare, R.J., van Swieten, J.C., Lopez, O.L., Harris, T.B., Chouraki, V., Breteler, M.M.B., De Jager, P.L., Becker, J.T., Vernooij, M.W., Knopman, D., Fazekas, F., Wolf, P. a, van der Lugt, A., Gudnason, V., Longstreth, W.T., Brown, M. a, Bennett, D. a, van Duijn, C.M., Mosley, T.H., Schmidt, R., Tzourio, C., Launer, L.J., Ikram, M.A., Seshadri, S., 2012. Common variants at 12q14 and 12q24 are associated with hippocampal volume. Nat. Genet. 44, 545–551. doi:10.1038/ng.2237

4. Bullmore, E.T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., Brammer, M.J., 1999. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. IEEE Trans. Med. Imaging 18, 32–42.

5. Chauhan, G., Adams, H.H.H., Bis, J.C., Weinstein, G., Yu, L., Töglhofer, A.M., Smith, A.V., J. van der Lee, S., Gottesman, R.F., Thomson, R., Wang, J., Yang, Q., Niessen, W.J., Lopez, O.L., Becker, J.T., Phan, T.G., Beare, R.J., Arfanakis, K., Fleischman, D., Vernooij, M.W., Mazoyer, B., Schmidt, H., Srikanth, V., Knopman, D.S., Jack, C.R., Amouyel, P., Hofman, A., DeCarli, C., Tzourio, C., van Duijn, C.M., Bennett, D. a., Schmidt, R., Longstreth, W.T., Mosley, T.H., Fornage, M., Launer, L.J., Seshadri, S., Ikram, M.A., Debette, S., 2015. Association of Alzheimer's disease GWAS loci with MRI markers of brain aging. Neurobiol. Aging 36, 1765.e7–1765.e16. doi:10.1016/j.neurobiolaging.2014.12.028

6. Churchill, G.A., Doerge, R.W., 1994. Empirical threshold values for quantitative trait mapping. Genetics 138, 963–71.

7. Davies, G., Armstrong, N., Bis, J.C., Bressler, J., Chouraki, V., Giddaluru, S., Hofer, E., Ibrahim-Verbaas, C. a, Kirin, M., Lahti, J., van der Lee, S.J., Le Hellard, S., Liu, T., Marioni, R.E., Oldmeadow, C., Postmus, I., Smith, a V, Smith, J. a, Thalamuthu, A., Thomson, R., Vitart, V., Wang, J., Yu, L., Zgaga, L., Zhao, W., Boxall, R., Harris, S.E., Hill, W.D., Liewald, D.C., Luciano, M., Adams, H., Ames, D., Amin, N., Amouyel, P., Assareh, a a, Au, R., Becker, J.T., Beiser, A., Berr, C., Bertram, L., Boerwinkle, E., Buckley, B.M., Campbell, H., Corley, J., De Jager, P.L., Dufouil, C., Eriksson, J.G., Espeseth, T., Faul, J.D., Ford, I., Scotland, G., Gottesman, R.F., Griswold, M.E., Gudnason, V., Harris, T.B., Heiss, G., Hofman, A., Holliday, E.G., Huffman, J., Kardia, S.L.R., Kochan, N., Knopman, D.S., Kwok, J.B., Lambert, J.-C., Lee, T., Li, G., Li, S.-C., Loitfelder, M., Lopez, O.L., Lundervold, a J, Lundqvist, A., Mather, K. a, Mirza, S.S., Nyberg, L., Oostra, B. a, Palotie, A., Papenberg, G., Pattie, A., Petrovic, K., Polasek, O., Psaty, B.M., Redmond, P., Reppermund,

S., Rotter, J.I., Schmidt, H., Schuur, M., Schofield, P.W., Scott, R.J., Steen, V.M., Stott, D.J., van Swieten, J.C., Taylor, K.D., Trollor, J., Trompet, S., Uitterlinden, a G., Weinstein, G., Widen, E., Windham, B.G., Jukema, J.W., Wright, a F., Wright, M.J., Yang, Q., Amieva, H., Attia, J.R., Bennett, D. a, Brodaty, H., de Craen, a J.M., Hayward, C., Ikram, M. a, Lindenberger, U., Nilsson, L.-G., Porteous, D.J., Räikkönen, K., Reinvang, I., Rudan, I., Sachdev, P.S., Schmidt, R., Schofield, P.R., Srikanth, V., Starr, J.M., Turner, S.T., Weir, D.R., Wilson, J.F., van Duijn, C., Launer, L., Fitzpatrick, a L., Seshadri, S., Mosley, T.H., Deary, I.J., 2015. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). Mol. Psychiatry 183–192. doi:10.1038/mp.2014.188

8.   de Bruijn, R.F.A.G., Bos, M.J., Portegies, M.L.P., Hofman, A., Franco, O.H., Koudstaal, P.J., Ikram, M.A., 2015. The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. BMC Med. 13, 132. doi:10.1186/s12916-015-0377-5

9.   Esiri, M., Chance, S., Joachim, C., Warden, D., Smallwood, A., Sloan, C., Christie, S., Wilcock, G., Smith, A.D., 2015. Cerebral amyloid angiopathy, subcortical white matter disease and dementia: literature review and study in OPTIMA. Brain Pathol. 25, 51–62. doi:10.1111/bpa.12221

10.   Fritsch, V., Da Mota, B., Loth, E., Varoquaux, G., Banaschewski, T., Barker, G.J., Bokde, A.L.W., Brühl, R., Butzek, B., Conrod, P., Flor, H., Garavan, H., Lemaitre, H., Mann, K., Nees, F., Paus, T., Schad, D.J., Schümann, G., Frouin, V., Poline, J.-B., Thirion, B., IMAGEN consortium, 2015. Robust regression for large-scale neuroimaging studies. Neuroimage 111, 431–41. doi:10.1016/j.neuroimage.2015.02.048

11.   Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage 14, 21–36. doi:10.1006/nimg.2001.0786

12.   Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum. Brain Mapp. 19, 224–47. doi:10.1002/hbm.10123

13.   Hofman, A., van Duijn, C.M., Franco, O.H., Ikram, M.A., Janssen, H.L. a, Klaver, C.C.W., Kuipers, E.J., Nijsten, T.E.C., Stricker, B.H.C., Tiemeier, H., Uitterlinden, A.G., Vernooij, M.W., Witteman, J.C.M., 2011. The Rotterdam Study: 2012 objectives and design update. Eur. J. Epidemiol. 26, 657–86. doi:10.1007/s10654-011-9610-5

14.   Holton, P., Ryten, M., Nalls, M., Trabzuni, D., Weale, M.E., Hernandez, D., Crehan, H., Gibbs, J.R., Mayeux, R., Haines, J.L., Farrer, L. a., Pericak-Vance, M. a., Schellenberg, G.D., Ramirez-Restrepo, M., Engel, A., Myers, A.J., Corneveaux, J.J., Huentelman, M.J., Dillman, A., Cookson, M.R., Reiman, E.M., Singleton, A., Hardy, J., Guerreiro, R., 2013. Initial Assessment of the Pathogenic Mechanisms of the Recently Identified Alzheimer Risk Loci. Ann. Hum. Genet. 77, 85–105. doi:10.1111/ahg.12000

15.   Ikram, M., Lugt, A. van der, 2011. The Rotterdam Scan Study: design and update up to 2012. Eur. J. Epidemiol. 26, 811–24. doi:10.1007/s10654-011-9624-z

16.   Jaenisch, R., Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. 33 Suppl, 245–254. doi:10.1038/ng1089

17.   Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. Neuroimage 17, 825–841. doi:10.1006/nimg.2002.1132

18.   Jones, L., 2015. Convergent genetic and expression data implicate immunity in Alzheimer's disease. Alzheimer's

Dement. 11, 658–671. doi:10.1016/j.jalz.2014.05.1757

19.   Karch, C.M., Jeng, A.T., Nowotny, P., Cady, J., Cruchaga, C., Goate, A.M., 2012. Expression of Novel Alzheimer's Disease Risk Genes in Control and Alzheimer's Disease Brains. PLoS One 7. doi:10.1371/journal.pone.0050976

20.   Lambert, J., Ibrahim-Verbaas, C., Harold, D., 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. 45, 1452–1458. doi:10.1038/ng.2802.Meta-analysis

21.   Liu, Y., Yu, J.-T., Wang, H.-F., Hao, X.-K., Yang, Y.-F., Jiang, T., Zhu, X.-C., Cao, L., Zhang, D.-Q., Tan, L., 2014. Association between NME8 Locus Polymorphism and Cognitive Decline, Cerebrospinal Fluid and Neuroimaging Biomarkers in Alzheimer's Disease. PLoS One 9, e114777. doi:10.1371/journal.pone.0114777

22.   Lovell, M.A., Xie, C., Markesbery, W.R., 1998. Decreased glutathione transferase activity in brain and ventricular fluid in Alzheimer's disease. Neurology 51, 1562–1566. doi:10.1212/WNL.51.6.1562

23.   Mattson, M.P., 2004. Pathways towards and away from Alzheimer's disease. Nature 430, 631–639. doi:10.1038/nature02940

24.   Medland, S.E., Jahanshad, N., Neale, B.M., Thompson, P.M., 2014. Whole-genome analyses of whole-brain data : working within an expanded search space. Nat. Publ. Gr. 17, 791–800. doi:10.1038/nn.3718

25.   Morgen, K., Ramirez, A., Frölich, L., Tost, H., Plichta, M.M., Kölsch, H., Rakebrandt, F., Rienhoff, O., Jessen, F., Peters, O., Jahn, H., Luckhaus, C., Hüll, M., Gertz, H.-J., Schröder, J., Hampel, H., Teipel, S.J., Pantel, J., Heuser, I., Wiltfang, J., Rüther, E., Kornhuber, J., Maier, W., Meyer-Lindenberg, A., 2014. Genetic interaction of PICALM and APOE is associated with brain atrophy and cognitive impairment in Alzheimer's disease. Alzheimers. Dement. 10, 1–8. doi:10.1016/j.jalz.2013.11.001

26.   Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P., Zismann, V.L., Joshipura, K., Huentelman, M.J., Hu-Lince, D., Coon, K.D., Craig, D.W., Pearson, J. V, Holmans, P., Heward, C.B., Reiman, E.M., Stephan, D., Hardy, J., 2007. A survey of genetic human cortical gene expression. Nat. Genet. 39, 1494–9. doi:10.1038/ng.2007.16

27.   Nelson, P.T., 2013. APOE- 2 and APOE- 4 Correlate with Increased Amyloid Accumulation in Cerebral Vasculature. J. Neuropathol. Exp. Neurol. 72, 708–715. doi:10.1016/j.biotechadv.2011.08.021.Secreted

28.   Pandey, J.P., Manolio, T. a, 2010. Genomewide association studies and assessment of the risk of disease. N. Engl. J. Med. 363, 166–176. doi:10.1056/NEJMra0905980

29.   Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., Hardy, J., Ryten, M., Trabzuni, D., Guelfi, S., Weale, M.E., Ramasamy, A., Forabosco, P., Smith, C., Walker, R., Arepalli, S., Cookson, M.R., Dillman, A., Gibbs, J.R., Hernandez, D.G., Nalls, M. a, Singleton, A.B., Traynor, B., van der Brug, M., Ferrucci, L., Johnson, R., Zielke, R., Longo, D.L., Troncoso, J., Zonderman, A., Coin, L., de Silva, R., Cookson, M.R., Singleton, A.B., Hardy, J., Ryten, M., Weale, M.E., 2014. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat. Neurosci. 17, 1418–1428. doi:10.1038/nn.3801

30.   Ringman, J.M., Sachs, M.C., Zhou, Y., Monsell, S.E., Saver, J.L., Vinters, H. V, 2014. Clinical Predictors of Severe Cerebral Amyloid Angiopathy and Influence of APOE Genotype in Persons With Pathologically Verified Alzheimer Disease. JAMA Neurol. 71, 878–83. doi:10.1001/jamaneurol.2014.681

31.   Shield, A.J., Murray, T.P., Cappello, J.Y., Coggan, M., Board, P.G., 2010. Polymorphisms in the human glutathione transferase Kappa (GSTK1) promoter alter gene expression. Genomics 95, 299–305. doi:10.1016/j.ygeno.2010.02.007

32. Smith, E.E., Greenberg, S.M., 2009. Amyloid, blood vessels, and brain function. Stroke 40, 2601–2606. doi:10.1161/STROKEAHA.108.536839

33. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1, S208–19. doi:10.1016/j.neuroimage.2004.07.051

34. Smith, S.M., Nichols, T.E., n.d. Threshold-Free Cluster Enhancement : Addressing problems of smoothing , threshold dependence and localisation in cluster inference FMRIB Technical Report TR08SS1 1, 1–20.

35. Steinberg, S., Stefansson, H., Jonsson, T., Johannsdottir, H., Ingason, A., Helgason, H., Sulem, P., Magnusson, O.T., Gudjonsson, S. a, Unnsteinsdottir, U., Kong, A., Helisalmi, S., Soininen, H., Lah, J.J., Almdahl, I.S., Andersen, F., Bogdanovic, N., Brækhus, A., Engedal, K., Rongve, A., Saltvedt, I., Stordal, E., Witoelar, A., Aarsland, D., Fladby, T., Ulstein, I.D., Djurovic, S., Sando, S.B., White, L.R., Knudsen, G.-P., Westlye, L.T., Selbæk, G., Giegling, I., Hampel, H., Hiltunen, M., Levey, A.I., Andreassen, O. a, Rujescu, D., Jonsson, P. V, Bjornsson, S., Snaedal, J., Stefansson, K., 2015. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat. Genet. 47, 445–447. doi:10.1038/ng.3246

36. Thathiah, A., De Strooper, B., 2011. The role of G protein-coupled receptors in the pathology of Alzheimer's disease. Nat. Rev. Neurosci. 12, 73–87. doi:10.1038/nrn2977

37. Thompson, P.M., Cannon, T.D., Narr, K.L., van Erp, T., Poutanen, V.P., Huttunen, M., Lönnqvist, J., Standertskjöld-Nordenstam, C.G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C.I., Toga, a W., 2001. Genetic influences on brain structure. Nat. Neurosci. 4, 1253–8. doi:10.1038/nn758

38. Verhaaren, B.F.J., Vernooij, M.W., Koudstaal, P.J., Uitterlinden, A.G., Van Duijn, C.M., Hofman, A., Breteler, M.M.B., Ikram, M.A., 2013. Alzheimer's disease genes and cognition in the nondemented general population. Biol. Psychiatry 73, 429–434. doi:10.1016/j.biopsych.2012.04.009

39. Vrooman, H.A., Cocosco, C.A., van der Lijn, F., Stokking, R., Ikram, M.A., Vernooij, M.W., Breteler, M.M.B., Niessen, W.J., 2007. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. Neuroimage 37, 71–81. doi:10.1016/j.neuroimage.2007.05.018

40. Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L. a, Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., Morris, J.C., Petersen, R.C., Saykin, A.J., Schmidt, M.E., Shaw, L., Siuciak, J. a, Soares, H., Toga, A.W., Trojanowski, J.Q., 2012. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimer's Dement. 8, S1–68. doi:10.1016/j.jalz.2011.09.172

41. Weyer, C., Funahashi, T., Tanaka, S., Hotta, K., Matsuzawa, Y., Pratley, R.E., Tataranni, P.A., 2001. Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia. J. Clin. Endocrinol. Metab. 86, 1930–1935. doi:10.1210/jcem.86.5.7463

42. Wright, I.C.C., McGuire, P.K., Poline, J.B., Travere, J.M., Murray, R.M., Frith, C.D., Frackowiak, R.S., Friston, K.J., 1995. A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. Neuroimage 2, 244–52. doi:10.1006/nimg.1995.1032

SUPPLEMENTARY MATERIAL

Fine-mapping the effects of Alzheimer's Disease risk loci on brain morphology.

CONTENT:

- **Supplementary Table 1:** Genetic variants with a genome wide significant effect on Alzheimer's disease.
- **Supplementary Table 2:** Top 3 associated clusters of voxels per each risk locus ordered by minimum p-value.
- **Supplementary Table 3:** Coded structure list for brain regions.
- **Supplementary Table 4:** Results of spatial overlap between VBM risk loci association and gene expression profiles of 3 Caucasian donors from the Allen Human Brain Atlas.
- **Supplementary Table 5:** The most significant voxel-wise association signals in APOEε4 stratified analysis for carrier.
- **Supplementary Table 6:** The most significant voxel-wise association signals in APOEε4 stratified analysis for non-carrier.
- **Supplementary Figure 1:** 3D brain VBM association maps.
- **Supplementary Figure 2:** Region brain Manhattan plots.

SUPPLEMENTARY TABLE 1: Genetic variants with a genome wide significant effect on Alzheimer's disease.

RS ID	Chr.	Gene *	Risk allele frequency	Risk allele	Other allele	OR
rs6656401	1	CR1	0.18	A	G	1.18
rs6733839	2	BIN1	0.42	T	C	1.22
rs35349669	2	INPP5D	0.49	T	C	1.08
rs190982	5	MEF2C	0.62	A	G	1.08
rs10948363	6	CD2AP	0.27	G	A	1.10
rs2718058	7	NME8	0.63	A	G	1.06
rs1476679	7	ZCWPW1	0.67	T	C	1.09
rs11771145	7	EPHA1	0.66	G	A	1.11
rs28834970	8	PTK2B	0.37	A	G	1.10
rs9331896	8	CLU	0.60	T	C	1.16
rs10838725	11	CELF1	0.31	G	C	1.08
rs983392	11	MS4A6A	0.59	A	G	1.90
rs10792832	11	PICALM	0.62	G	A	1.14
rs11218343	11	SORL1	0.95	C	T	1.37
rs17125944	14	FERMT2	0.10	C	T	1.14
rs10498633	14	SLC24A4	0.77	G	T	1.11
rs4147929	19	ABCA7	0.18	A	G	1.15
rs429358/rs7412	19	APOE	0.15	4	2/3	3.69
rs7274581	20	CASS4	0.91	C	T	1.15

* Assigned risk locus, risk allele and odds ratio (OR) according to Lambert et al [1]

SUPPLEMENTARY TABLE 2: Top 3 associated clusters of voxels per each risk locus ordered by minimum p-value. Cluster forming threshold was set to p-value 0.05. Cluster size calculated by summing all voxels in cluster. Effect direction indicates betas sign, and demonstrates risk loci associated with higher gray matter density (+) or lower gray matter density (-).

RS ID	Locus [a]	Brain Regions [b]	Minimum p-value	Cluster size	Effect direction
rs429358/rs7412	APOEε4	Postcentral gyrus right	6.92x10-6	1411	+
		Remainder of parietal lobe left	2.14x10-5	1051	+
		Posterior temporal lobe right	4.17x10-4	786	-
rs10792832	PICALM	Superior frontal gyrus left	4.47x10-4	1502	+
		Middle frontal gyrus left	9.33x10-5	1206	-
		Cingulate gyrus, anterior part left	2.24x10-4	1871	-
rs10838725	CELF1	Superior frontal gyrus right	4.68x10-5	2228	-
		Medial orbital gyrus left	1.41x10-4	1907	+
		Superior temporal gyrus left, posterior part	1.74x10-4	493	-
rs10948363	CD2AP	Superior parietal gyrus right	2.95x10-4	273	+
		Middle frontal gyrus right	3.39x10-4	280	+
		Middle frontal gyrus left	4.17x10-4	238	-
rs11218343	SORL1	Inferior frontal gyrus right	1.62x10-5	3150	+
		Insula left	6.46x10-5	1822	+
		Middle frontal gyrus left	1.07x10-4	2020	-
rs11771145	EPHA1	Precentral gyrus right	8.91x10-6	3185	-
		Remainder of parietal lobe right	4.27x10-5	385	-
		Insula right	3.02x10-4	2295	-
rs1476679	ZCWPW1	Superior frontal gyrus right	5.13x10-6	1255	+
		Superior frontal gyrus left	1.66x10-4	578	-
		Lateral remainder of occipital lobe left	1.86x10-4	3592	+
rs17125944	FERMT2	Lateral remainder of occipital lobe right	5.01x10-5	1733	+
		Superior parietal gyrus left	1.32x10-4	285	-
		Middle frontal gyrus right	1.45x10-4	1260	+
rs190982	MEF2C	Lateral remainder of occipital lobe right	9.55x10-6	2322	+
		Lateral remainder of occipital lobe left	5.88x10-5	461	-
		Middle frontal gyrus right	1.46x10-4	865	+
rs2718058	NME8	Postcentral gyrus right	2.88x10-5	528	+
		Medial and inferior temporal gyri right	6.46x10-5	2055	+
		Posterior temporal lobe right	1.46x10-4	786	+
rs28834970	PTK2B	Remainder of parietal lobe right	2.95x10-5	2014	+
		Lateral remainder of occipital lobe left	8.32x10-5	1461	+
		Lingual gyrus right	3.47x10-4	751	+

SUPPLEMENTARY TABLE 2 CONTINUED: Top 3 associated clusters of voxels per each risk locus ordered by minimum p-value. Cluster forming threshold was set to p-value 0.05. Cluster size calculated by summing all voxels in cluster. Effect direction indicates betas sign, and demonstrates risk loci associated with higher gray matter density (+) or lower gray matter density (-).

RS ID	Locus [a]	Brain Regions [b]	Minimum p-value	Cluster size	Effect direction
rs35349669	INPP5D	Superior frontal gyrus left	9.33×10^{-5}	820	-
		Posterior temporal lobe left	1.46×10^{-4}	798	+
		Insula right	1.59×10^{-4}	1252	+
rs4147929	ABCA7	Postcentral gyrus left	4.47×10^{-6}	2335	-
		Lateral remainder of occipital lobe left	1.35×10^{-4}	2223	+
		Middle frontal gyrus right	1.38×10^{-4}	498	-
rs6656401	CR1	Precentral gyrus right	4.17×10^{-5}	2802	-
		Superior frontal gyrus left	6.46×10^{-5}	1813	+
		Lateral remainder of occipital lobe right	1.32×10^{-4}	473	+
rs6733839	BIN1	Anterior orbital gyrus right	3.39×10^{-5}	2463	-
		Insula right	1.35×10^{-4}	1100	-
		Superior temporal gyrus anterior part right	2.24×10^{-4}	255	+
rs7274581	CASS4	Cingulate gyrus anterior part right	3.02×10^{-4}	2036	-
		Middle frontal gyrus left	3.09×10^{-4}	758	-
		Cingulate gyrus, posterior part left	4.68×10^{-4}	2382	-
rs9331896	CLU	Posterior temporal lobe right	10^{-4}	3176	-
		Middle frontal gyrus left	3.80×10^{-4}	680	-
		Cingulate gyrus, posterior part left	4.68×10^{-4}	1574	+
rs983392	MS4A6A	Lateral remainder of occipital lobe left	1.77×10^{-5}	791	+
		Posterior temporal lobe left	4.68×10^{-5}	1126	+
		Superior frontal gyrus right	6.03×10^{-4}	557	+
rs10498633	SLC24A4	Precentral gyrus left	2.95×10^{-5}	411	-
		Precentral gyrus right	8.51×10^{-5}	1555	-
		Remainder of parietal lobe right	10^{-4}	2037	+

[a] Assigned risk locus according to Lambert et al [1]

[b] Brain region labeling based on Hammer Atlas segmentation.

SUPPLEMENTARY TABLE 3: Coded structure list for brain regions according to Hammer Atlas segmentation.

Number on heatmap	Name of Structure
	Temporal Lobe
1	Hippocampus right, left
2	Amygdala right, left
3	Anterior temporal lobe, medial part right, left
4	Anterior temporal lobe, lateral part right, left
5	Parahippocampal and ambient gyri right, left
6	Superior temporal gyrus, posterior part right, left
7	Middle and inferior temporal gyrus right, left
8	Fusiform gyrus right, left
9	Posterior temporal lobe right, left
10	Superior temporal gyrus, anterior part right, left
	Insula and Cingulate gyri
11	Insula left, right
12	Cingulate gyrus (gyrus cinguli), anterior part left, right
13	Cingulate gyrus (gyrus cinguli), posterior part left, right
	Frontal Lobe
14	Middle frontal gyrus left, right
15	Precentral gyrus left, right
16	Straight gyrus left, right
17	Anterior orbital gyrus left, right
18	Inferior frontal gyrus left, right
19	Superior frontal gyrus left, right
20	Medial orbital gyrus left, right
21	Lateral orbital gyrus left, right
22	Posterior orbital gyrus left, right
23	Subgenual frontal cortex left, right
24	Subcallosal area left, right
25	Pre-subgenual frontal cortexleft, right
	Occipital Lobe
26	Lingual gyrus left, right
27	Cuneus left, right
28	Lateral remainder of occipital lobe left, right
	Parietal Lobe
29	Postcentral gyrus left, right
30	Superior parietal gyrus left, right
31	Inferio lateral remainder of parietal lobe left, right
	Central Structures
32	Caudate nucleus left, right
33	Nucleus accumbens left, right
34	Putamen left, right
35	Thalamus left, right
36	Pallidum left, right

SUPPLEMENTARY TABLE 4: Results of spatial overlap between VBM risk loci association and gene expression profiles of 3 Caucasian donors from the Allen Human Brain Atlas. The table shows genes in risk loci regions, for which expression differs significantly (at corrected threshold 1.7×10^{-5}) from background expression in regions associated by VBM analysis.

Risk variant	Putative causal gene[a]			Genes showing significant overlap				
	Locus	Minimum p-value	Distance from risk loci, bp	Significant gene expression	Minimum p-value	Distance from risk allele, bp	Significant donors/ total number of donors	Significant probes/total number of probes
rs6656401	CR1	5.58×10^{-4}	22.557	CD55	1.13×10^{-5}	197.196	1/3	1/4
rs6733839	BIN1	0.11	87.207	IWS1	8.07×10^{-7}	-300.973	1/3	1/2
rs35349669	INPP5D	3.80×10^{-4}	143.799	NGEF	7.57×10^{-16}	325.080	3/3	2/2
rs190982	MEF2C	1.41×10^{-5}	44.275	DGKD	1.69×10^{-8}	-194.677	3/3	2/2
rs10948363	CD2AP	0.04	42.237	MEF2C	1.41×10^{-5}	44.275	1/3	1/3
rs27180058	NME8	4.48×10^{-2}	-46.665	EPDR1*	9.05×10^{-5}	118.088	-	-
rs1476679	ZCWPW1	7.21×10^{-5}	5.997	PCOLCE	8.39×10^{-8}	-195.354	1/3	1/2
rs11771145	EPHA1	1.81×10^{-4}	23.380	GSTK1	1.01×10^{-5}	169.576	2/3	3/4
rs28834970	PTK2B	2.61×10^{-4}	26.122	-	-	-	-	-
rs9331896	CLU	4.43×10^{-7}	13.252	CLU	4.43×10^{-7}	13.252	1/3	1/1
				SCARA3	3.18×10^{-6}	-296.264	1/3	3/3
				CCDC25	7.95×10^{-6}	-123.149	1/3	1/3
rs10838725	CELF1	5.59×10^{-3}	70.375	-	-	-	-	-
rs983392	MS4A6A	1.61×10^{-2}	-15.573	-	-	-	-	-
rs10792832	PICALM	1.35×10^{-3}	199.148	SYTL2	5.33×10^{-8}	462.608	2/3	3/4
rs112183343	SORL1	1.99×10^{-4}	112.675	SC5D	1.76×10^{-6}	272.425	1/3	2/3
rs17125944	FERMT2	2.98×10^{-2}	76.643	-	-	-	-	-
rs10498633	SLC24A4	1.50×10^{-5}	138.027	SLC24A4	1.50×10^{-5}	138.027	1/3	1/2
rs4147929	ABCA7	8.35×10^{-5}	23.341	PALM	1.92×10^{-9}	354.490	2/3	2/3
rs429358/rs7412	APOE	3.94×10^{-4}	-	-	-	-	-	-
rs7274581	CASS4	9.34×10^{-4}	31.092	C20orf43	1.64×10^{-5}	-25.387	1/3	1/3

* Not significant at Bonferroni corrected threshold, but more significant, then originally assigned locus; [a] Assigned risk locus according to Lambert et al [1]

SUPPLEMENTARY TABLE 5: The most significant voxel-wise association signals with p-values$<10^{-5}$ in participants carrying the APOEε4 allele. Brain region labeling based on the Hammer Atlas segmentation. Effect direction indicates beta sign, and demonstrates risk loci associated with increasing gray matter tissue (+) or decreasing gray matter tissue (-).

Risk variant	Gene*	Minimum p-value	Effect direction	Brain Region
rs190982	MEF2C	7.36×10^{-8}	+	Superior temporal gyrus, central part left
rs10948363	CD2AP	8.14×10^{-7}	-	Substantia nigra left
rs10838725	CELF1	4.70×10^{-6}	-	Posterior temporal lobe right
rs2718058	NME8	8.78×10^{-6}	-	Lateral remainder of occipital lobe left
Genetic Risk Score	All	1.20×10^{-6}	+	Precentral gyrus right
Genetic Risk Score	Without APOE	0.0002**	-	Postcentral gyrus left

* Assigned risk gene according to Lambert et al [1]

** P-value is not less than 10^{-5}, shown to compare with GRS without exclusion APOE.

SUPPLEMENTARY TABLE 6: The most significant voxel-wise association signals in participants not carrying the APOEε4 allele. Brain region labeling based on the Hammer Atlas segmentation. Effect direction indicates beta sign, and demonstrates risk loci associated with increasing gray matter tissue (+) or decreasing gray matter tissue (-).

Risk variant	Gene*	Minimum p-value	Effect direction	Brain Region
rs11218343	SORL1	1.57×10^{-5}	+	Superior temporal gyrus, central part left
rs10838725	CELF1	1.99×10^{-5}	+	Medial orbital gyrus left
rs17125944	FERMT2	2.29×10^{-5}	-	Lateral remainder of occipital lobe left
Genetic Risk Score	Without APOE	0.0002**	-	Cingulate gyrus anterior (supragenual) part left

* Assigned risk gene according to Lambert et al [1]

SUPPLEMENTARY FIGURE 1: Projection of risk loci association clusters from VBM to cortical surface. Colors reflect regression association: blue for negative, red for positive. Clusters formed based on nominal significant p-value threshold 0.05.

VBM association maps of ABCA7**(A)**, ZCWPW1**(B)**, EPHA1**(C)** and MEF2C**(D)** risk loci.

A
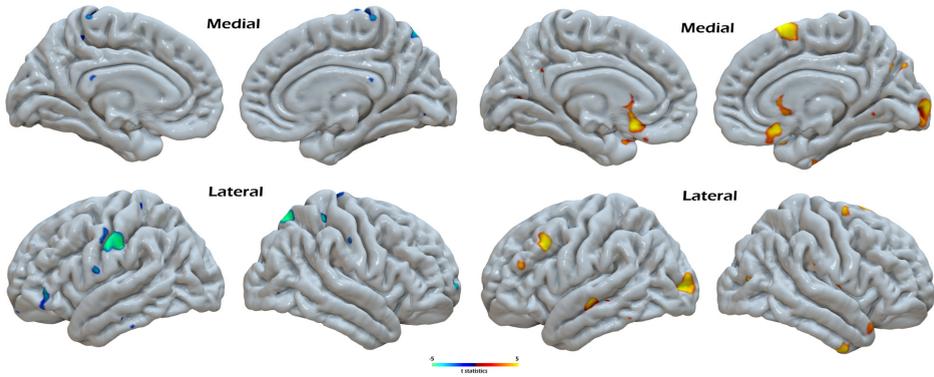
B
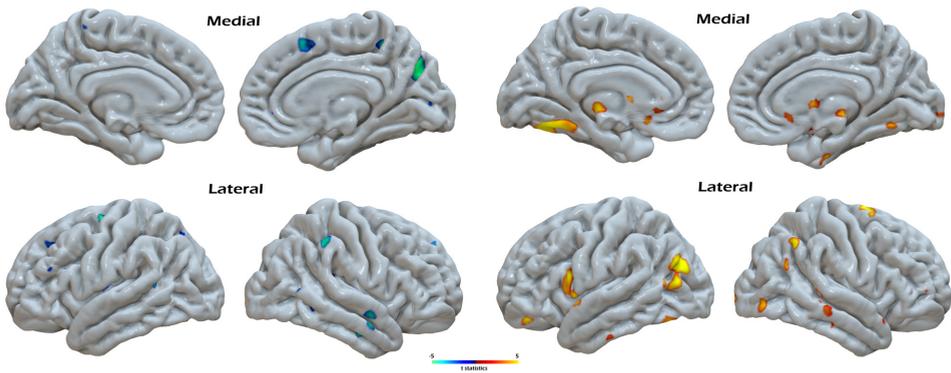
SUPPLEMENTARY FIGURE 1 CONTINUED: Projection of risk loci association clusters from VBM to cortical surface.

Colors reflect regression association: blue for negative, red for positive. Clusters formed based on nominal significant p-value threshold 0.05.

VBM association maps of ABCA7**(A)**, ZCWPW1**(B)**, EPHA1**(C)** and MEF2C**(D)** risk loci.

C
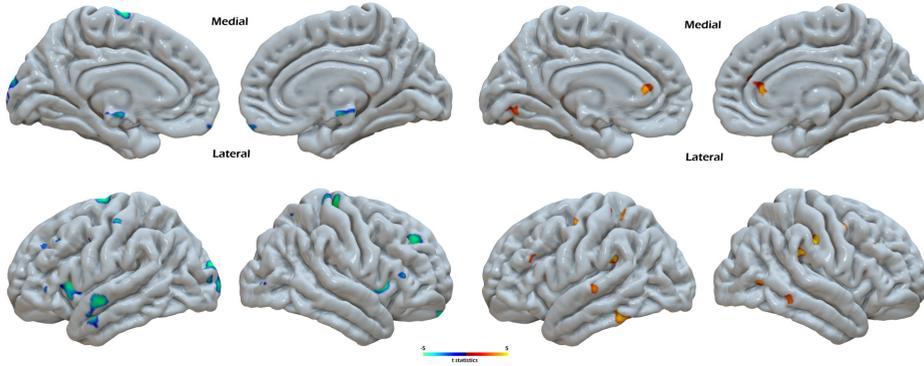
D
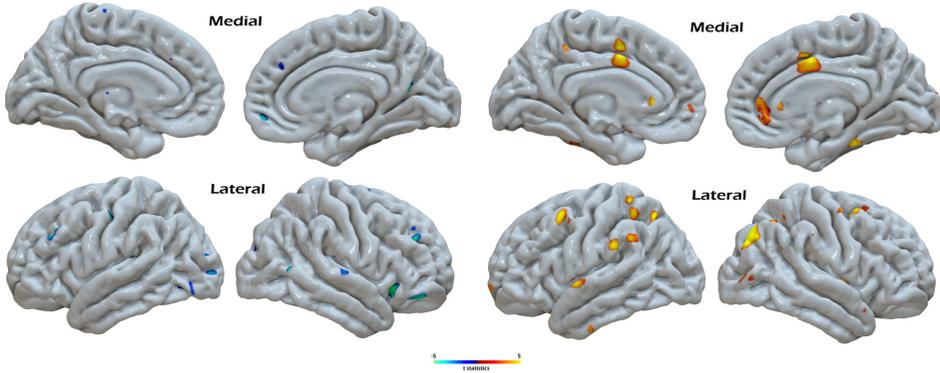
SUPPLEMENTARY FIGURE 2: Region brain Manhattan plots. Effect of AD SNPs shown for the same voxels from selected brain region. **A**- Postcentral gyrus right; **B** - Lateral remainder of occipital lobe right. The detail results are available online: www.imagene.nl/ADSNPs/.
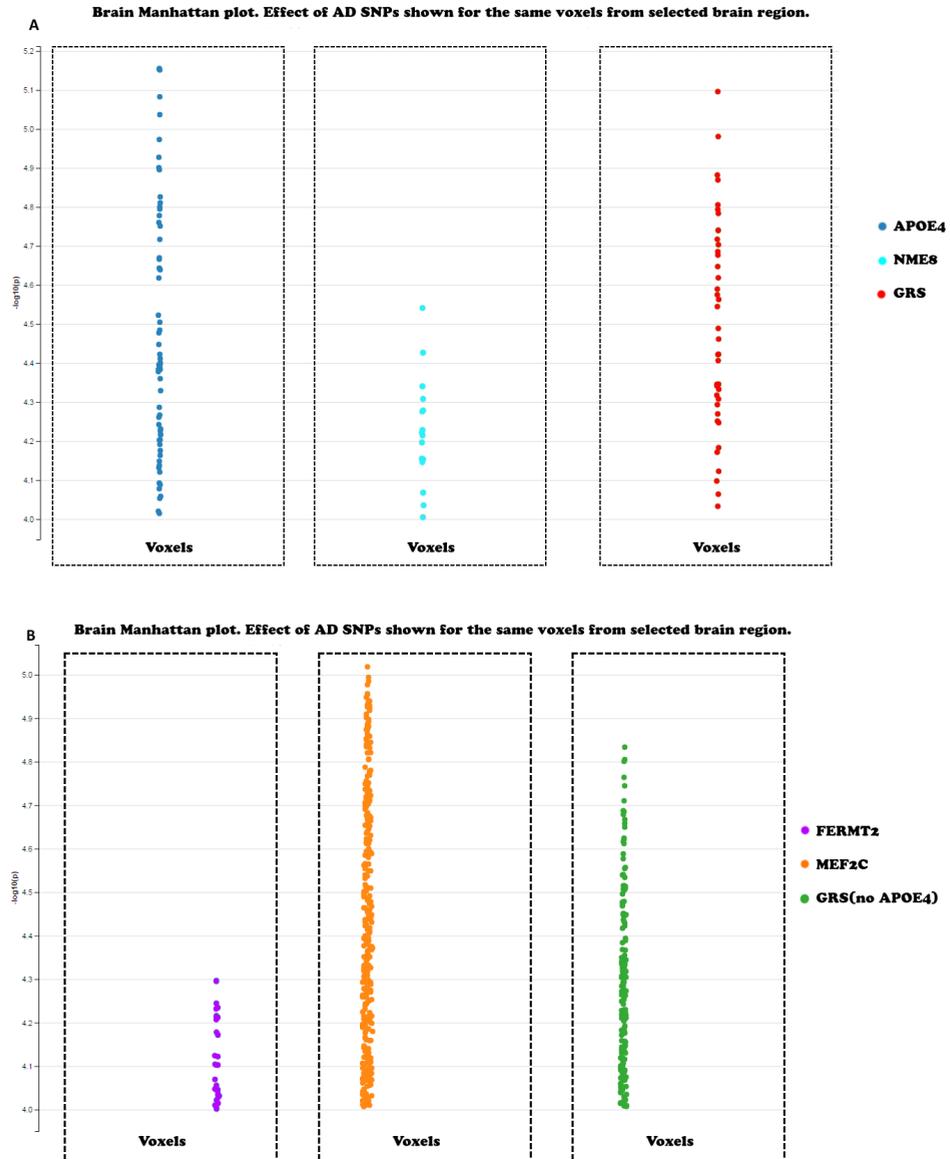
CHAPTER 3.2

High-dimensional mapping of cognition to the brain using voxel-based morphometry and subcortical shape analysis

ABSTRACT

Background

It is increasingly recognized that the complex functions of human cognition are not accurately represented by arbitrarily defined anatomical brain regions. A considerable functional specialization typically exists within such regions. Studies that examine the association between cognition and brain structure at the highest resolution would allow better map such localized associations. However, such analysis in a large community-dwelling population is lacking.

Methods

In 3,813 stroke-free and non-demented persons from the Rotterdam Study (mean age 69.1 ($\pm$8.8) years; 55.8% women) with cognitive assessments and brain MRI, we performed voxel-based morphometry and subcortical shape analysis on global cognition and separate tests that tapped into memory, information processing speed, fine motor speed, and executive function domains.

Results

We found that the different cognitive tests significantly associated with grey matter voxels in differential but also overlapping brain regions, primarily in the left hemisphere. Clusters of significantly associated voxels with global cognition were located within multiple anatomic regions: left amygdala, hippocampus, parietal lobule, superior temporal gyrus, insula and posterior temporal lobe. Subcortical shape analysis revealed associations primarily within the head and tail of caudate nucleus, putamen, ventral part of the thalamus, and nucleus accumbens, more equally distributed among the left and right hemisphere. Within caudate nucleus both positive (head) as well as negative (tail) associations were observed with global cognition.

Conclusions

In a large population-based sample, we mapped cognitive performance to (sub)cortical grey matter using a hypothesis-free approach with high-dimensional neuroimaging. Leveraging the power of our large sample size, we confirmed well-known associations as well as identified novel brain regions affecting cognition.

INTRODUCTION

Human cognition comprises a variety of important domains including memory, information processing speed and executive function. Cognitive ability is associated with important health outcomes(Eggermont, et al., 2012; Gale, et al., 2008; Halperin, et al., 2008) and varies between individuals and throughout life(Halperin, et al., 2008). It is determined by both genetic and environmental factors(Deary, et al., 2010; Haworth, et al., 2010), which are reflected in the structure of the brain(Davies, et al., 2015; Kramer, et al., 2004).

Many of the initial links between brain structure and cognition arose from clinical observations of patients with localized brain lesions or following surgical interventions(Newman, et al., 2007; Rorden and Karnath, 2004). Subsequent neuroimaging studies have used these observations in hypothesis-driven approaches to study the neural substrate of human cognitive ability including its various domains(Elderkin-Thompson, et al., 2008; Newman, et al., 2007). These studies have primarily focused on aggregate measures over the entire brain regions e.g., volumetric measures of the prefrontal cortex(Salat, et al., 2004; Tisserand, et al., 2004) , thalamus(Van Der Werf, et al., 2001) or hippocampus(Van Petten, 2004). However, it is increasingly recognized that the complex functions of human cognition are not accurately represented by anatomical regions that are arbitrarily defined based on macroscopical landmarks or histological microstructure, e.g. Brodmann areas(Bola and Sabel, 2015; Bressler and Menon, 2010; Park and Friston, 2013). Moreover a considerable functional specialization typically exists within such regions. For example, in Alzheimer's disease the size of hippocampal subfields contains information important for cognition beyond the gross hippocampal volume(de Flores, et al., 2015; La Joie, et al., 2013; Lindberg, et al., 2012). The thalamus comprises more than 60 cytoarchitectonically and functionally distinct nuclei, all of which have a different pattern of anatomical connections to other brain regions (Fama and Sullivan, 2015; Schmahmann and Pandya, 2008).

An alternative to hypothesis-driven analyses are hypothesis-free approaches that interrogate brain structure at the highest resolution and provide the opportunity to explore the association beyond just aggregated measures. For instance, voxel-based morphometry (VBM) studies volumetric differences at the level of the voxel, the smallest unit of measure of an MRI scan. In recent years, a large body of literature has emerged that employs VBM and related techniques to study how brain structures relate to cognition(Burgaleta, et al., 2014; Fleischman, et al., 2014; Ruscheweyh, et al., 2013; Squarzoni, et al., 2012; Tisserand, et al., 2004). (Frisoni GB et al. 2002; Guo X et al. 2010; Minkova L et al. 2017)However, still several knowledge gaps remain: First, many hypothesis-free brain imaging studies are performed in relatively small studies, thereby running the risk of false-positive findings or not significant results.

Larger sample sizes can overcome this restriction and yield more robust findings as well as generalize the previous. Second, in addition to VBM analysis, the shape of subcortical structures allows to study the brain regions beyond just volumetric measures(Roshchupkin, et al., 2016b) and may represent underlying subfield organization(Wang, et al., 2008).

Therefore, using hypothesis-free approaches of voxel-based morphometry and shape analysis we performed a fine mapping of cognitive ability to (sub)cortical grey matter on magnetic resonance imaging (MRI) in a large population-based sample of middle-aged and elderly subjects.

MATERIALS AND METHODS

STUDY POPULATION

This study was embedded within the Rotterdam Study, an ongoing prospective population-based cohort designed to investigate chronic diseases in the middle-aged and elderly population(Ikram, et al., 2017). The cohort started in 1990 and comprised 7,983 participants aged ≥55 years. In 2000 and 2006, the study was expanded and at present comprises 14,926 participants aged ≥45 years. Since 2005, brain MRI was implemented into the study protocol(Ikram, et al., 2015). Between 2009 and 2014 4,140 participants underwent brain MRI and cognitive testing. Examinations in this time period were conducted as one project. We excluded participants due to prevalent dementia (n=42), insufficient cognitive screening (n=21), with cortical infarcts (n=103) or clinical stroke (n=161). In total, 3,813 participants were available for analysis. The Rotterdam Study has been approved by the medical ethics committee according to the Population Study Act Rotterdam Study, executed by the Ministry of Health, Welfare and Sports of the Netherlands. A written informed consent was obtained from all participants.

MRI ACQUISITION

Brain MRI was performed on a 1.5-T MRI scanner (Signa Excite II, General Electric Healthcare, Milwaukee, WI,USA) using an eight-channel head coil. The protocol included T1-weighted sequence (T1), proton density-weighted sequence, and a T2-weighted fluid-attenuated inversion recovery (FLAIR) sequence, as described extensively in detail before(Ikram, et al., 2015).

VOXEL BASED MORPHOMETRY

Voxel based morphometry (VBM) was performed according to an optimized VBM protocol(Good, et al., 2001) and as previously described(Roshchupkin, et al., 2016a). Briefly, all T1-weighted images were segmented into supratentorial grey matter, white matter and cerebrospinal fluid using a previously described k-nearest neighbor algorithm, which was trained on six manually labeled atlases(Vrooman, et al., 2007). All grey matter (GM) density maps were non-linearly registered to the standard ICBM MNI152 grey matter template (Montreal Neurological Institute) with a 1x1x1 mm^3 voxel resolution. A spatial modulation procedure was used to avoid differences in absolute grey matter volume due to the registration, following by smoothing procedure, using a 3mm (FWHM 8mm) isotropic Gaussian kernel.

SUBCORTICAL SHAPES

The T1-weighted MRI scans were processed using FreeSurfer(Fischl, et al., 2004) (version 5.1) to obtain segmentations and volumetric summaries of the following seven subcortical structures for each hemisphere: nucleus accumbens, amygdala, caudate, hippocampus, pallidum,

putamen, and thalamus(Roshchupkin, et al., 2016c). Next, segmentations were processed using a previously described shape analysis pipeline(Gutman, 2015; Roshchupkin, et al., 2016c). Briefly, a mesh model was created for the boundary of each structure. Subcortical shapes were registered using the "Medial Demons" framework, which matches shape curvatures and medial features to a pre-computed template(Gutman, et al., 2013). The templates and mean medial curves were previously constructed and are distributed as part of the ENIGMA-Shape package (http://enigma.usc.edu/ongoing/enigma-shape-analysis/). The resulting meshes for the 14 structures consist of a total of 27,120 vertices(Roshchupkin, et al., 2016c). Two measures were used to quantify shape: the radial distance and the natural logarithm of the Jacobian determinant. The radial distance represents the distance of the vertex from the medial curve of the structure. The Jacobian determinant captures the deformation required to map the subject-specific vertex to a template and indicates shape dilation due to sub-regional volume change(Roshchupkin, et al., 2016c).

ASSESSMENT OF COGNITIVE FUNCTIONING

Cognitive function was assessed with a cognitive test battery comprising Stroop test(Houx, et al., 1993) (word reading, color naming and a reading/color naming interference task (error-adjusted time in seconds)), which tests information processing speed and executive function; 15-Word learning test (15-WLT)(Bleecker, et al., 1988), which taps into immediate and delayed recall to investigate memory; Letter-digit substitution task (LDST)(Prins, et al., 2005) and Word fluency test (WFT, animal categories), both of which test executive function. The three Stroop tests were natural log transformed due to a skewed distribution. To allow for comparison across cognitive tests, we calculated z scores (subtracting the population mean and dividing by the standard deviation) for each cognitive test. The z scores for the Stroop Tests were inverted because higher scores on the Stroop test indicate a poorer performance, whereas higher scores on the other cognitive tests indicate better cognitive performance. In addition, we also investigated global cognition by calculating a compound score (G-factor) using a principal component analysis on the delayed recall score of the 15-WLT, Stroop Interference Test, LDST, and WFT(Hoogendam, et al., 2014). The G-factor explained 57.2% of the variance in cognitive test scores in the population.

OTHER MEASUREMENTS

Attained level of education was collected and expressed in years. Prevalent clinical stroke and dementia were assessed based on a protocol as previously described.(Bos, et al., 2014; de Bruijn, et al., 2015)

STATISTICAL ANALYSIS

For VBM and shape analyis, linear regression models were fitted with age, sex, education and

cognitive test value as independent variables and voxel or vertex measure as the dependent variable. We corrected for level of education as a measure of cognitive reserve. For both VBM and shapes analysis we computed the significance threshold based on nonparametric statistic test by performing 10,000 random permutations.(Churchill and Doerge, 1994) After collecting the minimum p-value from every test, they were sorted and the 5% quantile was used (α=0.05) to estimate the p-value significant threshold, while controlling the family wise error (FWE). The resulting values were 2.99×10^{-7} for VBM and 9.63×10^{-6} for shapes, which were subsequently divided by the number of cognitive tests (n=8) to account for multiple hypothesis tests correction.

RESULTS

Characteristics of the study population are presented in **Table 1**. Of the 3,813 participants, 55.8% were women and mean age was 69.1 years (ranging from 51.9 to 97.9 years). Correlations between all cognitive test scores stratified by sex are shown in **Supplementary Figure 1**.

TABLE 1: Characteristics of the Study Population.

	Total population n= 3,813
Age, years	69.1 (8.8)
Female	2129 (55.8)
Education, years	12.7 (3.9)
Left hemisphere, cm³	
Nucleus accumbens	0.55 (0.09)
Amygdala	1.30 (0.21)
Caudate nucleus	3.37 (0.55)
Hippocampus	3.82 (0.63)
Pallidum	1.46 (0.23)
Putamen	4.56 (0.63)
Thalamus	6.15 (0.73)
Right hemisphere, cm³	
Nucleus accumbens	0.48 (0.09)
Amygdala	1.39 (0.22)
Caudate nucleus	3.48 (0.56)
Hippocampus	3.84 (0.59)
Pallidum	1.39 (0.24)
Putamen	4.40 (0.61)
Thalamus	6.16 (0.73)

Data presented as mean (standard deviation) for continuous variables and number (percentages) for categorical variables. The following variable had missing data: education (n=52).

VOXEL-BASED MORPHOMETRY ANALYSIS

In total 4,081 of the grey matter voxels were significant in relation to at least one cognitive tests and/or global cognition after correction for multiple-testing (**Table 2**). These significant voxels were clustered within different brain structures, and almost exclusively within the left hemisphere. The strongest positive associations of better global cognition (G-factor) with grey matter voxels were found in the left amygdala (156 voxels, minimum (min) p-value 4.2×10^{-12}), hippocampus (173 voxels, min p-value 9.6×10^{-12}), parietal lobule (517 voxels, min p-value 1.2×10^{-10}), superior temporal gyrus (313 voxels, min p-value 1.5×10^{-10}), insula (142 voxels, min p-value 7.4×10^{-10}), posterior temporal lobe left (101 voxels, min p-value 7.7×10^{-10}), postcentral gyrus, inferior and middle frontal gyrus, posterior orbital gyrus and right caudate nucleus (all <25 voxels, min p-value all $< 2.9 \times 10^{-8}$) (**Figure 1**).

TABLE 2: Association between cognitive tests and grey matter voxels.

Cotgnitive Test	Brain Region	Region size (#voxels)	min p-value	# negative voxels	# positive voxels
G-factor	Amygdala left	2192	4.24E-12	0	156
	Hippocampus left	3183	9.62E-12	0	173
	Parietal lobe left (including supramarginal and angular gyrus)	65794	1.19E-10	0	517
	Superior temporal gyrus central part left	20954	1.48E-10	0	313
	Insula left	21930	7.43E-10	0	142
	Posterior temporal lobe left	70660	7.68E-10	0	101
	Inferior frontal gyrus left	28631	9.18E-10	0	18
	Middle frontal gyrus left	80119	2.17E-09	0	15
	Postcentral gyrus left	46092	2.92E-09	0	23
	Posterior orbital gyrus left	7439	4.46E-09	0	2
	Caudate nucleus right	6170	2.86E-08	0	4
WLT delayed recall	Hippocampus left	3183	4.84E-10	0	466
	Amygdala left	2192	1.13E-09	0	385
	Insula left	21930	1.67E-08	0	10
	Gyri parahippocampalis et ambiens left	6823	3.45E-08	0	1
WLT immediate recall	Hippocampus left	3183	2.91E-08	0	6
WFT	Superior temporal gyrus, central part left	20954	8.75E-09	0	103
	Parietal lobe left (including supramarginal and angular gyrus)	65794	8.81E-09	0	81
Stroop interference	Superior temporal gyrus, central part left	20954	5.29E-11	0	442
	Parietal lobe left (including supramarginal and angular gyrus)	65794	5.67E-11	0	655
	Postcentral gyrus left	46092	1.86E-10	0	80
	Posterior temporal lobe left	70660	2.00E-10	0	152
	Insula left	21930	2.52E-10	0	137
	Amygdala left	2192	9.21E-10	0	46
	Hippocampus left	3183	1.37E-09	0	24
	Caudate nucleus left	6059	1.62E-08	0	21
LDST	Insula left	21930	2.62E-08	0	7
	Inferior frontal gyrus left	28631	2.72E-08	0	1
Total # positive voxels					4081
Total # negative voxels				0	
Total voxels					**4081**

All associations are adjusted for age, sex, and education. Abbreviations: G-factor global cognition; WLT Word learning test; WFT Word fluency test; LDST Letter-digit substitution task.
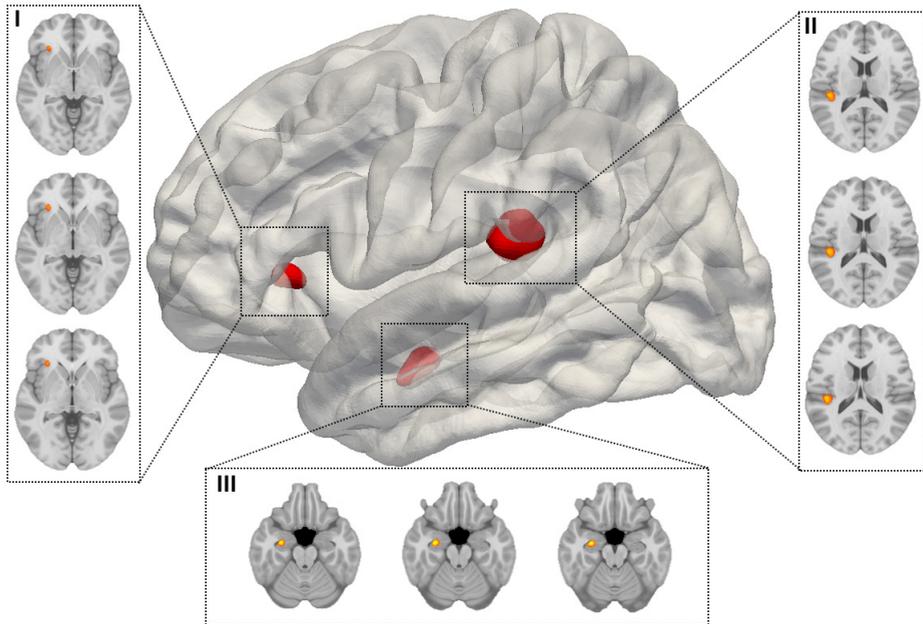
FIGURE 1: Association of grey matter voxel density with global cognition. Lateral view of the left hemisphere. FWE-significant voxels, indicated by red-yellow, cluster in the insular cortex (I), Wernicke's area (II), and the hippocampus (III). All associations are adjusted for age, sex, and education. Neurological orientation axial images: left = left.

With respect to the separate cognitive tests, a higher score on the delayed recall task of the 15-WLT was positively associated with grey matter voxels in parts of the left hippocampus (466 voxels, min p-value 4.9×10^{-10}), amygdala (385 voxels, min p-value 1.1×10^{-9}), insula (10 voxels, min p-value 1.7×10^{-8}), and gyri parahippocampalis et ambiens (1 voxel, min p-value 3.5×10^{-8}). A higher score on immediate recall score of the 15-WLT was positively related to a small portion of the left hippocampus (6 voxels, min p-value 2.9×10^{-8}). We observed that Word-Fluency test (WFT) was associated with superior temporal gyrus (103 voxels, min p-value 8.8×10^{-9}), and parietal lobule (81 voxels, min p-value 8.8×10^{-9}). We did not observe any association with the Stroop Reading test or the Stroop Color Naming test that survived correction for multiple testing. Stroop interference task harbored significant associations in the left hemisphere including superior temporal gyrus (442 voxels, min p-value 5.3×10^{-11}), the parietal lobule (655 voxels, min p-value 5.7×10^{-11}), postcentral gyrus (80 voxels, min p-value 1.9×10^{-10}), posterior temporal lobe (152 voxels, min p-value 2.0×10^{-10}), insula (137 voxels, min p-value 2.5×10^{-10}), amygdala (46 voxels, min p-value 9.2×10^{-10}), hippocampus and caudate nucleus (both <25 voxels, min p-value 1.6×10^{-8}). LDST was associated with a

small cluster of voxels in the left insula (7 voxels, min p-value 2.6×10^{-8}), and inferior frontal gyrus (1 voxel, min p-value 2.7×10^{-8}). All of these results are depicted in **Figure 2A-E**.
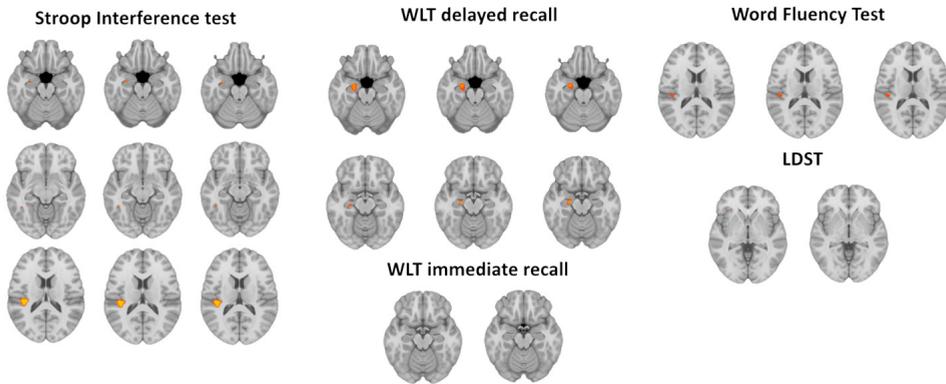

FIGURE 2: FWE-significant grey matter voxels in relation to cognitive tests. FWE-significant grey matter voxels, indicated by red-yellow, in relation to cognitive tests. Neurological orientation: left =left. All associations are adjusted for age, sex, and education. Abbreviations: WLT Word learning test; LDST Letter-digit substitution task.

SHAPE ANALYSIS

Jacobian determinant and radial distance showed clusters of significant FWE-corrected vertices in relation to cognitive tests, distributed among the left and right hemisphere: 2,819 and 2,298 respectively (**Supplementary Table 2**). The thalamus, caudate nucleus, and putamen harbored most significant associations (**Supplementary Table 2**). Largest significant clusters were found for the Jacobian determinant of the left and right thalamus with Stroop interference task (369 vertices, min p-value 5.8×10^{-11} and 324 vertices, min p-value 6.1×10^{-12} respectively). Global cognition harbored several significant associations, including the Jacobian determinant of the left and right thalamus (281 vertices, min p-value 4.4×10^{-12} and 159 vertices, min p-value 1.9×10^{-9} respectively), and the radial distance of the left and right caudate nucleus (133 vertices, min p-value 2.9×10^{-16} and 78 vertices, min p-value 2.9×10^{-13} respectively) (**Figure 3**). A few inverse associations were observed, primarily between the Jacobian determinant of the caudate nucleus and the WLT (both delayed and immediate recall), and G-factor. Small clusters of vertices (ranging from 1 to 22) were found in the hippocampus with global cognition, WFT and Stroop interference task, but not with the memory tests. **Supplementary Figure 2A-F** shows all significant findings of the shape analysis of subcortical brain structures in relation to the individual cognitive tests.
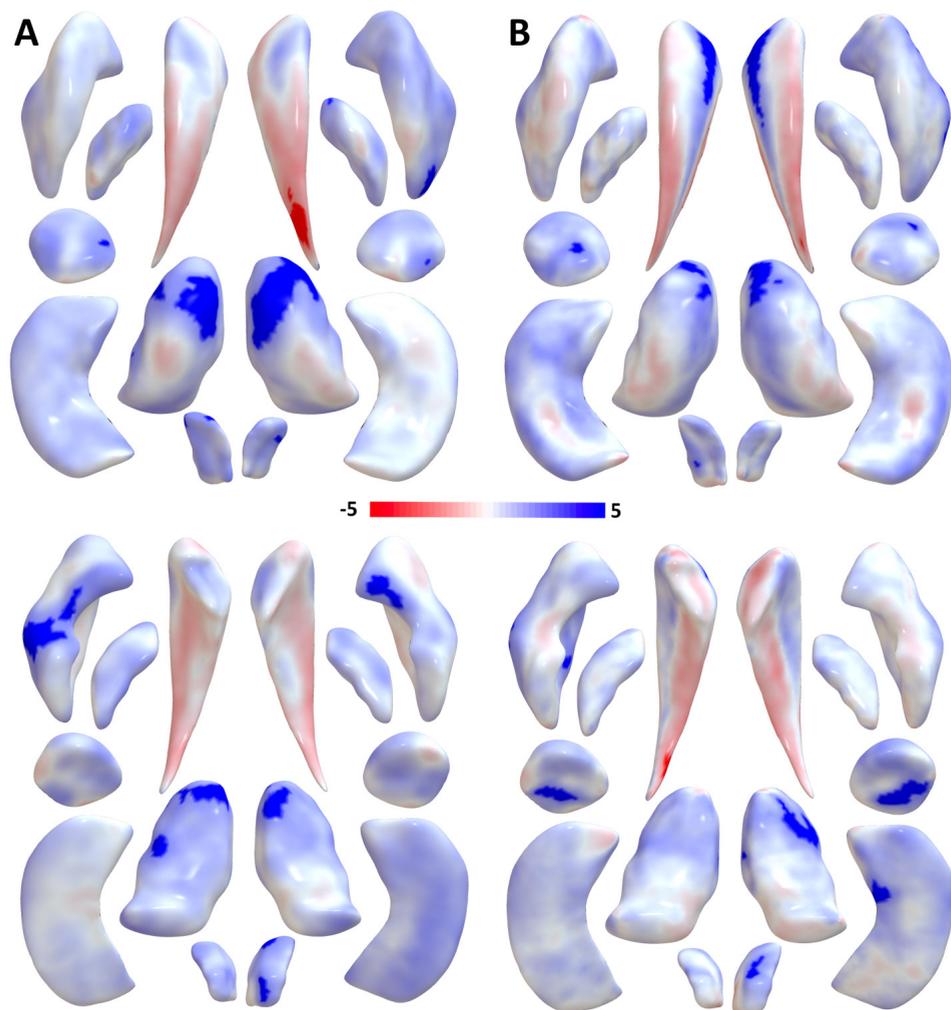
FIGURE 3: Maps of shape measures of subcortical brain regions in relation to global cognition. Maps show the associations of seven bilateral subcortical structures for the shape measures of Jacobian determinant (**Panel A**) and radial distance (**Panel B**), anterior (top row) and posterior (bottom row) view. All associations are adjusted for age, sex, and education. Color map represents the t-statistics and shows the direction of association, with red and blue indicating negative and positive associations respectively. Highlighted regions represent statistically significant vertices.

DISCUSSION

In this large study of community-dwelling adults, we presented the neuroanatomical fine-mapping of seven cognitive tests and global cognition using voxel-based morphometry and subcortical shape analysis. We found that the different cognitive tests significantly associated with grey matter voxels in different brain regions, primarily in the left hemisphere. Moreover, many of associated regions overlap between cognitive tests. Subcortical shape analysis revealed associations primarily within the head and tail of caudate nucleus, putamen, ventral part of the thalamus, and nucleus accumbens, more equally distributed among the left and right hemisphere. Within caudate nucleus both positive (head) as well as negative (tail) associations were observed with global cognition.

Regarding the voxel-based morphometry analysis, we observed three clusters of grey matter voxels to be associated with global cognition. These clusters were found in the left amygdala, hippocampus, parietal lobule, insula, posterior temporal lobe, inferior and middle frontal gyrus, postcentral gyrus and posterior orbital gyrus. Importantly, each of these three clusters was located within multiple anatomic regions. This may emphasize the importance of investigating the association with cognition beyond anatomically defined regions. Global cognition represents the shared variance of the individual cognitive tests, so it is therefore unsurprising that the three significant clusters are also significantly associated with the separate tests. We therefore will discuss our findings in more detail per individual cognitive test below.

Memory research has a long history(Squire and Zola-Morgan, 1988). The medial temporal lobes, and in particular the hippocampus, have long been implicated in episodic memory, with visuospatial memory predominantly associated with the right and verbal memory with the left hippocampus(Burgess, et al., 2002; Frisk and Milner, 1990; Smith and Milner, 1981). In line with this, we found that the 15-Word Learning test, with delayed recall more pronounced than immediate recall, was associated with clusters of grey matter voxels in particular the left hippocampus, as well as in the left amygdala. For decades there has been debate over the question of whether the amygdala is involved in memory(McGaugh, et al., 1996). Task-based resting-state functional MRI studies have shown that the amygdala is considered to play a role in emotional-related memory. However, its role in episodic memory is less known(Phelps, 2004). We did not observe an association between the shape of the hippocampus and tests measuring memory function. As was shown in a previous study, shape of subcortical structures represents different, a complimentary, phenotype compare to just volumetric measures, with its own genetic architecture (Roshchupkin, et al., 2016c). Therefore the absence of signal may be caused by the fact that the shape of hippocampus has also different functional specialization which is less sensitive to associations with cognition.

The Stroop interference test, Word Fluency test (WFT) and Letter-Digit Substitution Task (LDST) are all tapping into executive function. The Stroop interference task has been used extensively in studies designed to explore the efficiency of controlled attentional processes(Davidson, et al., 2003). The Stroop effect reflects slowing of response time or increase in error rate when persons are required to respond with the identity if an incongruent stimulus relative to a congruent stimulus(Bugg, et al., 2008). Interestingly, we observed that the Stroop interference task was positively associated with a cluster of grey matter voxels in the left hippocampus. Over the past decades, there has been increasing interest in the contribution of the hippocampus to processes beyond the memory domain(Rubin, et al., 2017). A study in healthy subjects explored the role of the hippocampus for response conflict in the Stroop task by combining intracranial electroencephalography with region of interest-based functional MRI. Researchers found that the hippocampus is recruited during response conflict. Importantly, it remains questionable whether conflict processing can be disentangled from circumstances in which there is conflicting valence or perceptual information, even in experimental studies that thoroughly control for the effect of memory(Ito and Lee, 2016). Moreover, WFT and Stroop interference test showed clusters of significant grey matter voxels in the left hemisphere where the posterior frontal lobe, upper segment of temporal lobe, and parietal lobule (including supramarginal and angular gyrus) intersect. These brain areas are part of Wernicke's area, a well-known functional language area(Pirmoradi, et al., 2016). The WFT being used in the current study tests semantic fluency. Semantic fluency requires searching for semantic associations within the lexicon(Shao, et al., 2014). Lower scores on semantic fluency tests may therefore also reflect problems with semantic memory, and not only executive function. In line with this, we found WFT to be associated with the left hippocampus, although non-significant after correction for multiple testing.

The LDST and Stroop interference test both showed associations with the left insula, more specifically the dorsal anterior insula. The left insular cortex is involved in consciousness and plays a role in diverse cognitive functions(Dupont, et al., 2003) such as higher cognitive processing and social-emotional processing(Chang, et al., 2013). Anterior insular cortices are among the most commonly activated brain regions across all cognitive tasks(Nelson, et al., 2010). It is also considered to be part of the cognitive control network and hypothesized that this network might form a pathway by which information in the insula, can affect decision making, and therefore influence information processing speed(Cauda, et al., 2011; Deen, et al., 2011).

In line with literature, our shape analysis results indicate that subcortical structures are heterogeneous and consist of functionally diverging sub-regions(de Flores, et al., 2015; Fama and Sullivan, 2015; Lin, et al., 2017; Roshchupkin, et al., 2016c). This is illustrated by, e.g., the caudate nucleus showing that its head and tail differ in their associations with global cog-

nition. It is thought that the head of the caudate nucleus interacts with medial, ventral, and dorsolateral prefrontal cortex as part of the 'cognitive' corticostriatal loop, whereas the tail interacts with inferior temporal areas and appears to be involved in visual stimulus processing(Haber, 2016; Lawrence, et al., 1998; Seger, 2013). In addition, our results suggest that the shape of other subcortical structures are involved in cognition as well, emphasizing the importance of subcortical shape analysis in understanding cognition.

Strengths of our study include the large sample size, the population-based setting and the hypothesis-free approach to be able to fine map cognition to grey matter. Some limitations deserve to be acknowledged. First, because of the cross-sectional design, no conclusions can be drawn regarding the directionality of causality of the associations. Second, our cognitive test battery, limited in time because of the population-based nature, yielded a less extensive cognitive assessment compared to other studies in smaller samples. Third, the current study mainly consists of Caucasians, therefore the generalizability to other ethnicities is limited. Finally, it is well-known that several cognitive processes are lateralized to a functionally dominant hemisphere and therefore it would have been interesting to investigate handedness as effect modifier. Unfortunately, in our study we did not have a reliable measure of handedness. In conclusion, in this population-based study of nearly 4000 subjects we mapped cognitive ability to grey matter by using hypothesis-free approaches of voxel-based morphometry and shape analysis. We made the maps of association publicly available (https://neurovault.org/) for any researcher to explore the results or to contrast their findings against. Our results propose that a more fine-grained analysis of brain structure adds to our understanding of cognitive function in normal aging. Future studies are needed to disentangle development and degeneration of the human brain. Additionally, longitudinal assessment of cognitive functioning and grey matter atrophy is needed to study causality.

REFERENCES

1. Bleecker, M.L., Bolla-Wilson, K., Agnew, J., Meyers, D.A. (1988) Age-related sex differences in verbal memory. J Clin Psychol, 44:403-11.

2. Bola, M., Sabel, B.A. (2015) Dynamic reorganization of brain functional networks during cognition. NeuroImage, 114:398-413.

3. Bos, M.J., Koudstaal, P.J., Hofman, A., Ikram, M.A. (2014) Modifiable etiological factors and the burden of stroke from the Rotterdam study: a population-based cohort study. PLoS Med, 11:e1001634.

4. Bressler, S.L., Menon, V. (2010) Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn Sci, 14:277-90.

5. Bugg, J.M., Jacoby, L.L., Toth, J.P. (2008) Multiple levels of control in the Stroop task. Mem Cognit, 36:1484-94.

6. Burgaleta, M., MacDonald, P.A., Martinez, K., Roman, F.J., Alvarez-Linera, J., Ramos Gonzalez, A., Karama, S., Colom, R. (2014) Subcortical regional morphology correlates with fluid and spatial intelligence. Hum Brain Mapp, 35:1957-68.

7. Burgess, N., Maguire, E.A., O'Keefe, J. (2002) The human hippocampus and spatial and episodic memory. Neuron, 35:625-41.

8. Cauda, F., D'Agata, F., Sacco, K., Duca, S., Geminiani, G., Vercelli, A. (2011) Functional connectivity of the insula in the resting brain. NeuroImage, 55:8-23.

9. Chang, L.J., Yarkoni, T., Khaw, M.W., Sanfey, A.G. (2013) Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. Cereb Cortex, 23:739-49.

10. Churchill, G.A., Doerge, R.W. (1994) Empirical threshold values for quantitative trait mapping. Genetics, 138:963-71.

11. Davidson, D.J., Zacks, R.T., Williams, C.C. (2003) Stroop interference, practice, and aging. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 10:85-98.

12. Davies, G., Armstrong, N., Bis, J.C., Bressler, J., Chouraki, V., Giddaluru, S., Hofer, E., Ibrahim-Verbaas, C.A., Kirin, M., Lahti, J., van der Lee, S.J., Le Hellard, S., Liu, T., Marioni, R.E., Oldmeadow, C., Postmus, I., Smith, A.V., Smith, J.A., Thalamuthu, A., Thomson, R., Vitart, V., Wang, J., Yu, L., Zgaga, L., Zhao, W., Boxall, R., Harris, S.E., Hill, W.D., Liewald, D.C., Luciano, M., Adams, H., Ames, D., Amin, N., Amouyel, P., Assareh, A.A., Au, R., Becker, J.T., Beiser, A., Berr, C., Bertram, L., Boerwinkle, E., Buckley, B.M., Campbell, H., Corley, J., De Jager, P.L., Dufouil, C., Eriksson, J.G., Espeseth, T., Faul, J.D., Ford, I., Generation, S., Gottesman, R.F., Griswold, M.E., Gudnason, V., Harris, T.B., Heiss, G., Hofman, A., Holliday, E.G., Huffman, J., Kardia, S.L., Kochan, N., Knopman, D.S., Kwok, J.B., Lambert, J.C., Lee, T., Li, G., Li, S.C., Loitfelder, M., Lopez, O.L., Lundervold, A.J., Lundqvist, A., Mather, K.A., Mirza, S.S., Nyberg, L., Oostra, B.A., Palotie, A., Papenberg, G., Pattie, A., Petrovic, K., Polasek, O., Psaty, B.M., Redmond, P., Reppermund, S., Rotter, J.I., Schmidt, H., Schuur, M., Schofield, P.W., Scott, R.J., Steen, V.M., Stott, D.J., van Swieten, J.C., Taylor, K.D., Trollor, J., Trompet, S., Uitterlinden, A.G., Weinstein, G., Widen, E., Windham, B.G., Jukema, J.W., Wright, A.F., Wright, M.J., Yang, Q., Amieva, H., Attia, J.R., Bennett, D.A., Brodaty, H., de Craen, A.J., Hayward, C., Ikram, M.A., Lindenberger, U., Nilsson, L.G., Porteous, D.J., Raikkonen, K., Reinvang, I., Rudan, I., Sachdev, P.S., Schmidt, R., Schofield, P.R., Srikanth, V., Starr, J.M., Turner, S.T., Weir, D.R.,

Wilson, J.F., van Duijn, C., Launer, L., Fitzpatrick, A.L., Seshadri, S., Mosley, T.H., Jr., Deary, I.J. (2015) Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). Mol Psychiatry, 20:183-92.

13. de Bruijn, R.F., Bos, M.J., Portegies, M.L., Hofman, A., Franco, O.H., Koudstaal, P.J., Ikram, M.A. (2015) The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. BMC Med, 13:132.

14. de Flores, R., La Joie, R., Chetelat, G. (2015) Structural imaging of hippocampal subfields in healthy aging and Alzheimer's disease. Neuroscience, 309:29-50.

15. Deary, I.J., Penke, L., Johnson, W. (2010) The neuroscience of human intelligence differences. Nat Rev Neurosci, 11:201-11.

16. Deen, B., Pitskel, N.B., Pelphrey, K.A. (2011) Three systems of insular functional connectivity identified with cluster analysis. Cereb Cortex, 21:1498-506.

17. Dupont, S., Bouilleret, V., Hasboun, D., Semah, F., Baulac, M. (2003) Functional anatomy of the insula: new insights from imaging. Surgical and radiologic anatomy : SRA, 25:113-9.

18. Eggermont, L.H., de Boer, K., Muller, M., Jaschke, A.C., Kamp, O., Scherder, E.J. (2012) Cardiac disease and cognitive impairment: a systematic review. Heart, 98:1334-40.

19. Elderkin-Thompson, V., Ballmaier, M., Hellemann, G., Pham, D., Kumar, A. (2008) Executive function and MRI prefrontal volumes among healthy older adults. Neuropsychology, 22:626-37.

20. Fama, R., Sullivan, E.V. (2015) Thalamic structures and associated cognitive functions: Relations with age and aging. Neurosci Biobehav Rev, 54:29-37.

21. Fischl, B., Salat, D.H., van der Kouwe, A.J., Makris, N., Segonne, F., Quinn, B.T., Dale, A.M. (2004) Sequence-independent segmentation of magnetic resonance images. Neuroimage, 23 Suppl 1:S69-84.

22. Fleischman, D.A., Leurgans, S., Arfanakis, K., Arvanitakis, Z., Barnes, L.L., Boyle, P.A., Han, S.D., Bennett, D.A. (2014) Gray-matter macrostructure in cognitively healthy older persons: associations with age and cognition. Brain structure & function, 219:2029-49.

23. Frisk, V., Milner, B. (1990) The role of the left hippocampal region in the acquisition and retention of story content. Neuropsychologia, 28:349-59.

24. Gale, C.R., Deary, I.J., Boyle, S.H., Barefoot, J., Mortensen, L.H., Batty, G.D. (2008) Cognitive ability in early adulthood and risk of 5 specific psychiatric disorders in middle age: the Vietnam experience study. Arch Gen Psychiatry, 65:1410-8.

25. Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S. (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage, 14:21-36.

26. Gutman, B. (2015) Information Processing in Medical Imaging. 205–218 Springer International Publishing.

27. Gutman, B.A., Hua, X., Rajagopalan, P., Chou, Y.Y., Wang, Y., Yanovsky, I., Toga, A.W., Jack, C.R., Jr., Weiner, M.W., Thompson, P.M., Alzheimer's Disease Neuroimaging, I. (2013) Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features. Neuroimage, 70:386-401.

28. Haber, S.N. (2016) Corticostriatal circuitry. Dialogues Clin Neurosci, 18:7-21.

29. Halperin, J.M., Trampush, J.W., Miller, C.J., Marks, D.J., Newcorn, J.H. (2008) Neuropsychological outcome in adolescents/young adults with childhood ADHD: profiles of persisters, remitters and controls. Journal of child psychology and psychiatry, and allied disciplines, 49:958-66.

30. Haworth, C.M., Wright, M.J., Luciano, M., Martin, N.G., de Geus, E.J., van Beijsterveldt, C.E., Bartels, M., Posthuma, D., Boomsma, D.I., Davis, O.S., Kovas, Y., Corley, R.P., Defries, J.C., Hewitt, J.K., Olson, R.K., Rhea, S.A., Wadsworth, S.J., Iacono, W.G., McGue, M., Thompson, L.A., Hart, S.A., Petrill, S.A., Lubinski, D., Plomin, R. (2010) The heritability of general cognitive ability increases linearly from childhood to young adulthood. Mol Psychiatry, 15:1112-20.

31. Hoogendam, Y.Y., Hofman, A., van der Geest, J.N., van der Lugt, A., Ikram, M.A. (2014) Patterns of cognitive function in aging: the Rotterdam Study. Eur J Epidemiol, 29:133-40.

32. Houx, P.J., Jolles, J., Vreeling, F.W. (1993) Stroop interference: aging effects assessed with the Stroop Color-Word Test. Exp Aging Res, 19:209-24.

33. Ikram, M.A., Brusselle, G.G.O., Murad, S.D., van Duijn, C.M., Franco, O.H., Goedegebure, A., Klaver, C.C.W., Nijsten, T.E.C., Peeters, R.P., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vernooij, M.W., Hofman, A. (2017) The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol, 32:807-850.

34. Ikram, M.A., van der Lugt, A., Niessen, W.J., Koudstaal, P.J., Krestin, G.P., Hofman, A., Bos, D., Vernooij, M.W. (2015) The Rotterdam Scan Study: design update 2016 and main findings. Eur J Epidemiol.

35. Ito, R., Lee, A.C. (2016) The role of the hippocampus in approach-avoidance conflict decision-making: Evidence from rodent and human studies. Behav Brain Res, 313:345-57.

36. Kramer, A.F., Bherer, L., Colcombe, S.J., Dong, W., Greenough, W.T. (2004) Environmental influences on cognitive and brain plasticity during aging. J Gerontol A Biol Sci Med Sci, 59:M940-57.

37. La Joie, R., Perrotin, A., de La Sayette, V., Egret, S., Doeuvre, L., Belliard, S., Eustache, F., Desgranges, B., Chetelat, G. (2013) Hippocampal subfield volumetry in mild cognitive impairment, Alzheimer's disease and semantic dementia. NeuroImage. Clinical, 3:155-62.

38. Lawrence, A.D., Sahakian, B.J., Robbins, T.W. (1998) Cognitive functions and corticostriatal circuits: insights from Huntington's disease. Trends Cogn Sci, 2:379-88.

39. Lin, A., Ching, C.R.K., Vajdi, A., Sun, D., Jonas, R.K., Jalbrzikowski, M., Kushan-Wells, L., Pacheco Hansen, L., Krikorian, E., Gutman, B., Dokoru, D., Helleman, G., Thompson, P.M., Bearden, C.E. (2017) Mapping 22q11.2 Gene Dosage Effects on Brain Morphometry. J Neurosci, 37:6183-6199.

40. Lindberg, O., Walterfang, M., Looi, J.C., Malykhin, N., Ostberg, P., Zandbelt, B., Styner, M., Paniagua, B., Velakoulis, D., Orndahl, E., Wahlund, L.O. (2012) Hippocampal shape analysis in Alzheimer's disease and frontotemporal lobar degeneration subtypes. J Alzheimers Dis, 30:355-65.

41. McGaugh, J.L., Cahill, L., Roozendaal, B. (1996) Involvement of the amygdala in memory storage: interaction with other brain systems. Proc Natl Acad Sci U S A, 93:13508-14.

42. Nelson, S.M., Dosenbach, N.U., Cohen, A.L., Wheeler, M.E., Schlaggar, B.L., Petersen, S.E. (2010) Role of the anterior insula in task-level control and focal attention. Brain structure & function, 214:669-80.

43. Newman, L.M., Trivedi, M.A., Bendlin, B.B., Ries, M.L., Johnson, S.C. (2007) The Relationship Between Gray Matter Morphometry and Neuropsychological Performance in a Large Sample of Cognitively Healthy Adults. Brain

Imaging Behav, 1:3-10.

44. Park, H.J., Friston, K. (2013) Structural and functional brain networks: from connections to cognition. Science, 342:1238411.

45. Phelps, E.A. (2004) Human emotion and memory: interactions of the amygdala and hippocampal complex. Curr Opin Neurobiol, 14:198-202.

46. Pirmoradi, M., Jemel, B., Gallagher, A., Tremblay, J., D'Hondt, F., Nguyen, D.K., Beland, R., Lassonde, M. (2016) Verbal memory and verbal fluency tasks used for language localization and lateralization during magnetoencephalography. Epilepsy Res, 119:1-9.

47. Prins, N.D., van Dijk, E.J., den Heijer, T., Vermeer, S.E., Jolles, J., Koudstaal, P.J., Hofman, A., Breteler, M.M. (2005) Cerebral small-vessel disease and decline in information processing speed, executive function and memory. Brain, 128:2034-41.

48. Rorden, C., Karnath, H.O. (2004) Using human brain lesions to infer function: a relic from a past era in the fMRI age? Nat Rev Neurosci, 5:813-9.

49. Roshchupkin, G.V., Adams, H.H., van der Lee, S.J., Vernooij, M.W., van Duijn, C.M., Uitterlinden, A.G., van der Lugt, A., Hofman, A., Niessen, W.J., Ikram, M.A. (2016a) Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology. Neurobiol Aging, 48:204-211.

50. Roshchupkin, G.V., Gutman, B.A., Vernooij, M.W., Jahanshad, N., Martin, N.G., Hofman, A., McMahon, K.L., Van Der Lee, S.J., Van Duijn, C.M., De Zubicaray, G.I. (2016b) Heritability of the shape of subcortical brain structures in the general population. Nature Communications, 7.

51. Roshchupkin, G.V., Gutman, B.A., Vernooij, M.W., Jahanshad, N., Martin, N.G., Hofman, A., McMahon, K.L., van der Lee, S.J., van Duijn, C.M., de Zubicaray, G.I., Uitterlinden, A.G., Wright, M.J., Niessen, W.J., Thompson, P.M., Ikram, M.A., Adams, H.H. (2016c) Heritability of the shape of subcortical brain structures in the general population. Nat Commun, 7:13738.

52. Rubin, R.D., Schwarb, H., Lucas, H.D., Dulas, M.R., Cohen, N.J. (2017) Dynamic Hippocampal and Prefrontal Contributions to Memory Processes and Representations Blur the Boundaries of Traditional Cognitive Domains. Brain Sci, 7.

53. Ruscheweyh, R., Deppe, M., Lohmann, H., Wersching, H., Korsukewitz, C., Duning, T., Bluhm, S., Stehling, C., Keller, S.S., Knecht, S. (2013) Executive performance is related to regional gray matter volume in healthy older individuals. Hum Brain Mapp, 34:3333-46.

54. Salat, D.H., Buckner, R.L., Snyder, A.Z., Greve, D.N., Desikan, R.S., Busa, E., Morris, J.C., Dale, A.M., Fischl, B. (2004) Thinning of the cerebral cortex in aging. Cereb Cortex, 14:721-30.

55. Schmahmann, J.D., Pandya, D.N. (2008) Disconnection syndromes of basal ganglia, thalamus, and cerebrocerebellar systems. Cortex; a journal devoted to the study of the nervous system and behavior, 44:1037-66.

56. Seger, C.A. (2013) The visual corticostriatal loop through the tail of the caudate: circuitry and function. Front Syst Neurosci, 7:104.

57. Shao, Z., Janse, E., Visser, K., Meyer, A.S. (2014) What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. Front Psychol, 5:772.

58. Smith, M.L., Milner, B. (1981) The role of the right hippocampus in the recall of spatial location. Neuropsycho-

logia, 19:781-93.

59. Squarzoni, P., Tamashiro-Duran, J., Souza Duran, F.L., Santos, L.C., Vallada, H.P., Menezes, P.R., Scazufca, M., Filho, G.B., Alves, T.C. (2012) Relationship between regional brain volumes and cognitive performance in the healthy aging: an MRI study using voxel-based morphometry. J Alzheimers Dis, 31:45-58.

60. Squire, L.R., Zola-Morgan, S. (1988) Memory: brain systems and behavior. Trends Neurosci, 11:170-5.

61. Tisserand, D.J., van Boxtel, M.P., Pruessner, J.C., Hofman, P., Evans, A.C., Jolles, J. (2004) A voxel-based morphometric study to determine individual differences in gray matter density associated with age and cognitive change over time. Cereb Cortex, 14:966-73.

62. Van Der Werf, Y.D., Tisserand, D.J., Visser, P.J., Hofman, P.A., Vuurman, E., Uylings, H.B., Jolles, J. (2001) Thalamic volume predicts performance on tests of cognitive speed and decreases in healthy aging. A magnetic resonance imaging-based volumetric analysis. Brain research. Cognitive brain research, 11:377-85.

63. Van Petten, C. (2004) Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. Neuropsychologia, 42:1394-413.

64. Vrooman, H.A., Cocosco, C.A., van der Lijn, F., Stokking, R., Ikram, M.A., Vernooij, M.W., Breteler, M.M., Niessen, W.J. (2007) Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. Neuroimage, 37:71-81.

65. Wang, L., Mamah, D., Harms, M.P., Karnik, M., Price, J.L., Gado, M.H., Thompson, P.A., Barch, D.M., Miller, M.I., Csernansky, J.G. (2008) Progressive deformation of deep brain nuclei and hippocampal-amygdala formation in schizophrenia. Biol Psychiatry, 64:1060-8.

SUPPLEMENTARY MATERIAL

High-dimensional mapping of cognition to the brain using voxel-based morphometry and subcortical shape analysis.

CONTENT:

- **Supplementary Figure 1:** Correlation between cognitive tests stratified by sex.
- **Supplementary Figure 2.** FWE-significant vertices of subcortical brain structures in relation to cognitive tests
- **Supplementary Table 1:** Association between cognitive tests and shape measures in the left and right hemisphere.

SUPPLEMENTARY FIGURE 1: Correlation between cognitive tests stratified by sex.
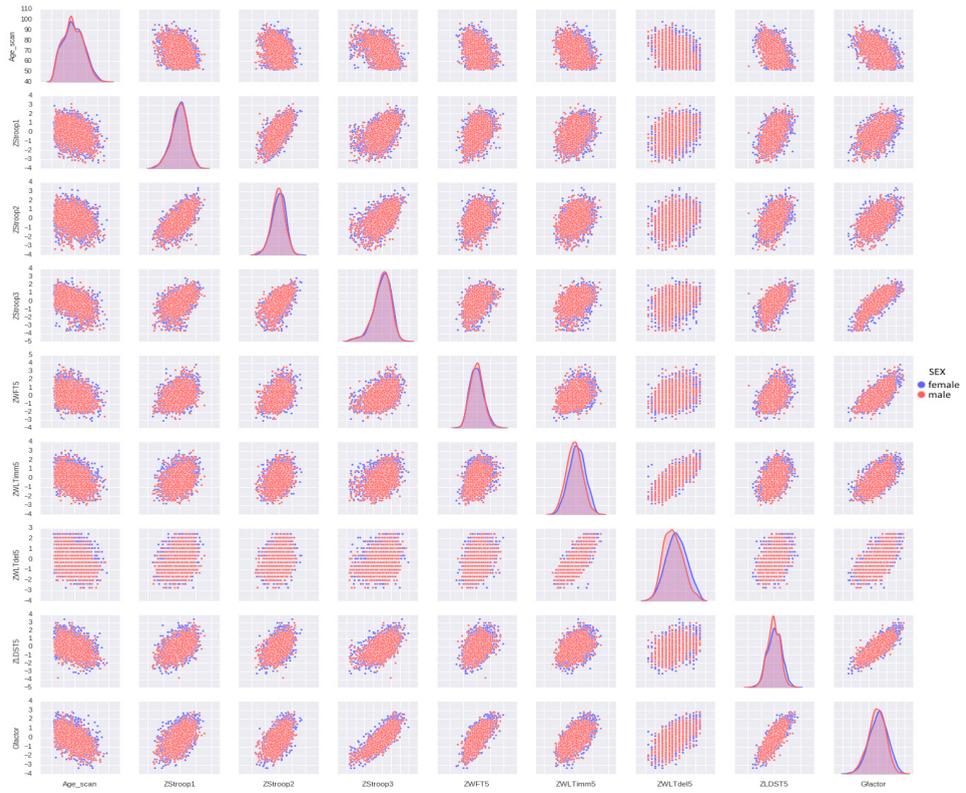
SUPPLEMENTARY FIGURE 2A: FWE-significant vertices of subcortical brain structures in relation to Word learning test, delayed recall.
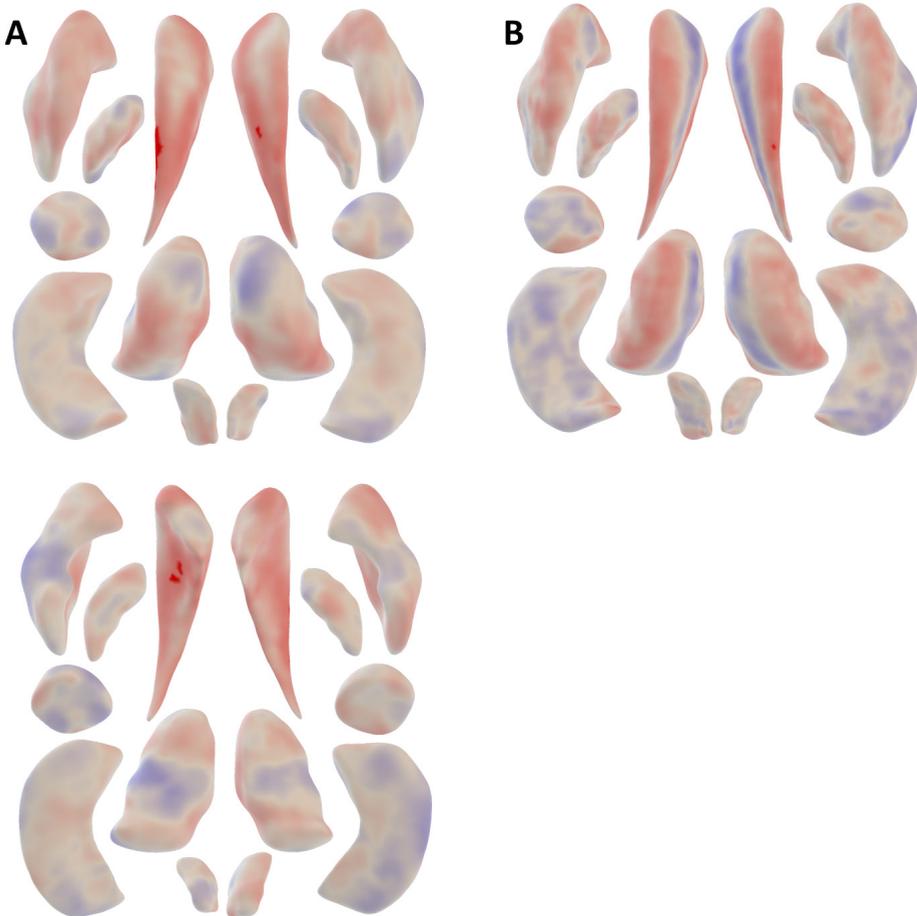
Maps show the associations of seven bilateral subcortical structures for the shape measures of Jacobian determinant **(Panel A)** and radial distance **(Panel B)**, anterior (top row) and posterior (bottom row) view. All associations are adjusted for age, sex, and education. Color map represents the t-statistics and shows the direction of association, with red and blue indicating negative and positive associations respectively.

SUPPLEMENTARY FIGURE 2B: FWE-significant vertices of subcortical brain structures in relation to Word learning test, immediate recall.
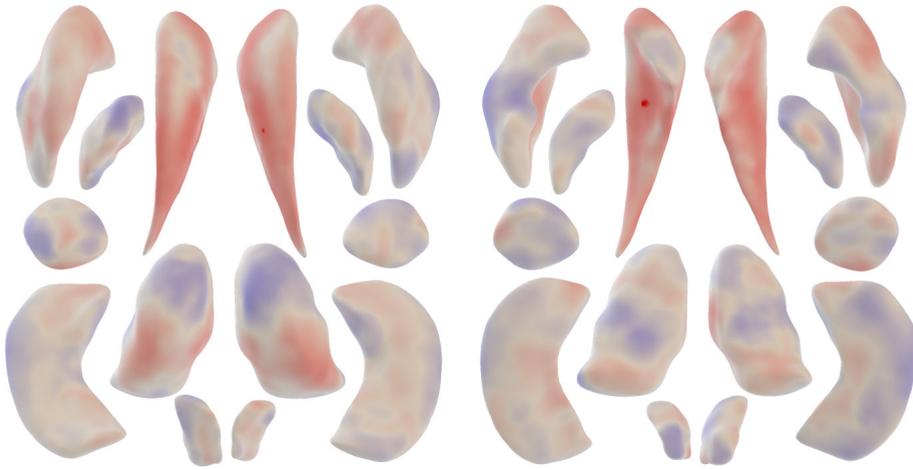
Maps show the associations of seven bilateral subcortical structures for the shape measures of Jacobian determinant, anterior (left) and posterior (right) view. All associations are adjusted for age, sex, and education. Color map represents the t-statistics and shows the direction of association, with red and blue indicating negative and positive associations respectively.

SUPPLEMENTARY FIGURE 2C. FWE-significant vertices of subcortical brain structures in relation to Word fluency test.
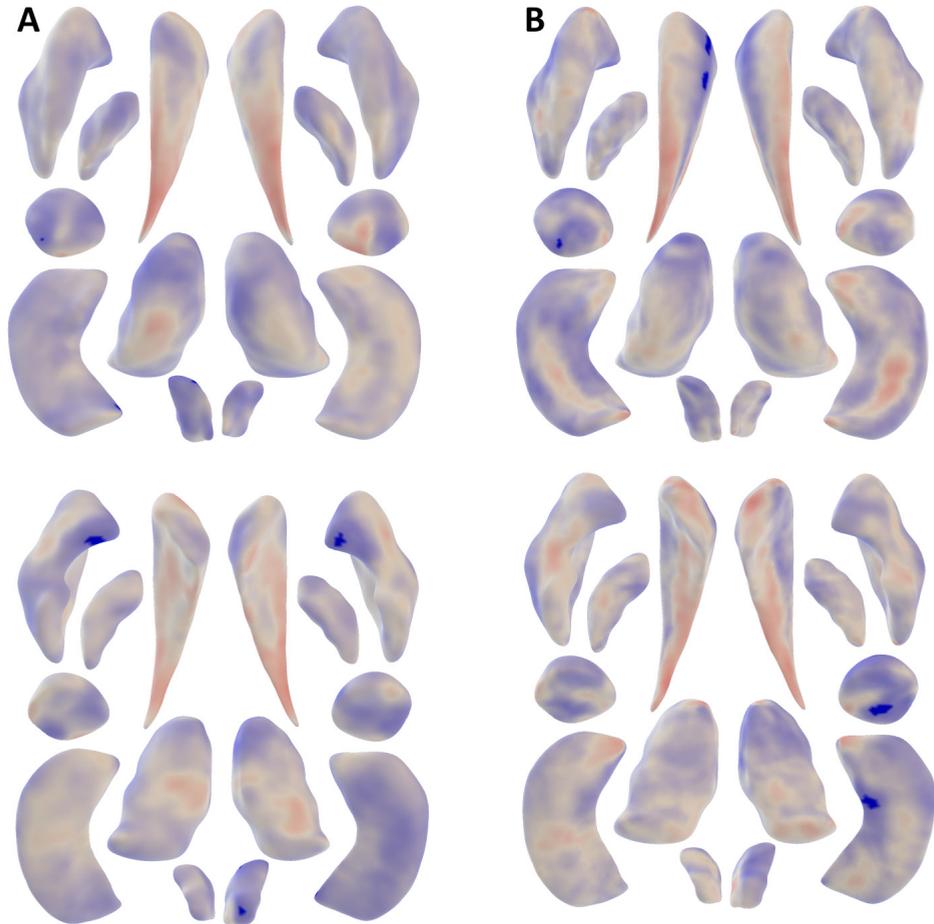
Maps show the associations of seven bilateral subcortical structures for the shape measures of Jacobian determinant **(Panel A)** and radial distance **(Panel B)**, anterior (top row) and posterior (bottom row) view. All associations are adjusted for age, sex, and education. Color map represents the t-statistics and shows the direction of association, with red and blue indicating negative and positive associations respectively.

SUPPLEMENTARY FIGURE 2D: FWE-significant vertices of subcortical brain structures in relation to Stroop color reading test.
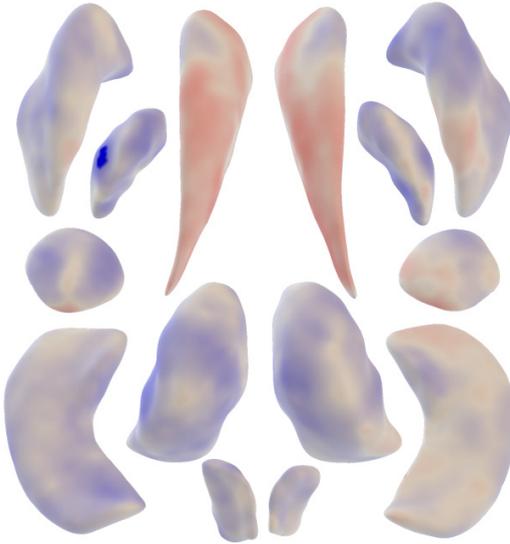
Maps show the associations of seven bilateral subcortical structures for the shape measures of Jacobian determinant, anterior view. All associations are adjusted for age, sex, and education. Color map represents the t-statistics and shows the direction of association, with red and blue indicating negative and positive associations respectively.

SUPPLEMENTARY FIGURE 2E. FWE-significant vertices of subcortical brain structures in relation to Stroop color naming test.
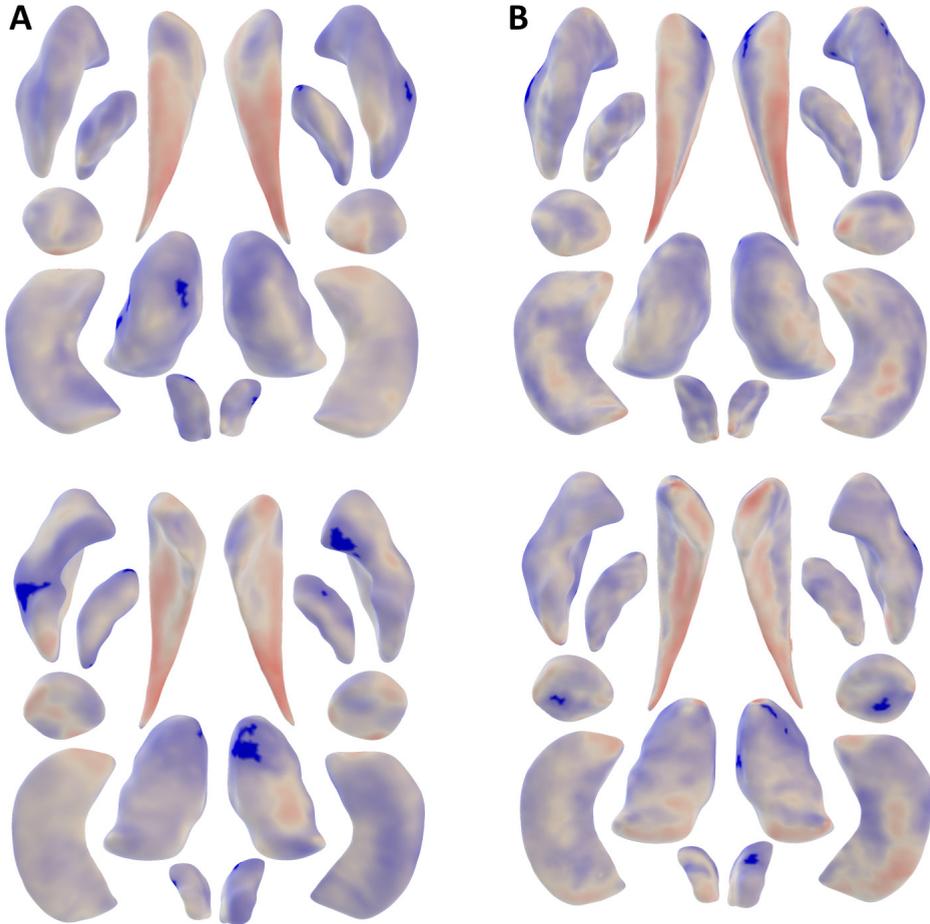

Maps show the associations of seven bilateral subcortical structures for the shape measures of Jacobian determinant **(Panel A)** and radial distance **(Panel B)**, anterior (top row) and posterior (bottom row) view. All associations are adjusted for age, sex, and education. Color map represents the t-statistics and shows the direction of association, with red and blue indicating negative and positive associations respectively.

SUPPLEMENTARY FIGURE 2F: FWE-significant vertices of subcortical brain structures in relation to Stroop interference test.
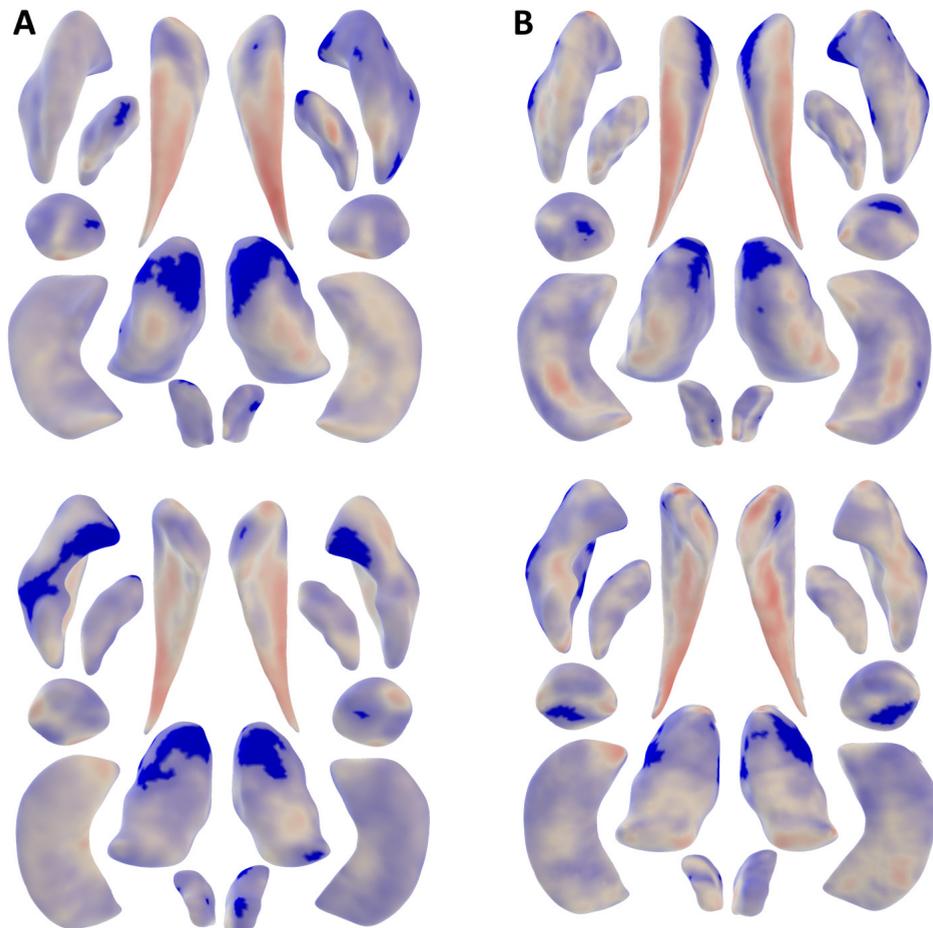
Maps show the associations of seven bilateral subcortical structures for the shape measures of Jacobian determinant **(Panel A)** and radial distance **(Panel B)**, anterior (top row) and posterior (bottom row) view. All associations are adjusted for age, sex, and education. Color map represents the t-statistics and shows the direction of association, with red and blue indicating negative and positive associations respectively.

**SUPPLEMENTARY FIGURE 2G:** FWE-significant vertices of subcortical brain structures in relation to Letter-digit substitution test.
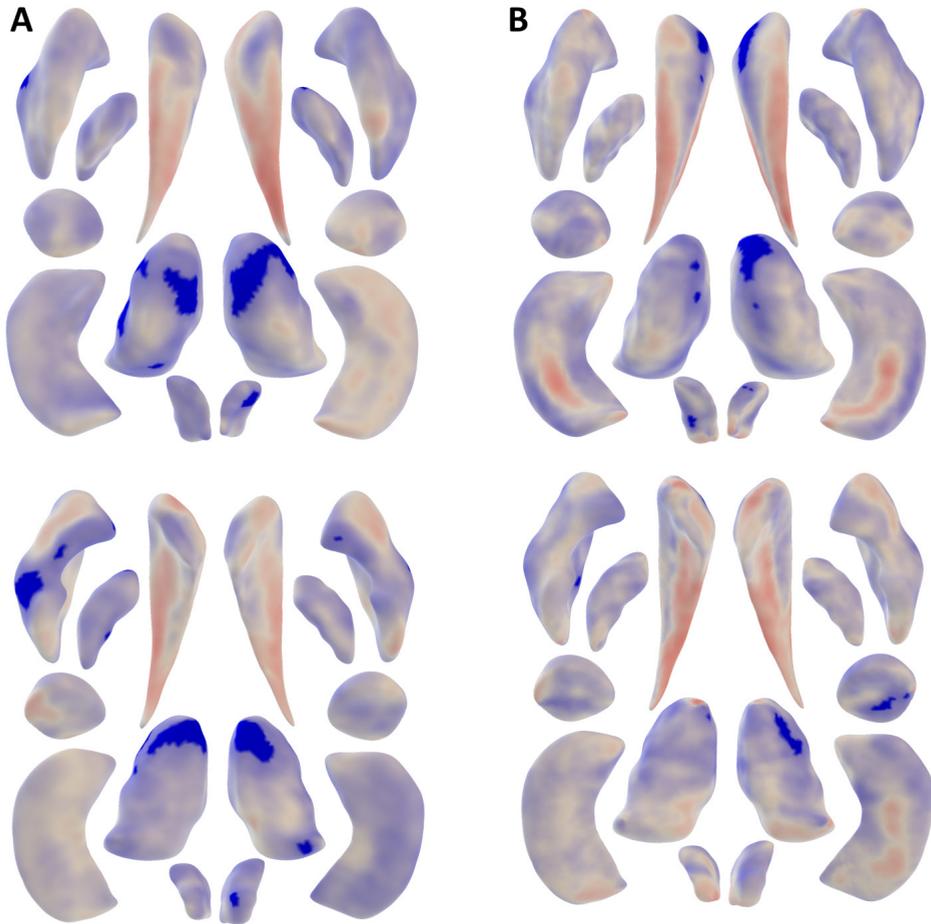

Maps show the associations of seven bilateral subcortical structures for the shape measures of Jacobian determinant **(Panel A)** and radial distance **(Panel B)**, anterior (top row) and posterior (bottom row) view. All associations are adjusted for age, sex, and education. Color map represents the t-statistics and shows the direction of association, with red and blue indicating negative and positive associations respectively.

Supplementary Table 1A: Association between cognitive tests and shape measures in the left hemisphere.

Cognitive Test	Measure	Subcortical Structure	min p-value	# negative vertices	# positive vertices
G-factor	Jacobian determinant	Thalamus	4.36E-12	0	281
	Jacobian determinant	Caudate nucleus	2.11E-08	68	0
	Jacobian determinant	Putamen	2.71E-08	0	122
	Jacobian determinant	Nucleus Accumbens	7.87E-07	0	7
	Jacobian determinant	Pallidum	1.43E-06	0	4
	Jacobian determinant	Amygdala	1.84E-06	0	2
	radial distance	Caudate nucleus	2.89E-16	20	113
	radial distance	Amygdala	1.09E-10	0	37
	radial distance	Thalamus	1.17E-10	0	70
	radial distance	Putamen	1.08E-08	0	41
WLT delayed recall	Jacobian determinant	Caudate nucleus	1.03E-06	13	0
	radial distance	Caudate nucleus	6.96E-07	1	0
WLT immediate recall	Jacobian determinant	Caudate nucleus	7.83E-07	5	0
WFT	Jacobian determinant	Putamen	3.67E-07	0	12
Stroop color naming	Jacobian determinant	Putamen	2.22E-08	0	46
	Jacobian determinant	Nucleus Accumbens	1.26E-07	0	11
	Jacobian determinant	Pallidum	2.04E-07	0	17
	Jacobian determinant	Thalamus	1.13E-06	0	3
	radial distance	Putamen	1.66E-07	0	5
	radial distance	Caudate nucleus	1.97E-07	0	21
	radial distance	Amygdala	2.46E-07	0	6
	radial distance	Thalamus	1.02E-06	0	1
Stroop interference	Jacobian determinant	Thalamus	5.84E-11	0	396
	Jacobian determinant	Putamen	2.28E-09	0	298
	Jacobian determinant	Pallidum	8.64E-09	0	33
	Jacobian determinant	Nucleus Accumbens	9.26E-08	0	26
	Jacobian determinant	Caudate nucleus	1.91E-06	0	3
	radial distance	Caudate nucleus	2.08E-16	0	102
	radial distance	Thalamus	1.34E-12	0	209
	radial distance	Amygdala	7.51E-12	0	60
	radial distance	Putamen	1.26E-09	0	215
	radial distance	Nucleus Accumbens	1.33E-07	0	7
	radial distance	Hippocampus	6.79E-07	0	1
	radial distance	Pallidum	9.47E-07	0	3
LDST	Jacobian determinant	Thalamus	4.50E-12	0	274
	Jacobian determinant	Nucleus Accumbens	2.54E-09	0	30
	Jacobian determinant	Putamen	6.76E-09	0	72
	Jacobian determinant	Pallidum	6.62E-07	0	12
	radial distance	Caudate nucleus	9.39E-13	0	70
	radial distance	Thalamus	7.74E-12	0	90
	radial distance	Putamen	1.01E-07	0	9
	radial distance	Nucleus Accumbens	4.60E-07	0	3
Total # vertices				107	2712

SUPPLEMENTARY TABLE 1B: Association between cognitive tests and shape measures in the right hemisphere.

Cognitive Test	Measure	Subcortical Structure	min p-value	# negative vertices	# positive vertices
G-factor	Jacobian determinant	Thalamus	1.93E-09	0	159
	Jacobian determinant	Putamen	2.03E-07	0	41
	Jacobian determinant	Nucleus accumbens	3.04E-07	0	56
	Jacobian determinant	Amygdala	1.17E-06	0	6
	radial distance	Caudate nucleus	9.68E-13	0	78
	radial distance	Amygdala	3.74E-11	0	68
	radial distance	Nucleus accumbens	1.03E-09	0	24
	radial distance	Hippocampus	2.17E-09	0	22
	radial distance	Thalamus	6.19E-09	0	102
WLT delayed recall	Jacobian determinant	Caudate nucleus	1.57E-07	37	0
WFT	Jacobian determinant	Nucleus accumbens	8.97E-07	0	11
	Jacobian determinant	Hippocampus	1.10E-06	0	5
	Jacobian determinant	Putamen	1.13E-06	0	8
	Jacobian determinant	Amygdala	2.26E-06	0	1
	radial distance	Amygdala	7.61E-09	0	27
	radial distance	Caudate nucleus	6.66E-08	0	23
	radial distance	Hippocampus	1.36E-07	0	8
Stroop color reading	Jacobian determinant	Pallidum	6.49E-08	0	19
Stroop color naming	Jacobian determinant	Nucleus accumbens	1.05E-07	0	19
	Jacobian determinant	Thalamus	1.09E-07	0	75
	Jacobian determinant	Putamen	2.27E-07	0	32
	Jacobian determinant	Pallidum	1.65E-06	0	2
	radial distance	Thalamus	1.16E-08	0	17
	radial distance	Nucleus accumbens	1.35E-08	0	11
	radial distance	Amygdala	2.24E-08	0	13
	radial distance	Putamen	4.09E-08	0	45
	radial distance	Caudate nucleus	2.74E-07	1	4
Stroop interference	Jacobian determinant	Thalamus	6.11E-12	0	324
	Jacobian determinant	Putamen	6.07E-09	0	95
	Jacobian determinant	Nucleus accumbens	5.98E-08	0	71
	Jacobian determinant	Pallidum	9.13E-08	0	22
	Jacobian determinant	Amygdala	6.34E-07	0	19
	Jacobian determinant	Caudate nucleus	8.36E-07	0	3
	radial distance	Amygdala	2.25E-12	0	75
	radial distance	Thalamus	2.38E-11	0	221
	radial distance	Caudate nucleus	9.21E-10	0	96
	radial distance	Putamen	2.17E-09	0	60
	radial distance	Nucleus accumbens	8.71E-07	0	1
LDST	Jacobian determinant	Thalamus	1.56E-11	0	249
	Jacobian determinant	Putamen	5.45E-08	0	26
	Jacobian determinant	Nucleus accumbens	1.34E-07	0	18
	radial distance	Caudate nucleus	3.42E-10	0	27
	radial distance	Nucleus accumbens	5.01E-09	0	8
	radial distance	Thalamus	1.08E-08	0	51
	radial distance	Amygdala	1.96E-08	0	18
Total # vertices				38	2260
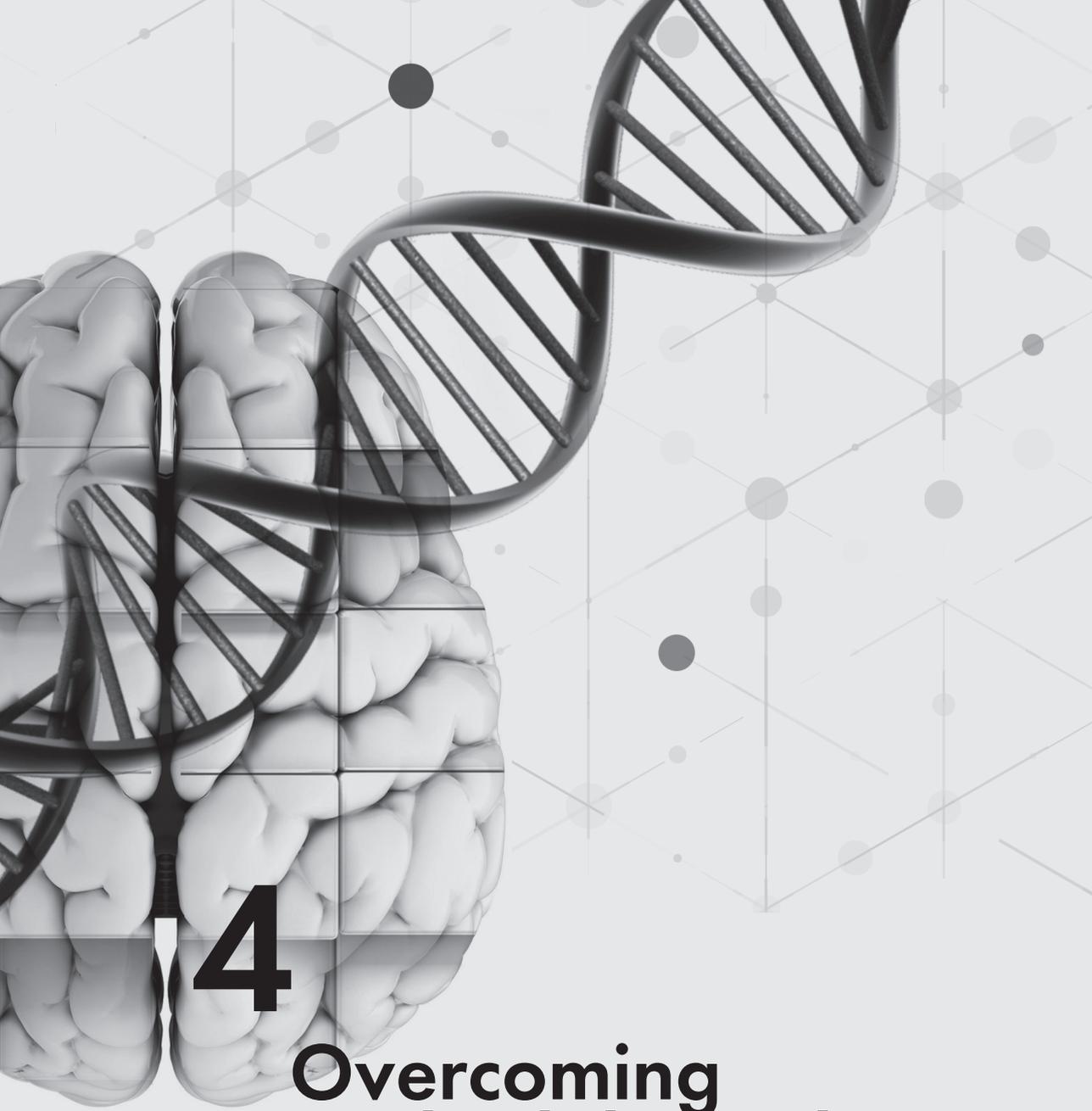
4

Overcoming
methodological
issues

CHAPTER 4.1

HASE: Framework for
efficient high-dimensional
association analyses

ABSTRACT

High-throughput technology can now provide rich information on a person's biological makeup and environmental surroundings. Important discoveries have been made by relating these data to various health outcomes in fields such as genomics, proteomics, and medical imaging. However, cross-investigations between several high-throughput technologies remain impractical due to demanding computational requirements (hundreds of years of computing resources) and unsuitability for collaborative settings (terabytes of data to share).Here we introduce the HASE framework that overcomes both of these issues. Our approach dramatically reduces computational time from years to only hours and also requires several gigabytes to be exchanged between collaborators. We implemented a novel meta-analytical method that yields identical power as pooled analyses without the need of sharing individual participant data. The efficiency of the framework is illustrated by associating 9 million genetic variants with 1.5 million brain imaging voxels in three cohorts (total N=4,034) followed by meta-analysis, on a standard computational infrastructure. These experiments indicate that HASE facilitates high-dimensional association studies enabling large multicenter association studies for future discoveries.

INTRODUCTION

Technological innovations have enabled the large-scale acquisition of biological information from human subjects. The emergence of these big datasets has resulted in various 'omics' fields. Systematic and large-scale investigations of DNA sequence variations (genomics)[1], gene expression (transcriptomics)[2], proteins (proteomics)[3]including disease risk prediction, prevention and targeted medication. One of the major challenges that researchers face on the path between the initial identification of an association and precision treatment of patients is the comprehension of the biological mechanisms that underlie these associations. Currently, the focus to solve these questions lies on the integrative analysis of system-wide data on global genome variation, gene expression, transcription factor binding, epigenetic profiles and chromatin conformation. The generation of this data mainly relies on next-generation sequencing. However, due to multiple recent developments, mass spectrometry-based proteomics now offers additional, by the GWAS field so far hardly recognized possibilities for the identification of functional genome variants and, in particular, for the identification and characterization of (differentially, small molecule metabolites (metabolomics)[4], and medical images (radiomics)[5], among other data, lie at the basis of many recent biological insights. These analyses are typically unidimensional, i.e. studying only a single disease or trait of interest.

Although this approach has proven its scientific merit through many discoveries, jointly investigating multiple big datasets would allow for their full exploitation, as is increasingly recognized throughout the 'omics' world[5–8]. However, the high-dimensional nature of these analyses makes them challenging and often unfeasible in current research settings. Specifically, the computational requirements for analyzing high-dimensional data are far beyond the infrastructural capabilities for single sites. Furthermore, it is incompatible with the typical collaborative approach of distributed multi-site analyses followed by meta-analysis, since the amount of generated data at every site is too large to transfer.

Some studies have attempted to combine multiple big datasets[5,8–10], but these methods generally rely on reducing the dimensionality or making assumptions to approximate the results, which leads to a loss of information.

Here we present the framework for efficient high-dimensional association analyses (HASE), which is capable of analyzing high-dimensional data at full resolution, yielding exact association statistics (i.e. no approximations), and requiring only standard computational facilities. Additionally, the major computational burden in collaborative efforts is shifted from the individual sites to the meta-analytical level while at the same time reducing the amount of data

needed to be exchanged and preserving participant privacy. HASE thus removes the current computational and logistic barriers for single- and multi-center analyses of big data.

RESULTS

OVERVIEW OF THE METHODS

The methods are described in detail in the Methods. Essentially, HASE implements a high-throughput multiple linear regression algorithm that is computationally efficient when analyzing high-dimensional data of any quantitative trait. Prior to analysis, data are converted to an optimized storage format to reduce reading and writing time. Redundant calculations are removed and the high-dimensional operations are simplified into a set of matrix operations that are computationally inexpensive, thereby reducing overall computational overhead. While deriving summary statistics (e.g., beta coefficients, p-values) for every combination in the high-dimensional analysis would be computationally feasible at individual sites with our approach, it would be too large to share the intermediate results (>200GB per thousand phenotypes) in a multi-center setting. Therefore, extending from a recently proposed method, partial derivatives meta-analysis[17], we additionally developed a method that generates two relatively small datasets (e.g. 5GB for genetics data of 9 million variants and 20MB of thousand phenotypes for 4000 individuals) that are easily transferred and can subsequently be combined to calculate the full set of summary statistics, without making any approximation. This meta-analysis method additionally reduces computational overhead at individual sites by shifting the most expensive calculation to the central site. The total computational burden thus becomes even more efficient relative to conventional methods with additional sites. The HASE software is freely available from our website www.imagene.nl/HASE/.

COMPARISON OF COMPLEXITY AND SPEED

We compared the complexity and speed of HASE with a classical workflow, based on linear regression analyses with PLINK (version 1.9)[11] followed by meta-analysis with METAL[12]; two of the most popular software packages for these tasks.

Table 1 shows that HASE dramatically reduces the complexity for the single site analysis and data transfer stages. For conventional methods, the single site analysis and data transfer have a multiplicative complexity (dependent on the number of phenotypes and determinants), whereas this is only additive for HASE. Our approach requires 2×10^6-fold less time on the single site stage and 3.500-fold less data to transfer for a high-dimensional association study. Additionally, the time for single site analysis does not increase significantly from analyzing a single phenotype to a million phenotypes (**Table 1**). This is due to the fact that speed is determined by the highest number of either the determinants or phenotypes. Therefore, in this case with nine million genetic variants, the complexity of is the primary factor influencing the speed, whereas plays a secondary role.

TABLE 1: Comparison of complexity and speed between the HASE framework and a classical workflow.

Stage	Complexity[c]		Time[a,b] (hours)			
			$n_p=1$		$n_p=10^6$	
	Classical workflow	HASE	Classical workflow	HASE	Classical workflow	HASE
Single site analysis	$O(n_i n_p n_t)$	$\max(O(n_i n_p), O(n_i n_t))$	2.46	0.63	2.46×10^6	0.70
Data transfer	$O(n_p n_t)$	$O(n_i n_p + n_i n_t)$	0.04	0.07[*]	4×10^4	11.6
Meta-Analysis	$O(n_p n_t)$	$O(n_i n_p n_t)$	0.06	0.03	6×10^4	1.7×10^3

[a] Based on a model with three covariates and 9 million genetic variants, for a total of 4034 participants from three sites. For the classical workflow we used the PLINK software for single site analysis and METAL for the meta-analysis.

[b] For single site analysis and meta-analysis the time is given in CPU hours; for the data transfer stage this is in hours using an average network speed of 10Mbps.

[c] Complexity for CPU hours is given in terms of classical computation time complexity; complexity for data transfer is shown in terms of how the size of the to be transferred data depends on the size of the input data.

[*] This time is derived from the transfer of partial derivatives only, because for an association analysis with relatively few phenotypes it is not necessary to transfer encoded data.

- number of individuals in the study; n_p - number of phenotypes of interest; n_t - number of tests (genetic variants); N_s - number of sites in the meta-analysis. In standard analysis $n_i < < n_p$ and $n_i < < \overline{n_t}$

This drastic increase in performance is made possible through the shift of the computationally most expensive regression operation to the meta-analytical stage. For the meta-analytical stage, the HASE complexity is therefore slightly higher. However, it outperforms the classical meta-analysis using METAL (total computation time reduced 35 times), owing to the efficient implementation of our algorithm. Additionally, if HASE is only used for a high-dimensional association study of a single site, i.e. without subsequent meta-analysis with other sites, the computation time would be reduced 1400 times due to the removal of redundant calculations (for details see **Methods**).

APPLICATION TO REAL DATA

We used HASE to perform a high-dimensional association study in 4,034 individuals from the population-based Rotterdam Study. In this proof of principle study, we relate 8,723,231 million imputed genetic variants to 1,534,602 million brain magnetic resonance imaging (MRI) voxel densities (see **Supplementary Note**). The analysis was performed on a small cluster of 100 CPUs and took 17 hours to complete.

To demonstrate the potential of such high-dimensional analyses, we screened all genetic association results for both hippocampi (7,030 voxels) and identified the voxel with the lowest p-value. The most significant association (rs77956314; p = 3 x 10^{-9}) corresponded to a locus on chromosome 12q24 (**Figure 1**), which was recently discovered in a genome-wide association study of hippocampal volume encompassing 30,717 participants[13].
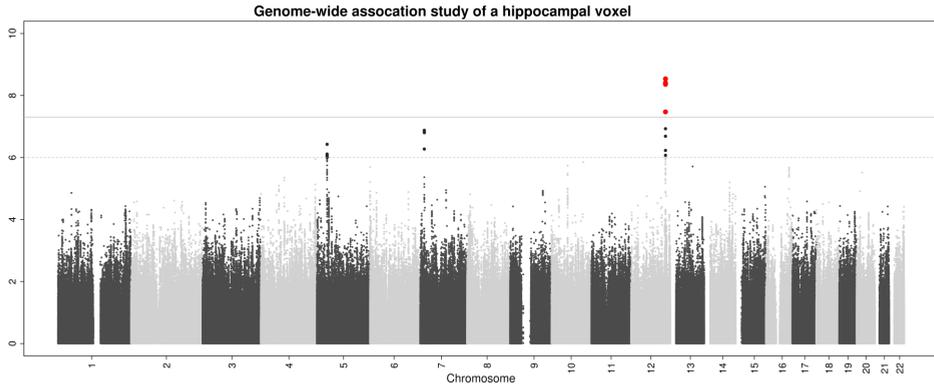
FIGURE 1: Manhattan plot of the hippocampus voxel with the most significant association after screening all 7030 hippocampal voxels. The most significant association (rs77956314; p = 3 x 10-9) corresponded to a previously identified locus on chromosome 12q24. Such voxel-wise hippocampus screening would take less than 8 hours on standard laptop.

Additionally, we performed the high-dimensional association studies separately in the three subcohorts of the Rotterdam Study and meta-analyzed the results using the HASE data reduction approach. It took on average 40 minutes for each subcohort to generate intermediate data for subsequent meta-analysis on a single CPU for all genetic variants and voxels. The meta-analysis was performed on the same cluster and took 17 hours to complete. Next, we compared the association results of the pooled analysis with the meta-analysis. **Figure 2** shows that the results are identical as it was predicted by theory (see **Methods**).
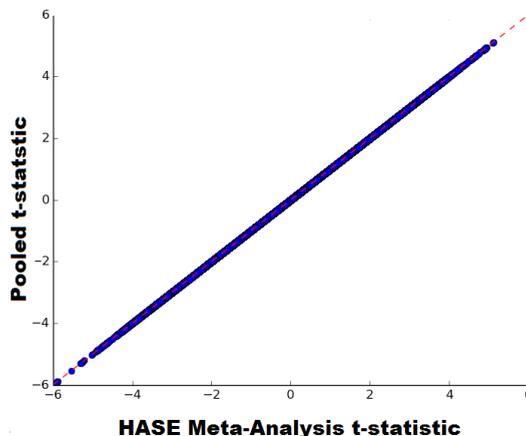
FIGURE 2: Correlation plot of voxel GWAS t-statistic estimated from pooled together data and voxel GWAS t-statistic estimated from meta-analysis of partial derivatives and encoded matrix. It took 40 min for single site to pre-compute data instead of 280 years to compute summary statistics.

DISCUSSION

We describe a framework that allows for (i) computationally-efficient high-dimensional association studies within individual sites using standard computational infrastructure and (ii) facilitates the exchange of compact summary statistics for subsequent meta-analysis for association studies in a collaborative setting. Using HASE, we performed a genome-wide and brain-wide search for genetic influences on voxel densities (more than 1.5 million GWAS analysis in total), and illustrate both its feasibility and potential for driving scientific discoveries.

A large improvement in efficiency comes from the reduced computational complexity. High-dimensional analyses contain many redundant calculations, which were removed in the HASE. Also, we were able to further increase efficiency by simplifying the calculations to a set of matrix operations, which are computationally inexpensive, compared to conventional linear regression algorithms. Furthermore, the implementation of partial derivatives meta-analysis allowed us to greatly reduce the size of the summary statistics that need to be shared for performing a meta-analysis. Another advantage of this approach is that it only needs to calculate the partial derivatives for each site instead of the parameter estimates (i.e., beta coefficients and standard errors). This enabled us to develop within HASE a reduction approach that encodes data prior to exchange between sites, while yielding the exact same results after meta-analysis as if the original data were used. The encoding is performed such that tracing back to original data is impossible. This guarantees protection of participant privacy and circumvents restrictions on data sharing that are unfortunately common in many research institutions.

When using HASE, it is first necessary to convert the multi-dimensional data to a format that is optimized for fast reading and writing. This particular format, «hdf5», is not dependent on the architecture of the file system and can therefore be implemented on a wide range of hardware and software infrastructures. To facilitate this initial conversion step, we have built-in tools within the HASE framework for processing common file format of such big data. Furthermore, this is easily generalizable to other large data matrices in general and we foresee this initial conversion step not to form an obstacle for researchers to implement HASE.

Alternative methods for solving the issues with high-dimensional data take one of two approaches. One approach is to reduce the dimensionality of the big datasets by summarizing the large amount of data into fewer variables[2]. Although this increases the speed, it comes at the price of losing valuable information, which these big data were primarily intended to capture. The second approach is to not perform a full analysis of all combinations of the big

datasets, but instead make certain assumptions (e.g., a certain underlying pattern, or a lack of dependency on potential confounders) that allow for using statistical models that require less computing time. Again, this is a tradeoff between speed and accuracy, which is not necessary in the HASE framework, where computational efficiency is increased without introducing any approximations.

Unidimensional analyses of big data, such as genome-wide association studies, have already elucidated to some extent the genetic architecture of complex diseases and other traits of interest[1,14–16], but much remains unknown. Cross-investigations between multiple big datasets potentially hold the key to fulfill the promise of big data in understanding of biology[7]. Using the HASE framework to perform high-dimensional association studies, this hypothesis is now testable.

METHODS

HASE

In high-dimensional associations analyses we test the following simple regression model:

$$Y = X\beta + \varepsilon$$

(1)

where **Y** is a $n_i \times n_p$ matrix of phenotypes of interest, n_i denotes the number of samples in the study, n_p the number of phenotypes of interest, and ε denotes the residual effect. **X** is a three dimensional matrix $n_i \times n_c \times n_t$ of independent variables, with n_c representing the number of covariates, such as the intercept, age, sex and, for example genotype as number of alleles, and n_t the number of independent determinants.

In association analyses we are interested in estimating the p-value to test the null hypothesis that $\beta = 0$. The p-values can be directly derived from the t-statistic of our test determinants. We will rewrite the classical equation for calculating t-statistics for our multi-dimensional matrices, which will lead to a simple matrix form solution for high-dimensional association analysis:

$$RSS(\beta) = (Y - X\beta)^T(Y - X\beta)$$

(2)

$$\frac{\partial RSS}{\partial \beta} = -2X^T(Y - X\beta)$$

(3)

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

(4)

$$RSS(\beta) = Y^TY - Y^TX(X^TX)^{-1}X^TY$$

(5)

$$T = \frac{\beta}{SE} = \frac{\beta}{\sqrt{diag\left((X^TX)^{-1}\right)\frac{RSS}{df}}} = \frac{(X^TX)^{-1}X^TY}{\sqrt{diag\left((X^TX)^{-1}\right)\frac{Y^TY - Y^TX(X^TX)^{-1}X^TY}{df}}}$$

(6)

Where **T** is $n_p \times n_c \times n_t$ matrix of t-statistics and **df** is degree of freedom of our regression model. Let's define $A = X^TX$, $B = X^TY$ and $C = Y^TY$ so that we can write our final equation for t-statistics:

$$T = A^{-1}B\sqrt{\frac{df}{diag\left(A^{-1}\right)(C - B^T A^{-1}B)}}$$

(7)

The result of this derivation is that, rather than computing all combinations of covariates and independent determinants, we only need to know three matrices: A, B and C, to calculate t-statistics and perform the full analysis. These results will be used in the section about meta-analysis.

The most computationally expensive operations here are the two multi-dimensional matrix multiplications $(A^{-1}B)$ and $(B^T A^{-1}B)$, where A^{-1} is a three dimensional matrix $\mathbf{n_c} \times \mathbf{n_c} \times \mathbf{n_t}$ and $\mathbf{B}$ is three dimensional matrix $\mathbf{n_c} \times \mathbf{n_p} \times \mathbf{n_t}$. Without knowledge of the data structure of these matrices, the simplest way to write the results of their multiplication would be to use Einstein's notation for tensor multiplication:

$$(A^{-1}\text{B})^i_{jk} = (A^{-1})^i_{ck} B^c_{jk} \tag{8}$$

$$(B^T A^{-1}B)^j_{\ k} = (B^T)^{jk}_i (A^{-1}\text{B})^i_{jk} \tag{9}$$

$$\text{where } i = \overrightarrow{1, n_c}; \ j = \overrightarrow{1, n_p}; \ k = \overrightarrow{1, n_t} \text{ and } c = \overrightarrow{1, n_c}$$

As you can see, the result is two matrices of $\mathbf{n_c} \times \mathbf{n_p} \times \mathbf{n_t}$ and $\mathbf{n_p} \times \mathbf{n_t}$ elements respectively. Despite the seemingly complex notation, the first matrix just represents the beta coefficients for all combinations of covariates ($\mathbf{n_c}$ by $\mathbf{n_p} \times \mathbf{n_t}$ combinations) and the second is fitting values of the dependent variable for every test ($\mathbf{n_p} \times \mathbf{n_t}$ independent determinants).

However, insight into the data structure of **A** and **B** can dramatically reduce the computational burden and simplify operations. First of all, matrix **A** depends only on the covariates and number of determinants, making it unnecessary to compute it for every phenotype of interest, so we just need to calculate it once. Additionally, only the last covariate (i.e., the variable of interest) is different between tests, meaning that the $(\mathbf{n_p}\text{-}1)\times(\mathbf{n_p}\text{-}1)\times\mathbf{n_t}$ part of matrix **A** remains constant during high-dimensional analyses. Matrix **B** consists of the dot product of every combination of the covariate and phenotype of interest. However, as we mentioned before, there are only $(\mathbf{n_t} + \mathbf{n_c}\ 1)$ different covariates, and thus we can split matrix **B** in two low dimensional matrices: the first includes dot products of non-tested covariates - $(\mathbf{n_c}\text{-}1)\times \mathbf{n_p}$ elements; the second includes the dot products only of the tested covariates - $\mathbf{n_p} \times \mathbf{n_t}$ elements. All this allows us to achieve large gain in computation efficiency and memory usage. In **Figure 3** we show a 2D schematic representation of these two matrices for standard genome association study with the covariates being an intercept, age, sex, and genotype. This example could be easily extrapolated to any linear regression model.

Applying the same splitting operation to B^T it is possible to simplify tensor multiplication equation **(8, 9)** to a low-dimensional matrix operation and rewrite the equation for t-statistics:

$$(A^{-1}\text{B})^i_{jk} = (A^{-1})^i_{\delta k} B^\delta_{\ j} + (A^{-1})^i_{\theta k} B^\theta_{\ jk} \tag{10}$$

$$(B^T A^{-1}B)^j_{\ k} = (B^T)^j_\delta (A^{-1}\text{B})^\delta_{jk} + (B^T)^{jk}_\theta (A^{-1}\text{B})^\theta_{jk} \tag{11}$$

$$\Gamma = ((A^{-1})^i_{\delta k} B^\delta_{\ j} + (A^{-1})^i_{\theta k} B^\theta_{\ jk}) \sqrt{\frac{df}{diag(A^{-1})}} \times \tag{12}$$

$$\times \left(C - \left((B^T)^j_\delta ((A^{-1})^i_{\delta k} B^\delta_{\ j} + (A^{-1})^i_{\theta k} B^\theta_{\ jk})^\delta_{ik} + (B^T)^{jk}_\theta ((A^{-1})^i_{\delta k} B^\delta_{\ j} + (A^{-1})^i_{\theta k} B^\theta_{\ jk})^\theta_{ik} \right) \right)^{-\frac{1}{2}}$$

Then, to compute t-statistics for high-dimensional association analyses we just need to perform several matrix multiplications.

$$A= \begin{bmatrix} \sum_{i}^{N} 1 & \sum_{i}^{N} age_i & \sum_{i}^{N} sex_i & \sum_{i}^{N} SNP_i \\ \sum_{i}^{N} age_i & \sum_{i}^{N} (age_i)^2 & \sum_{i}^{N} age_i sex_i & \sum_{i}^{N} age_i SNP_i \\ \sum_{i}^{N} sex_i & \sum_{i}^{N} age_i sex_i & \sum_{i}^{N} (sex_i)^2 & \sum_{i}^{N} sex_i SNP_i \\ \sum_{i}^{N} SNP_i & \sum_{i}^{N} age_i SNP_i & \sum_{i}^{N} sex_i SNP_i & \sum_{i}^{N} (SNP_i)^2 \end{bmatrix}$$

$$B= \begin{bmatrix} \overset{b_1}{\sum_{i}^{N} y_i} & \overset{b_2}{\sum_{i}^{N} y_i age_i} & \overset{b_3}{\sum_{i}^{N} y_i sex_i} & \overset{b_4}{\sum_{i}^{N} y_i SNP_i} \end{bmatrix}$$
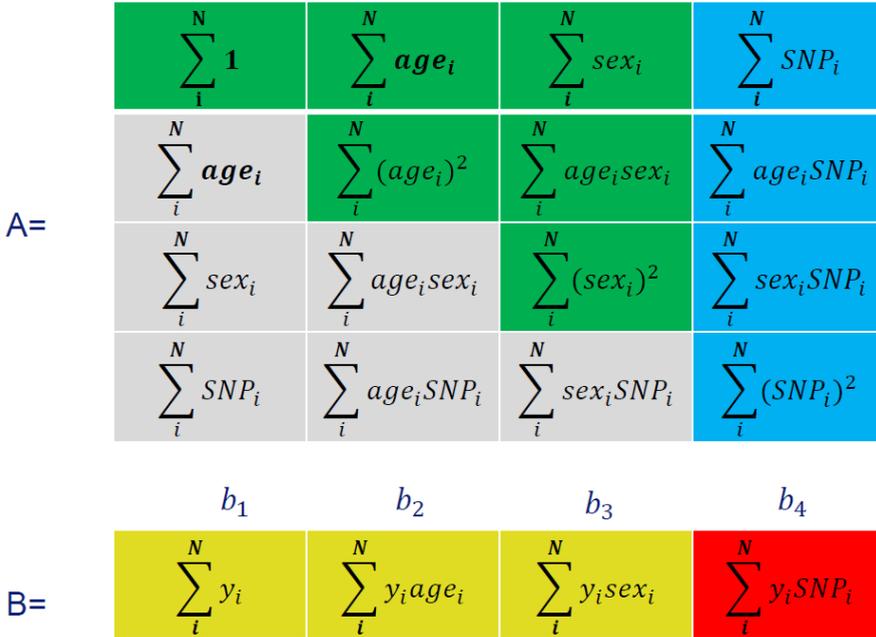
FIGURE 3: Explanation of the achieved speed reduction in HASE framework by removing redundant computations. In HASE multi-dimensional **A** and **B** matrices need to be calculated to perform GWAS studies. In the figure grey color means elements are parts of the matrix that are not necessary to calculate, as the **A** matrix is symmetric. The green color indicates elements that need to be calculated only once. Blue elements only have to be calculated for every SNP and yellow only for every phenotype. The red color indicates the most computationally expensive element, which needs to be calculated for every combination of phenotype and genotype. N denotes the number subjects in study.

META-ANALYSIS

In classical meta-analysis, summary statistics such as beta coefficients and p-values are exchanged between sites. For 1.5 million phenotypes, this would yield around 400TB of data at each site, making data transfer to a centralized site impractical.

In the previous section we showed that, to compute all statistics for an association study, we just need to know the **A, B** and **C** matrices. As we demonstrated before[17], by exchanging these matrices between sites, it is possible to gain the same statistical power as with a pooled

analysis, without sharing individual participant data, because these matrices consist of aggregate data (**Figure 3**). However, in high-dimensional association analyses, matrix **B** grows very fast, particularly the part that depends on the number of determinants and phenotypes (b_4 in **Figure 3**).

If **Y** is a $n_i \times n_p$ matrix of phenotypes of interest and **G** is a $n_i \times n_t$ matrix of determinants which we want to test (e.g., a genotype matrix in GWAS), then $b_4 = Y^T \times G$. These two matrices, **Y** and **G**, separately are not so large, but their product matrix has $n_p \times n_t$ elements, which in a real application could be $10^6 \times 10^7 = 10^{13}$ elements and thus too large to share between sites. We propose to create a random $n_i \times n_i$ nonsingular square matrix **F** and calculate its inverse matrix **F^{-1}**. Then by definition **F$\times$ F^{-1}=I**, where **I** is a $n_i \times n_i$ elements identity matrix with ones on main diagonal and zeros elsewhere. Using this property, we can rewrite the equation for b_4:

$$b_4 = Y^T \times G \tag{13}$$

$$b_4 = Y^T \times (F \times F^{-1}) \times G \tag{14}$$

$$b_4 = (Y^T \times F) \times (F^{-1} \times G) \tag{15}$$

$$b_4 = Y_F{}^T \times G_F \tag{16}$$

where Y_F and G_F are matrices carrying phenotypic and determinant information in encoded form respectively

Therefore, instead of transferring TBs of intermediate statistics (b_4), each side just needs to compute **A, C, Y$_F$ and G$_F$**.

Sharing just the encoded matrices does not provide information on individual participants and without knowing matrix **F** it is impossible to reconstruct the real data. However, it will be possible to calculate b_4, perform a high-dimensional meta-analysis, and avoid problems with data transfer. Additionally, this method dramatically reduces computation time by shifting all complex computations to central site, where the HASE regression algorithm should be used to handle the association analysis in time efficient way.

AVAILABILITY

Framework for efficient high-dimensional association analyses (HASE), https://github.com/roshchupkin/HASE/; description of the framework and protocol for meta-analysis, www.imagene.nl/HASE;

REFERENCE

1. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–1186 (2014).

2. Sambasivarao, S. V. Spatial patterns of genome wide expression profiles reflect anatomic and fiber connectivity architecture of healthy human brain. **18,** 1199–1216 (2013).

3. Stunnenberg, H. G. & Hubner, N. C. Genomics meets proteomics: Identifying the culprits in disease. *Hum. Genet.* **133,** 689–700 (2014).

4. Krumsiek, J. *et al.* Mining the Unknown: A Systems Approach to Metabolite Identification Combining Genetic and Metabolic Information. *PLoS Genet.* **8,** e1003005 (2012).

5. Parmar, C. *et al.* Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Sci. Rep.* **5,** 11044 (2015).

6. Medland, S. E., Jahanshad, N., Neale, B. M. & Thompson, P. M. Whole-genome analyses of whole-brain data : working within an expanded search space. *Nat. Publ. Gr.* **17,** 791–800 (2014).

7. Robinson, M. R., Wray, N. R. & Visscher, P. M. Explaining additional genetic variation in complex traits. *Trends Genet.* **30,** 124–32 (2014).

8. Zou, F. *et al.* Brain Expression Genome-Wide Association Study (eGWAS) Identifies Human Disease-Associated Variants. *PLoS Genet.* **8,** e1002707 (2012).

9. Stein, J. L. *et al.* Voxelwise genome-wide association study (vGWAS). *Neuroimage* **53,** 1160–74 (2010).

10. Huang, M., Nichols, T., Huang, C., Yang, Y. & Lu, Z. FVGWAS : Fast Voxelwise Genome Wide Association Analysis of Large-scale Imaging Genetic Data 1.

11. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** (2007).

12. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).

13. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **8,** (2015).

14. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460,** 748–52 (2009).

15. Polychronakos, C. & Alriyami, M. Diabetes in the post-GWAS era. *Nat. Genet.* **47,** 1373–1374 (2015).

16. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).

17. Adams, H.H.H. *et al.* Partial derivatives meta-analysis: pooled analyses without sharing individual participant data. BioRxiv (2016).

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

GVR and HHHA jointly conceived of the study, participated in its design, performed the statistical analysis, interpreted the data, and drafted the manuscript. GVR additionally developed the software. MWV, AH, and CMvD acquired data and revised the manuscript critically for important intellectual content. MAI and WJN participated in its design, interpreted the data, and revised the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

COMPETING FINANCIAL INTERESTS

WJN is co-founder and shareholder of Quantib BV. None of the other authors declare any competing financial interests.

SUPPLEMENTARY NOTE

HASE: Framework for efficient high-dimensional association analyses

STUDY POPULATION

The Rotterdam Study is an ongoing population-based cohort study in the Netherlands investigating diseases in the elderly and currently consists of 14,926 residents of Rotterdam who were aged 45 years or more at baseline [1,2]. The initial cohort was started in 1990 and expanded in 2000 and 2005. The whole population is subject to a set of multidisciplinary examinations every four years. MRI was implemented in 2005 and 5430 persons scanned until 2011 were eligible for this study. We excluded individuals with incomplete acquisitions, scans with artifacts hampering automated processing, participants with MRI-defined cortical infarcts, and subjects with dementia or stroke at the time of scanning. This resulted in a final study population of 4071 non-demented persons with information available on both genome-wide genotyping and MRI data. The Medical Ethics Committee of the Erasmus MC, University Medical Center Rotterdam and the review board of the Netherlands Ministry of Health, Welfare and Sports both approved the study. Informed consent was obtained from all subjects.

IMPUTATION OF GENOTYPES

The Illumina 550K and 550K duo arrays were used for genotyping. Samples with low call rate (<97.5%), with excess autosomal heterozygosity (>0.336) or with sex-mismatch were excluded, as were outliers identified by the identity-by-state clustering analysis (outliers were defined as being >3 standard deviation (SD) from population mean or having identity-by-state probabilities >97%). A set of genotyped input SNPs with call rate >98%, MAF >0.001 and Hardy–Weinberg equilibrium (HWE) P-value $> 10^{-6}$ was used for imputation. The Markov Chain Haplotyping (MACH) package version 1.0 software (Imputed to plus strand of NCBI build 37, 1000 Genomes phase I version 3) and minimac version 2012.8.6 were used for imputation.

MRI DATA

From August 2005 onwards, a dedicated 1.5 Tesla MRI scanner (GE Healthcare, Milwaukee, Wisconsin, USA) is operational in the Rotterdam Study research center in Ommoord. This scanner is operated by trained research technicians and all imaging data are collected according to standardized imaging protocols[2]. Brain MRI scans included a high-resolution 3D T1-weighted fast RF spoiled gradient recalled acquisition in steady state with an inversion recovery pre-pulse (FASTSPGR-IR) sequence with thin slices (voxel size<1mm^3)[2]*Image processing*

Voxel based morphometry (VBM) was performed according to an optimized VBM protocol [3]. First, all T1-weighted images were segmented into supratentorial gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) using a previously described k-nearest neighbor (kNN) algorithm, which was trained on six manually labeled atlases [4]. FSL software [5] was used for VBM data processing. First, all GM density maps were non-linearly registered to the standard GM probability template. For this study we chose the ICBM MNI152 GM template (Montreal Neurological Institute) with a 1x1x1 mm^3 voxel resolution. The MNI152 standard-space T1-weighted average structural template is derived from 152 structural images, which have been warped and averaged into the common MNI152 co-ordinate system after high-dimensional nonlinear registration.

A spatial modulation procedure was used to avoid differences in absolute GM volume due to the registration. This involved multiplying voxel density values by the Jacobian determinants estimated during spatial normalization. To gain more statistic power and decrease signal to noise ratio, all images were smoothed using a 3mm (FWHM 8mm) isotropic Gaussian kernel.

STATISTICAL ANALYSIS

Linear regression models were fitted with voxel values of GM modulation density as the dependent variable and age, sex, and the number of minor alleles as independent variables. In total 1,534,602 voxels were processed.

REFERENCES

1. Albert Hofman et al. "The Rotterdam Study: 2012 objectives and design update." In: European journal of epidemiology 26.8 (Aug. 2011), pp. 657

2. M. Arfan Ikram et al. "The Rotterdam Scan Study: design and update up to 2012." In: European journal of epidemiology 26.10 (Oct. 2011), pp. 811

3. C. D. Good et al. "A voxel-based morphometric study of ageing in 465 normal adult human brains." In: NeuroImage 14.1 Pt 1 (July 2001), pp. 21

4. Henri a Vrooman et al. "Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification." In: NeuroImage 37.1 (Aug. 2007)

5. Stephen M. Smith et al. "Advances in functional and structural MR image analysis and implementation as FSL." In: NeuroImage 23 Suppl 1 (Jan. 2004)

5

Application of the new imaging genetics approaches

CHAPTER 5.1

Genetic architecture
of the human
anterior commissure

ABSTRACT

The anterior commissure is a nerve fiber bundle that interconnects the brain's hemispheres. It is important for higher-order cognitive functions and has a presumed role in neurodegeneration, but its genetic architecture remains elusive. Uncovering the genes underlying variation in the anterior commissure could aid understanding of its development and it potentially provides mechanistic insight into diseases that results from its dysfunction. We performed a two-stage genome-wide association study (N = 18,828) of the size of the anterior commissure and identified six independent variants at four loci (p-values ranging from 4.1×10^{-8} to 9.4×10^{-22}). Using *in silico* and *in vitro* approaches, we mapped the loci to probable causal genes involved in axon guidance (*EPHA3* and *SEMA6A*), cognitive disorders (*CTNND2*), and growth factor signaling (*RIT2*). Also, voxel-based morphometry revealed distinct associations of the variants with connected grey matter regions in the brain. Genome-wide analyses revealed an enrichment for H3K4me1 peaks (marking enhancer sites), introns, and conserved sequences, as well as cell-type-specific annotations from the central nervous system and cardiovascular system. Furthermore, we identify pleiotropy between genes known to increase risk of neurodegenerative diseases and loci of the anterior commissure, including frontotemporal lobar degeneration gene *TMEM106B*. These analyses shed light on the genetic architecture of commissural tracts and establish the size of the anterior commissure as a relevant marker of neurodegeneration.

INTRODUCTION

Commissural tracts in the brain facilitate the cross-talk between the cerebral hemispheres. As such, they are essential for coordinating bilateral sensory and motor functions,[1-4] as well as for integrating lateralized cognitive and behavioral functions.[1,5-7] Furthermore, abnormalities in commissures are present in patients suffering from neurodegenerative diseases.[8-11] Our knowledge on the function and dysfunction of these tracts may be expanded by identifying factors that influence them, with genetics playing an important role.

The white matter of the brain as a whole is highly heritable,[12] but the commissural tracts in particular are under tight genetic control.[13,14] Most of our genetic understanding of brain commissures stems from studies of model organisms. These studies, mainly in mice and fruit flies, show an important role for several families of axon guidance proteins, including Ephrins, Semaphorins, Robos, Slits, and Netrins.[15-19] By contrast, genes influencing the commissures in humans remain elusive. Some mutations leading to midline crossing defects have been identified in axon guidance proteins, hinting at a similarity with animal models.[20] Additionally, these proteins are active beyond brain development with roles in neuronal repair and regeneration,[21] and could therefore also be important for neurological disorders in adults.[22] To advance our understanding of the genes influencing brain commissures in humans, we performed genome-wide association studies in 18,828 persons, followed by functional validation through *in silico* and *in vitro* experiments. We focused on the anterior commissure, the second largest commissure in humans, which connects frontal, temporal, and occipital cortical regions as well as the olfactory bulbs and amygdalae.[23] The identified genes underlying variation in the cross-sectional area of the anterior commissure highlight axon guidance (*EPHA3* and *SEMA6A*) as well as pleiotropy with neurodegenerative diseases (*TMEM106B* and *GRN*).

RESULTS

GENOME-WIDE ASSOCIATION STUDIES

After determining that the anterior commissure cross-sectional area was significantly heritable (population-based h2 = 36%, p = 3.0 x 10^{-4}; family-based h2 = 32%, p = 6.1 x 10^{-25}; see **Online Methods**), we performed a two-stage meta-analysis of genome-wide association studies. The discovery sample consisted of seven cohorts (N = 7,935) and was subsequently meta-analyzed with ten replication cohorts (N = 10,893; total N = 18,828), all of European ancestry (**Supplementary Table S1**).

Both the discovery and replication results are shown in a single 'Manhat-twin' plot (**Figure 1A**). Four loci reached genome-wide significance in the combined sample (in red; 3p11.1, 5p15.2, 5q23.1, and 18q12.3; **Figures 1B-E**, respectively), whereas two loci were no longer significant after adding the replication sample (in turquoise; 7p21.3 and 8q24.22; **Supplementary Figures S1A-B**, respectively). To determine whether these associations were confounded by disease status, we repeated all analyses after excluding persons with stroke and dementia (N = 1,240; remaining N = 17,588), but this did not affect the findings (**Supplementary Figure S2**). More details on all six loci are provided in **Table 1**.

TABLE 1: Genetic loci associated with the anterior commissure cross-sectional area.

Locus	Lead variant	Position	A1	A2	Freq	Discovery (N = 7,935) β	P	Combined (N = 18,828) β	P	Heterogeneity (N = 17) I²	P
3p11.1	rs7650184	89530057	A	C	0.40	-.149	2.0 x 10^{-6}	-.172	2.5 x 10^{-19}	0	0.75
5p15.2	rs11748929	11370494	C	G	0.70	-.095	5.0 x 10^{-3}	-.114	4.1 x 10^{-8}	0	0.48
5q23.1	rs11948331	116673915	T	G	0.61	-.226	8.2 x 10^{-12}	-.187	9.4 x 10^{-22}	20.4	0.22
5q23.1	rs148925592	117348902	A	T	0.96	-.437	2.8 x 10^{-6}	-.333	5.8 x 10^{-9}	45.8	0.024
18q12.3	rs346205	40013715	A	G	0.64	-.110	8.2 x 10^{-4}	-.119	1.1 x 10^{-9}	10.8	0.33
7p21.3	*rs3807865*	*12250402*	*A*	*G*	*0.41*	*.179*	*1.7 x 10^{-8}*	*.095*	*6.2 x 10^{-7}*	*52.1*	*0.0079*
8q24.22	*rs1159153*	*132617311*	*T*	*C*	*0.32*	*.182*	*3.4 x 10^{-8}*	*.058*	*4.8 x 10^{-3}*	*46.2*	*0.022*

ᵃ Variant missing in some cohorts due to insufficient imputation quality.

Abbreviations: A1 = effect allele, A2 = reference allele, Freq = frequency of the effect allele, N = sample size.

FUNCTIONAL CHARACTERIZATION

To elucidate potential mechanisms underlying these genetic associations we investigated the functional role of variants at the four significant loci. For each locus, we examined all genes lying within a 2 Mb window of the lead variant and prioritized genes based on their spatial expression patterns in six human brains (**Figure 2A, Supplementary Figure S3**; see **Online Methods**). Of particular interest were genes expressed in the white matter and in the grey matter of frontal, temporal, and central structures (**Figure 2B**). Next, we investigated whether the lead variants and those in linkage disequilibrium (LD) were quantitative trait loci (QTL) or predicted to be damaging by searching publicly available databases (**Supplementary Table S2,** see **Online Methods**). Finally, we screened the putative causal genes for (rare) functional variants using exome sequencing (N = 1,479) and a dedicated exome chip (N = 8,087) in subsamples of our study populations (**Supplementary Table S3, see Online Methods**).

3p11.1 and EPHA3

The 3p11.1 signal surrounds the *EPHA3* gene and its downstream region (**Figure 1B)**, with no other genes in close proximity of the lead variant. *EPHA3* belongs to the Eph family of receptor tyrosine kinases, which is one of the five canonical families of axon guidance proteins. The mouse homolog EphA3 mediates segregation and pathfinding of callosal axons.[24] In humans, it is expressed in grey matter structures of the brain that are connected by the anterior commissure, including the piriform cortex (**Figures 2A-B**). The lead variant rs7650184 lies in the 3'-UTR and influences expression of *EPHA3* in brain tissue obtained from the temporal cortex (p = 3.3 x 10^{-7}). Also, the missense variant rs35124509 is in high linkage disequilibrium (r^2 = 0.93; D' = 0.99). Exome analyses of the *EPHA3* region supported the association of another variant in the 3'-UTR (rs73139148; p = 3.3 x 10^{-7}), but did not reveal any additional functional variants (**Supplementary Table S3**).

5P15.2 AND CTNND2

For 5p15.2, the signal is located in a narrow region inside the *CTNND2* gene (**Figure 1C**). The lead variant rs11748929 and those in LD are all intronic and have enhancer histone marks (H3K4me1, H3K27ac, and H3K9ac) almost exclusively in brain tissues (**Supplementary Table S2**). The *CTNND2* gene is also expressed in the brain (**Figure 2A-B**), but largely undetected in other parts of the body.[25] *CTNND2* encodes a neuron-specific member armadillo protein, also known as δ-Catenin or NPRAP, that is a member of the β-catenin superfamily. It was initially discovered as an interaction partner of Presenilin-1,[25,26] mutations of which cause familial Alzheimer's Disease. Subsequently, *CTNND2* has been related to various cognitive disorders, including Cri-du-Chat syndrome,[27] reading problems and mild intellectual disability,[28] and a variety of psychiatric disorders.[29] Zebrafish knockdowns show migration defects of neuron subpopulations,[28] and mice mutants reveal that *CTNND2* is critical for

cognitive function *in vivo*.[30] Exome sequencing of *CTNND2* identified a rare, but synonymous exonic variant (chr5:10973781; minor allele frequency (MAF) = 0.4%; p = 9.2 x 10^{-4}).

5q23.1 and *SEMA6A*

Also on chromosome 5, we find the strongest signal of our genome-wide association study at 5q23.1 (**Figure 1D**). The lead variant rs11948331 lies close to *LINC00992*, a long non-coding RNA that is not expressed in the brain and on which little is known. Additionally, there was an apparent second signal on 5q23.1 from a less common variant (rs148925592; minor allele frequency = 4%; r^2 = 0.002; D' = 0.21). Therefore, we performed conditional analyses in a subset of studies, and found this to be an independent signal (**Supplementary Table S4**). This variant has an almost two-fold larger effect and lies within another long non-coding RNA, *LOC102467224*, which remains uncharacterized.

The nearest coding gene to the primary signal is *SEMA6A*, which is 762 kb further in the proximal direction. *SEMA6A* is highly expressed in the human brain, particularly in the white matter (**Figure 2A-B**). Interestingly, murine Sema6a is crucial for neurons to cross the midline, and misrouted axons of the anterior commissure are still present in adult mice mutants.[31] However, we found no evidence for QTLs nor any enhancer marks that might implicate the identified variants in regulating the expression of *SEMA6A* or other genes. One rare exonic nonsynonymous variant with a predicted damaging effect was nominally significant (rs200578077, MAF = 1.0%, p=0.046), but it was not in LD with rs11948331 (r^2 = 0.002; D' = 0.01) and thus unlikely to explain its effect. However, given the expression pattern of *SEMA6A* and the phenotype of Sema6a mutants, but lack of a direct link with our association signal, we hypothesized there might exist long-distance interaction between the region harboring the associated variants and the *SEMA6A* promotor. We first searched a curated database of genome-wide chromosome interaction and found that two long-distance interactions were described using Hi-C for both regions, but these did not include interaction with the promotor (**Supplementary Table S5**).[32] However, since none of the available cell lines for Hi-C were derived from the brain, we performed Targeted Chromatin Capture (T2C)[33] in human neural progenitor cells to selectively interrogate the 5q23.1 region at a high resolution (see **Online Methods**). We identified three topological associated domains (TADs) in a 5 Mb region surrounding the variants of interest (**Figure 2D**). The larger TAD contains both the *SEMA6A* promotor and the associated variants, indeed showing that, three-dimensionally, these genomic regions are in close proximity (**Figure 2E**).

18q12.3 and *RIT2*

The final locus spans an intergenic region on chromosome 18 that is flanked by two recombination hotspots (**Figure 1E**). Conditional analyses revealed that, independent of the lead

variant rs346205, there was another association at 18q12.3 with rs144695388 (**Supplementary Table S4**). The lead variant lies inside the long non-coding RNA *LINC00907* and between the coding genes *PIK3C3* and *RIT2*. Brain expression data shows *PIK3C3* is highly expressed in the white matter (**Figure 2A-B**), whereas *RIT2* as well as the more distal gene *SYT4* have a particularly low expression (**Figure 2A**). Putative functional variants in LD have enhancer marks in the brain and some acting as expression QTLs (**Supplementary Table S2**), including rs346212 that influences *RIT2* expression in the anterior cingulate cortex (p = 8.6 x 10^{-7}). *RIT2* belongs to the Ras superfamily of small guanosine triphosphate binding proteins and expressed in neurons, particularly in the substantia nigra.[34] A large GWAS of Parkinson's disease (PD) discovered rs12456492,[35] which is 660 kb distal of our lead variant, but it was not associated with the anterior commissure (p = 0.30). Through exome analyses of all nearby genes a rare missense mutation was found in *RIT2* (rs142911081; MAF = 1.5%; p = 1.0 x 10^{-3}) that was evolutionarily constrained and had an active transcription start site chromatin state in all samples taken from the brain, but in none of the other tissues (**Supplementary Table S5**).

NEURAL SUBSTRATE

Next, we set out to gain insight into the neural effects of the identified loci. The anterior commissure contains multiple fiber tracts that connect various cortical and subcortical grey matter regions. Therefore, we aimed to determine which grey matter regions are connected by the anterior commissure using voxel-based morphometry in the discovery sample (N = 9,934; includes individuals without genotyping). A larger anterior commissure cross-sectional area was associated with more tissue in the many subcortical and cortical regions, particularly the thalamus, insula, and the posterior temporal lobe (**Figure 3; Supplementary Table S6**).Next, we explored the effect of the four loci on grey matter within these regions that are presumably connected by the anterior commissure (N = 7,579; **Figure 3A-D**). Regions that were significantly associated with both the anterior commissure and the individual loci (p < 3.0 x 10^{-7}) included the anterior temporal lobe (3p11.1, 5p15.2, 18q12.3), insula (5q23.1), superior frontal gyrus (5p15.2, 5q23.1, 18q12.3), putamen (5p15.2, 5q23.1), and caudate (3p11.1, 5p15.2).

ENRICHMENT ANALYSES

Besides studying the top loci, we also investigated whether there is genome-wide enrichment for certain groups of genetic variants (**Supplementary Figure S4**). We found nominally significant enrichment for a variety of functional groups, most notably variants in enhancers (marked by H3K4me1 peaks), introns, and conserved regions (**Supplementary Figure S4A**), but these did not survive multiple testing correction. Furthermore, there were cell-type-specific enrichments for the central nervous system (p = 0.014) and cardiovascular system (0.0048)

(**Supplementary Figure S4B**). The proportion of heritability explained was divided fairly homogenously across chromosomes (**Supplementary Figure S4C**)

PLEIOTROPY WITH NEURODEGENERATIVE DISEASES

Given the relevance of commissural tracts and their connected brain regions for neurodegeneration, we determined whether genetic loci for AC show pleiotropy with neurodegenerative diseases. Using published genome-wide association studies of AD (n_{cases} / $N_{controls}$ =17,008 / 37,154),[36] PD (13,708 / 95,282),[35] and frontotemporal lobar degeneration (FTLD; 567 / 3,380)[37] we took two approaches to investigate pleiotropy.

First we investigated whether there is genome-wide overlap between these traits (see **Online Methods**). We found no genetic correlation with AD ($\rho_{genetic}$ = 0.116, p = 0.38) and a positive correlation with PD ($\rho_{genetic}$ = 0.156, p = 0.032), while LD score regression was not suited for analyzing FTLD given the little enrichment (mean χ^2 < 1.02).

In the second approach we aimed to identify specific pleiotropic loci by ranking all overlapping variants between anterior commissure, AD, PD, and FTLD (N = 5,188,883) by their respective p-value order for each of these traits, a recently developed method that is powerful for traits that are not on the same scale (see **Online Methods**). Next, we summed these four ranks using seven different combinations of traits: three pairs of the anterior commissure and a disease, three triples of the anterior commissure and two diseases, and a single combination of all four traits. Two loci were significantly pleiotropic for three combinations: 7p21.3 (p = 3.4 x 10^{-14} with FTLD), and 17q21.3 (p = 9.8 x 10^{-10} with FTLD and PD, and p = 3.1 x 10^{-9} with AD, FTLD, and PD).

The 7p21.3 locus was genome-wide significant at the discovery stage (**Table 1**) and previously reported to be associated with FTLD.[37] Interestingly, it also nearly reached genome-wide significant for the pleiotropy with AD (p = 4.0 x 10^{-7}), as well as both diseases together (p = 8.4 x 10^{-8}). The association signal of all three traits overlaps and spans the *TMEM106B* gene (**Figure 4A**). Since the 7p21.3 association with the AC was attenuated after addition of the replication samples and there was significant heterogeneity across the cohorts (**Table 1**), we set out to identify factors underlying these apparent inconsistencies. We previously showed that the effect of 7p21.3 increases with age (in a subset of the discovery sample; N = 4,413),[38] and the mean age within the cohorts was indeed related to the magnitude of the effect in a meta-regression analysis (**Figure 4x**). Additionally, restricting to the discovery cohorts with a mean age > 65 years considerably increased the signal even though the sample size was much smaller (N = 3,015; p = 3.8 x 10^{-11}). Furthermore, we formally tested the presence of a gene-by-age interaction for the 7p21.3 locus by modelling an interaction term between

age and the lead variant, rs3807865. Since exchanging individual participant data was not possible due to restrictions in data sharing, we performed a partial derivatives meta-analysis,[39] which provides identical results as a pooled analysis while only using summary statistics (see **Online Methods**). Indeed, this showed a significant interaction between rs3807865 and age (N = 7,459; p = 7.0 x 10^{-3}).

The second locus was on 17q21.31, near the genes *GRN*, *FAM171A2*, and *ITGA2B* (**Figure 4B**). The T allele of the lead pleiotropic variant, rs5848, was associated with a smaller anterior commissure (p = 7.6 x 10^{-4}) and an increased risk of AD (p = 0.01), PD (p = 9.3 x 10^{-6}), and FTLD (p = 8.2 x 10^{-5}). The variant rs5848 lies in the 3'-UTR of *GRN*, and is correlated with its expression in multiple tissues.[40] This association is independent of the known PD risk locus (lead variant rs2867316) near the *MAPT* gene, approximately 1Mb further.

DISCUSSION

We elucidated the genetic architecture of the human anterior commissure in this comprehensive study, which extended from large-scale population imaging and genetics to *in silico* and *in vitro* experiments. We identified robust associations of six variants at four distinct loci for which we assign the most likely causative genes, and show the remaining genetic component can be explained by many additional common variants with small effects. Across the whole genome, there is enrichment for variants in regulatory regions, and those in genomic sequences with functional marks specific to the central nervous system and cardiovascular cell lines. Finally, we found that loci affecting the anterior commissure in the general population have clear links to neurodegenerative diseases, highlighting the genetic pleiotropy.

Brain circuitry develops through complex processes that include cell fate specification and migration, axon guidance, and synaptogenesis.[41] These processes are tightly regulated by genetics, especially for the brain's commissural tracts. We therefore studied the size of the anterior commissure in order to uncover how of the variance is due to genetics and which specific factors are important. Similar to other brain structures,[42] we find that the anterior commissure has a polygenic architecture, with most of the variance remaining unexplained by the four identified loci. Larger GWAS will likely shed light on the additional loci that involved, as has been successfully achieved for other complex traits.

The combined meta-analysis of the discovery and replication cohorts yielded four genome-wide significant loci. Although GWAS only points to genomic regions associated with a trait, we subsequently performed *in silico* and *in vitro* analyses aimed at determining the causal genes. First, we prioritized genes that are particularly expressed in the white matter of the brain and in cortical and subcortical grey matter regions that are connected by the anterior commissure. Although this pointed to certain genes in some instances (*EPHA3* and *SEMA6A*), the results were ambiguous for 18q12.3, where multiple genes have plausible expression pattern. As a second step, however, we investigated whether our specific variants (or those in high LD) themselves affect gene expression. We found that variants at two loci show direct evidence of influencing expression of *EPHA3* (3p11.1) and *RIT2* (18q12.3) in brain tissue samples. No eQTL effects were observed for 5p15.2 or 5q23.1, but indirect evidence indicates these might be regulatory regions as well. The 5p15.2 locus contains enhancer marks in neural cell lines, whereas the T2C experiment in neural stem cells showed that 5q23.1 and the *SEMA6A* promotor are in close proximity. The third approach to identify causal genes was by studying the exomes of the nearby genes, allowing the study of variants not captured by genotyping chips nor well imputed using 1000 Genomes reference panel. This could reveal (functional) coding variants that drive the primary GWAS signal or, alternatively, independent risk variants

that would further implicate putative causal genes. For 3p11.1, exome analyses supported the variant discovered through GWAS, suggesting that this locus indeed exerts its affect by influencing expression of *EPHA3*. Screening of the other loci revealed other, rare variants with nominally significant effects on the anterior commissure. Most notably was the missense mutation in *RIT2* that also showed pronounced functional marks exclusively in brain tissue.

While gene expression and eQTL analyses provided valuable functional information on the identified loci, an important limitation is that white matter tissue is underrepresented in most expression and chromatin state databases. The candidate gene of our top locus, *SEMA6A*, is expressed in the white matter but we found no evidence of eQTL effects in any of the brain tissue samples, which were mostly obtained from grey matter structures. However, *in silico* evidence indicated that this region might interact with distant genomic regions, and we hypothesized this could be with the *SEMA6A* promotor. Using T2C, a high-resolution method to investigate these interaction, we indeed found that the region harboring the risk variants is in close proximity with the *SEMA6A* promotor. These experiments were done in human neural stem cells, which can be cultured to a sufficient scale. Besides the spatial gene expression in the brain, it is also important to consider differential temporal expression, which is particularly crucial for axon guidance.[43,44] Further explorations of *SEMA6A* and the effect of the common variants in this locus should take these factors into account.

The identified loci highlight the concordance between animal models and findings in humans. In mice and fruit flies, experimental studies have revealed gene families that are crucial for the development of commissural tracts. Interestingly, two of our four loci are likely to regulate genes from these families, namely the Semaphorin *SEMA6A* and the Ephrin *EPHA3*. However, while there are similarities with model systems, each of the families contains many members and it is unclear which are relevant for humans. Here we specifically identify genes for the human anterior commissure. It is not remarkable that human genes have remained elusive, since it is currently not feasible to study this complex organization of brain circuitry *in vitro*. Genetic studies thus rely on *in vivo* imaging, an approach requiring large sample sizes that has only become attainable in the past years with the formation of imaging genetics consortia such as CHARGE and ENIGMA.

Our measure of interest was the cross-sectional area of the anterior commissure, which we used as a proxy for the amount of fibers crossing through the anterior commissure. Theoretically, a larger area could be due to an increase in the total number of fibers or, alternatively, due to more space between the fibers. However, the density of callosal fibers has been shown not to correlate with the callosal area, thus making the area a good proxy for the total amount of fibers.[45] Another consideration is that a smaller area might not necessarily translate into

less interhemispheric connections, but could also correspond to a different anatomical organization, e.g. fibers instead passing through the corpus callosum. Furthermore, although for example *SEMA6A* is expressed in the white matter, some of the proposed causative genes are expressed in the grey matter (*EPHA3* and *CTNND2*), in regions connected by the anterior commissure, as we have shown in our voxel-based morphometry analyses. It is therefore possible that not all identified loci are directly affecting the anterior commissure, but some might influence it secondarily by their effect on related cortical and/or subcortical grey matter regions.

Regardless of the underlying mechanism through which genes affect the anterior commissure, there seems to be pleiotropy with various neurodegenerative diseases. Genetic variants for AD, PD, and FTLD have been shown to jointly affect cognitive status,[46] and some have been reported to be shared risk factors for disease.[47] We now find an additional link through the anterior commissure. While commissures are formed during development by axon guidance molecules,[44] these remain important for adults with roles in neuronal repair and regeneration. Furthermore, commissural genes seem relevant for adult neurological disorders; as evidenced by the presence of misrouted fibers in AD and PD.[48] Here, we leveraged this potential link by looking at genome-wide overlap and identified two loci affecting the anterior commissure and at least one neurodegenerative disease: *TMEM106B* and *GRN*. Both of these loci were previously implicated in FTLD,[37] a disease affecting the anterior commissure as well as frontal and temporal brain regions interconnected by it. Furthermore, these loci have been found to also associate with other neurodegenerative diseases. [49]

While these specific pleiotropic loci exist, we found no evidence for this to be a genome-wide trend. This could be a true observation or represent a lack of power to detect this pleiotropy using GWAS summary statistics. LD score regression requires a minimal enrichment in association signal (mean $\chi^2 > 1.02$) that was for example not met by the FTLD GWAS, making genome-wide comparisons futile. However, it is clear that the anterior commissure and FTLD share at least several genetic risk factors, and larger studies might therefore reveal additional loci. Focusing on older population might be a fruitful approach for neurodegenerative diseases given the gene-by-age interaction of the *TMEM106B* locus. The pronounced effect at older age indicates that variation in the anterior commissure cross-sectional area is more due to neurodegeneration rather than developmental differences, allowing us to pick up effects of risk loci for FTLD, AD, and PD. Since the (dys)function of the anterior commissure remains understudied in humans, its relevance could very well extend to other neurological or psychiatric disorders, for which different study populations are more informative.

We demonstrated that the combination of imaging and genetics can be powerful to yield in-

sight into the biological underpinnings of complex human traits that are not easily studied or manipulated experimentally. Our results will hopefully lead to additional follow-up laboratory work aimed at characterizing the identified candidate genes. As a potential clinical application, though the area of the anterior commissure itself is not (yet) a diagnostic or predictive tool for a disease, it might serve as a marker of the underlying disease process, e.g. as an endophenotype. A quantitative measure could be statistically more powerful to detect genetic risk loci compared to a dichotomous diagnosis of disease. Indeed, we were able to identify the *TMEM106B* locus, which was originally discovered in a GWAS of FTLD (567 cases and 3,380 controls; $p = 2.7 \times 10^{-9}$), at a higher level of significance and a lower total sample size when restricting to older cohorts only ($N = 3,015$; $p = 3.8 \times 10^{-11}$). Particularly for rare diseases such as FTLD, where samples are obtained with difficulty, this approach could be helpful. Furthermore, the identification of risk genes for multiple disorders makes them attractive targets for pharmacological intervention.

In conclusion, we show that the anterior commissure has a considerable polygenic basis and we identified loci that begin to unravel this complex genetic architecture. The successful approach of imaging genetics to understanding developmental biology and disease pathophysiology paves the way for applications to other brain structures.

REFERENCES

1. Paul, L. K. *et al.* Agenesis of the corpus callosum: genetic, developmental and functional aspects of connectivity. *Nature Reviews Neuroscience* **8**, 287-299 (2007).

2. Fabri, M. *et al.* Contribution of posterior corpus callosum to the interhemispheric transfer of tactile information. *Cognitive Brain Research* **24**, 73-80 (2005).

3. Berlucchi, G. Some effects of cortical and callosal damage on conscious and unconscious processing of visual information and other sensory inputs. *Progress in brain research* **144**, 77-93 (2004).

4. Zaidel, D. & Sperry, R. W. Some long-term motor effects of cerebral commissurotomy in man. *Neuropsychologia* **15**, 193-204 (1977).

5. Häberling, I. S., Badzakova-Trajkov, G. & Corballis, M. C. Callosal tracts and patterns of hemispheric dominance: a combined fMRI and DTI study. *Neuroimage* **54**, 779-786 (2011).

6. Schulte, T. & Müller-Oehring, E. M. Contribution of callosal connections to the interhemispheric integration of visuomotor and cognitive processes. *Neuropsychology review* **20**, 174-190 (2010).

7. Schutter, D. J. L. G. & Harmon-Jones, E. The corpus callosum: A commissural road to anger and aggression. *Neuroscience & Biobehavioral Reviews* **37**, 2481-2488 (2013).

8. Rosas, H. D. *et al.* Altered white matter microstructure in the corpus callosum in Huntington's disease: implications for cortical "disconnection". *Neuroimage* **49**, 2995-3004 (2010).

9. Teipel, S. J. *et al.* Regional pattern of hippocampus and corpus callosum atrophy in Alzheimer's disease in relation to dementia severity: evidence for early neocortical degeneration. *Neurobiology of aging* **24**, 85-94 (2003).

10. Wiltshire, K., Foster, S., Kaye, J. A., Small, B. J. & Camicioli, R. Corpus callosum in neurodegenerative diseases: findings in Parkinson's disease. *Dementia and geriatric cognitive disorders* **20**, 345-351 (2005).

11. Yamauchi, H. *et al.* Comparison of the pattern of atrophy of the corpus callosum in frontotemporal dementia, progressive supranuclear palsy, and Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry* **69**, 623-629 (2000).

12. Posthuma, D. *et al.* The association between brain volume and intelligence is of genetic origin. *Nature neuroscience* **5**, 83-84 (2002).

13. Kochunov, P. *et al.* Genetics of microstructure of cerebral white matter using diffusion tensor imaging. *Neuroimage* **53**, 1109-1116 (2010).

14. Jahanshad, N. *et al.* Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the ENIGMA–DTI working group. *Neuroimage* **81**, 455-469 (2013).

15. Okada, A. *et al.* Boc is a receptor for sonic hedgehog in the guidance of commissural axons. *Nature* **444**, 369-373 (2006).

16. Brankatschk, M. & Dickson, B. J. Netrins guide Drosophila commissural axons at short range. *Nature neuroscience* **9**, 188-194 (2006).

17. Eichmann, A., Makinen, T. & Alitalo, K. Neural guidance molecules regulate vascular remodeling and vessel navigation. *Genes & development* **19**, 1013-1021 (2005).

18. Dickson, B. J. & Gilestro, G. F. Regulation of commissural axon pathfinding by slit and its Robo receptors. *Annu.*

Rev. Cell Dev. Biol. **22**, 651-675 (2006).

19. Dickson, B. J. & Senti, K.-A. Axon guidance: growth cones make an unexpected turn. *Current biology* **12**, R218-R220 (2002).

20. Izzi, L. & Charron, F. Midline axon guidance and human genetic disorders. *Clinical genetics* **80**, 226-234 (2011).

21. Curinga, G. & Smith, G. M. Molecular/genetic manipulation of extrinsic axon guidance factors for CNS repair and regeneration. *Experimental neurology* **209**, 333-342 (2008).

22. Van Battum, E. Y., Brignani, S. & Pasterkamp, R. J. Axon guidance proteins in neurological disorders. *The Lancet Neurology* **14**, 532-546 (2015).

23. Berlucchi, G. Visual interhemispheric communication and callosal connections of the occipital lobes. *Cortex* **56**, 1-13 (2014).

24. Nishikimi, M., Oishi, K., Tabata, H., Torii, K. & Nakajima, K. Segregation and pathfinding of callosal axons through EphA3 signaling. *The Journal of Neuroscience* **31**, 16251-16260 (2011).

25. Zhou, J. *et al.* Presenilin 1 interaction in the brain with a novel member of the Armadillo family. *Neuroreport* **8**, 2085-2090 (1997).

26. Levesque, G. *et al.* Presenilins Interact with Armadillo Proteins Including Neural-Specific Plakophilin-Related Protein and β-Catenin. *Journal of neurochemistry* **72**, 999-1008 (1999).

27. Medina, M., Marinescu, R. C., Overhauser, J. & Kosik, K. S. Hemizygosity of δ-catenin (CTNND2) is associated with severe mental retardation in cri-du-chat syndrome. *Genomics* **63**, 157-164 (2000).

28. Hofmeister, W. *et al.* CTNND2—a candidate gene for reading problems and mild intellectual disability. *Journal of medical genetics* **52**, 111-122 (2015).

29. Nivard, M. G. *et al.* Further confirmation of the association between anxiety and CTNND2: replication in humans. *Genes, Brain and Behavior* **13**, 195-201 (2014).

30. Israely, I. *et al.* Deletion of the neuron-specific protein delta-catenin leads to severe cognitive and synaptic dysfunction. *Current biology* **14**, 1657-1663 (2004).

31. Rünker, A. E. *et al.* Mutation of Semaphorin-6A disrupts limbic and cortical connectivity and models neurodevelopmental psychopathology. *International Journal of Developmental Neuroscience* **28**, 690 (2010).

32. Li, M. J., Wang, L. Y., Xia, Z., Sham, P. C. & Wang, J. GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. *Nucleic acids research*, gkt456 (2013).

33. Kolovos, P. *et al.* Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements. *Epigenetics Chromatin* **16**, 10 (2014).

34. Zhou, Q., Li, J., Wang, H., Yin, Y. & Zhou, J. Identification of nigral dopaminergic neuron-enriched genes in adult rats. *Neurobiology of aging* **32**, 313-326 (2011).

35. Nalls, M. A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics* (2014).

36. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* **45**, 1452-1458 (2013).

37. Van Deerlin, V. M. *et al.* Common variants at 7p21 are associated with frontotemporal lobar degeneration with

TDP-43 inclusions. *Nature genetics* **42**, 234-239 (2010).

38. Adams, H. H. H. *et al*. TMEM106B Influences Volume of Left-Sided Temporal Lobe and Interhemispheric Structures in the General Population. *Biological psychiatry* **76**, 503-508 (2014).

39. Adams, H. H. H. *et al*. Partial derivatives meta-analysis: pooled analyses when individual participant data cannot be shared. *bioRxiv*, 038893 (2016).

40. Consortium, G. T. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).

41. Jessell, T. M. & Sanes, J. R. Development: The decade of the developing brain. *Current opinion in neurobiology* **10**, 599-611 (2000).

42. Hibar, D. P. *et al*. Common genetic variants influence human subcortical brain structures. *Nature* (2015).

43. Barnes, G., Puranam, R. S., Luo, Y. & McNamara, J. O. Temporal specific patterns of semaphorin gene expression in rat brain after kainic acid induced status epilepticus. *Hippocampus* **13**, 1-20 (2003).

44. Brose, K. & Tessier-Lavigne, M. Slit proteins: key regulators of axon guidance, axonal branching, and cell migration. *Current opinion in neurobiology* **10**, 95-102 (2000).

45. Aboitiz, F., Scheibel, A. B., Fisher, R. S. & Zaidel, E. Fiber composition of the human corpus callosum. *Brain research* **598**, 143-153 (1992).

46. Adams, H. H. *et al*. Genetic risk of neurodegenerative diseases is associated with mild cognitive impairment and conversion to dementia. *Alzheimers Dement*, doi:10.1016/j.jalz.2014.12.008 (2015).

47. Moskvina, V. *et al*. Analysis of genome-wide association studies of Alzheimer disease and of Parkinson disease to determine if these 2 diseases share a common genetic risk. *JAMA neurology* **70**, 1268-1276 (2013).

48. Hoogland, P. V., Van Den Berg, R. & Huisman, E. Misrouted olfactory fibres and ectopic olfactory glomeruli in normal humans and in Parkinson and Alzheimer patients. *Neuropathology and applied neurobiology* **29**, 303-311 (2003).

49. Sheng, J., Su, L., Xu, Z. & Chen, G. Progranulin polymorphism rs5848 is associated with increased risk of Alzheimer's disease. *Gene* **542**, 141-145 (2014).

CHAPTER 5.2

Genome-wide association
study of cortical regions:
thickness, surface area and volume

ABSTRACT

Background

Cortical thickness, surface area and volume relate to ageing, cognitive function as well as neurological and psychiatric diseases. These morphological measures are heritable and genetically heterogeneous.

Methods

We performed genome-wide association studies of cortical thickness, surface area and volume in the whole cortex as well as in 34 regions of interest in 22822 individuals from 20 cohorts within the Cohorts of Heart and Aging Research in Genomic Epidemiology consortium and the United Kingdom Biobank. Significant findings were replicated in the Enhancing Neuroimaging Genetics through Meta-analysis consortium. Additionally we calculated heritability and genetic correlation between traits and regions.

Results

We identified 161 genome-wide significant associations for cortical thickness in the whole cortex and in 26 cortical regions. Genetic correlation was strong between surface area and volume, weak between surface area and thickness, and varied per region between thickness and volume.

Conclusions

Our findings indicate that the genetic architecture of the cerebral cortex varies by region. These data create a basis for targeted functional studies to identify the biological mechanisms behind these genetic signals.

Novel genetic discovery in this unpublished manuscript have been anonymized or removed following consortium-wide practices. As a result, the text will not go in-depth about gene functions and the identified pathways.

INTRODUCTION

Cortical thickness (CT), cortical surface area (CSA) and cortical volume (CV) are morphological markers of the cortex obtained by magnetic resonance imaging (MRI). These measures change with age[1-6] and are linked to cognitive functioning[5,7-10]. Abnormalities in global or regional CT, CSA and CV have been observed in neurological and psychiatric disorders such as Alzheimer's disease[11-17], Parkinson's disease[18-22], multiple sclerosis[23-26], migraine[27-29], schizophrenia[30-34], bipolar disorder[35-38], autism[39-41] and depression[42,43]. CT and CSA reflect different mechanisms in cortical development[44,45]. Neurons in the neocortex are organized in columns which run perpendicular to the surface of the cerebral cortex[46] and, according to the radial unit hypothesis, CT is determined by the number of cells within the columns and CSA is determined by the number of columns[44].

Previous studies indicate that the genetic background of CT and CSA varies by region[47-50] and that they are genetically uncorrelated and influenced by different genetic factors[51,52]. Thus, CV, which is the product of CT and CSA, is likely affected by a combination of these distinct genetic determinants. However, Winkler et al.[52] reported a strong genetic correlation between global CV and global CSA, but a weak genetic correlation between global CV and global CT. CT, CSA and CV are heritable traits[47,49,51-60]. The estimated heritability for global measures is high, ranging from 0.69 to 0.81 for global CT, and from 0.42 to 0.90 for global CSA[49,51,52,57]. In cortical regions, there is substantial variability in heritability estimates of CT, CSA and CV[47,49,51-56,58-60].

As CT, CSA and CV are differentially heritable and as they seem to be genetically heterogeneous these measures are promising targets for genome-wide association studies (GWAS). We analyzed genetic data of up to 22 822 individuals from 21 cohorts to investigate genetic influences on CT, CSA and CV in the whole cortex and in 34 cortical regions. Our data contribute to the understanding of cortical development and related traits and diseases.

METHODS

STUDY POPULATION

The sample of this study comprises of up to 22822 participants from 20 population-based cohort studies collaborating in the Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium[84] and the UK Biobank (UKBB)[85]. All the individuals were stroke- and dementia free, aged between 20 and 90 years, and of European ancestry, except for ARIC AA with African ancestry. Table S1 provides population characteristics of each cohort and Supplementary Section 1 provides a short description of each study. Each study secured approval from institutional review boards or equivalent organizations, and all participants provided written informed consent. Our results were replicated using summary GWAS findings of 22635 individuals from the Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA) consortium[86].

GENOTYPING AND IMPUTATION

Genome-wide genotyping was conducted using various commercially available genotyping arrays across the study cohorts. Prior to imputation, extensive quality control was performed in each cohort. Genotype data were imputed to the 1000 Genomes reference panel[87] (mainly phase 1, version 3) using validated software. Details on genotyping, quality control and imputation can be found in Table S2.

PHENOTYPE DEFINITION

This study investigated CT, CSA and CV globally in the whole cortex and in 34 cortical regions. Global and regional CT was defined as the mean thickness of the left and the right hemisphere in mm. Global CSA was defined as the total surface area of the left and the right hemisphere in mm^2, while regional CSA was defined as the mean surface area of the left and the right hemisphere in mm^2. Global and regional CV was defined as the mean volume of the left and the right hemisphere in mm^3. The 34 cortical regions are listed in Table S3. High resolution brain magnetic resonance imaging (MRI) data was obtained in each cohort using a range of MRI scanners, field strengths and protocols. CT, CSA and CV were generated using the Freesurfer software package[88,89] in all cohorts except for FHSucd, where an in-house segmentation method was used. MRI protocols of each cohort can be found in Table S4 and descriptive statistics of CT, CSA and CV can be found in Tables S5, S6 and S7.

Genome-wide associations, meta-analysis, replication and annotation

Based on a predefined analysis plan, each study fitted linear regression models to determine the association between global and regional CT, CSA and CV and allele dosages of single nucleotide polymorphisms (SNPs). Additive genetic effects were assumed and the models

were adjusted for sex, age, age^2, and if needed for study site and for principal components to correct for population stratification. Cohorts including related individuals calculated linear mixed models to account for family structure. Details on association software and covariates for each cohort are shown in Table S2. Models investigating regional CT, CAS and CV were additionally adjusted for global CT, global CSA and global CV, respectively. Quality control of the summary statistics shared by each cohort was performed using EasyQC[90]. Genetic Variants with a minor allele frequency (MAF) < 0.05, low imputation quality ($R^2 < 0.4$), and which were available in less than 10000 individuals were removed from the analyses. Details on quality control are provided in Supplementary Section 2.

We then used METAL[91] to perform meta-analyses using the z-scores method, based on p-values, sample size and direction of effect, with genomic control correction. There was no evidence of inflated test statistics; QQ plots are shown in Supplementary Figure 1.

We performed 10.000 permutation tests based on cortical measurements from Rotterdam Study to estimate the number of independent tests. Based on permutation test results, the genome-wide significance threshold was set a priori at 1.0869×10^{-9} ($= 5 \times 10^{-8} / 46$). We used the clumping function in plink[92] (linkage disequilibrium (LD) threshold: 0.2, distance: 300Kb) to identify the most significant SNP in each LD block.

For replication of our genome-wide significant CT and CSA associations, we used GWAS meta-analysis results from the ENIGMA consortium for all SNPs that were associated at a p-value $< 5 \times 10^{-8}$ and performed a pooled meta-analysis. The p-value threshold for replication was set to 3.1×10^{-4} (nominal significance threshold (0.05) divided by total number of lead SNPs (161)). CV was not available in the ENIGMA results.

Annotation of genome-wide significant variants was performed using the ANNOVAR software package[93] and the FUMA web application[94]. Regional association plots were generated with LocusZoom[95]. Gene-based analyses, to combine the effects of SNPs assigned to a gene, and gene set analyses, to find out if genes assigned to significant SNPs were involved in biological pathways, were performed using MAGMA[96] as implemented in FUMA. The significance threshold was set to 5.87×10^{-8} for gene-based analyses (nominal threshold (0.05) divided by number of genes (18522) and number of independent tests (46)), and to 1.02×10^{-7} for the gene-set analyses (nominal threshold (0.05) divided by the number of gene sets (10651) and by the number of independent tests (46)). The NHGRI-EBI Catalog of published GWAS[97] was searched for previous SNP-trait associations at a p-value of 5×10^{-8} of lead SNPs. We investigated cis (<1Mb) and trans (>1 MB or on a different chromosome) expression quantitative trait loci (eQTL) for genome-wide significant SNPs in 724 post-mortem brains

from the Religious Order Study and the Rush Memory and Aging Project (ROSMAP)[98,99] stored in the AMP-AD database. The samples were collected from the gray matter of the dorsolateral prefrontal cortex.

HERITABILITY

Additive genetic heritability (h^2) of CT, CSA and CV was estimated in two studies: the Austrian Stroke Prevention Family Study (ASPS-Fam; n=365) and the Rotterdam Study (RS, n=4472). In the population based family study ASPS-Fam, the ratio of the genotypic variance to the phenotypic variance was calculated using variance components models in SOLAR[100]. In case of non-normalty, phenotype data were inverse-normal transformed. In RS, SNP-based heritability was computed with GCTA[101,102]. These heritability analyses were adjusted for age and sex. Heritability and partitioned heritability based on GWAS summary statistics was calculated from GWAS summary statistics using LD score regression (LDSR) implemented in the ldst tool (https://github.com/bulik/ldsc). Partitioned heritability analysis splits genome-wide SNP heritability into 53 functional annotation classes (e.g. coding, 3' UTR, promoter, transcription factor binding sites, conserved regions etc.) and additionally to 10 cell-type specific classes (e.g. CNS, cardiovascular, liver, skeletal muscle etc.) as defined by Finucane et al.[103] to estimate their contributions to heritability. The significance threshold was set to 2.05×10^{-5} (0.05/Nr of functional annotation classes (53) / nr of independent tests (46)) for heritability partitioned on functional annotation classes and $2.05 < 10^{-6}$ (0.05/Nr of functional annotation classes (53) / number of cell types (10) / nr of idependent tests (46)) for heritability partitioned on annotation classes and cell types.

RESULTS

GENOME-WIDE ASSOCIATION ANALYSIS

The meta-analyses of global CT, CSA and CV included 22163, 18617 and 22822 individuals respectively. There were no significant associations ($p_{Discovery} < 1.09 \times 10^{-9}$) with global CT. As displayed in Table 1 and the Supplementary Figures 2 and 3, for global CSA and CV there were 6 and 7 independent lead SNPs. All but one for global CSA have been replicated in the corresponding ENIGMA analysis.

TABLE 1: Genome-wide significant associations ($p_{Discovery} < 1.09 \times 10$-9) of the global cortex

	CHR	A1	A2	Freq(A1)	Annotation	N	P Discovery	P Replication	P pooled
	17	C	G	0.20	intronic	18617	1.78E-23	**4.35E-20**	1.09E-41
	6	A	G	0.77	intergenic	18617	5.21E-22	**3.33E-14**	6.75E-34
CSA	17	A	G	0.21	intronic	17425	2.38E-20	**3.55E-19**	7.02E-38
	6	A	G	0.83	intergenic	18617	4.86E-13	**7.15E-08**	5.81E-19
	6	A	G	0.25	intergenic	18617	1.26E-10	**1.47E-06**	2.43E-15
	6	T	C	0.33	intronic	18617	6.84E-10	0.001047	3.78E-11
	6	A	G	0.77	intergenic	22448	5.86E-19	NA	NA
	17	A	G	0.79	intronic	22822	1.53E-13	NA	NA
	17	A	G	0.18	intronic	22137	6.46E-13	NA	NA
CV	17	T	C	0.79	intronic	21062	3.64E-11	NA	NA
	12	T	C	0.50	3'UTR	22822	8.56E-11	NA	NA
	12	T	C	0.78	intergenic	22822	4.83E-10	NA	NA
	17	A	G	0.76	intronic	22822	7.18E-10	NA	NA

CSA=cortical surface area, CV= cortical volume, A1: allele 1, A2: allele 2, Freq(A1): frequency of A1, N: number of individuals in analyses, $p_{Discovery}$: p-value of discovery GWAS meta-analysis in CHARGE, $p_{Replication}$: p-value of replication meta-analysis in ENIGMA, p_{pooled}: p-value of pooled discovery and replication meta-analysis, NA: not availbable, in bold: $p < 3.1 \times 10^{-4}$ (=0.05/NI, NI=161 (total number of lead SNPs))

GWAS for CT, CSA and CV of cortical regions of interest (ROIs) identified 161 significant associations. There were 16 independent hits on 8 chromosomes in 9 cortical regions for CT (Table S9), 54 hits on 16 chromosomes in 21 cortical regions for CSA (Table S10), and 78 associations on 17 chromosomes in 23 cortical regions for CV (Table S11). As can also be seen from Tables S9 – S11, 13 ROI-specific lead SNPs were replicated for CT and 49 for CSA. Chromosomal ideograms showing genome-wide significant associations of the discovery stage are presented in Figure 1.

In the discovery stage, the lowest p-values for CT and CV were found for loci in chromosome15 ($p_{Discovery, CT} = 1.17 \times 10^{-73}$ and $p_{Discovery, CV} = 4.34 \times 10^{-133}$) in the postcentral cortex, and for CSA loci also on chromosome 15 ($p_{Discovery, CSA} = 8.45 \times 10^{-109}$) in the precentral cortex.

The postcentral cortex was also the region with the largest number of independent associations with 5, 4 and 9 index SNPs for CT, CSA and CV. We found 4 and 25 common lead

SNPs within the same cortical region for CT and CV, and CSA and CV, respectively , but not for CT and CSA.
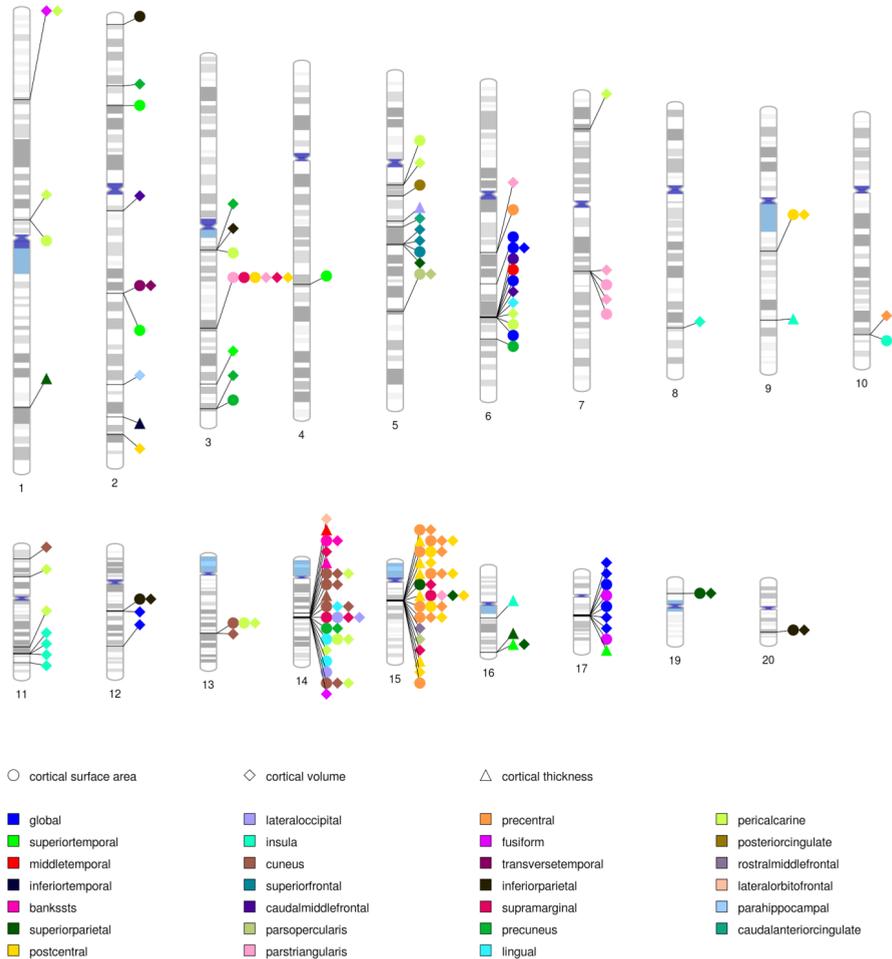
FIGURE 1: Chromosomal ideogram annotated with genome-wide significant assocations ($p_{Discovery} < 1.09 \times 10^{-9}$).

Positional mapping based on ANNOVAR showed that the majority of lead SNPs was intergenic ($N_{CT}=12$, $N_{CSA}=38$, $N_{CV}=57$) and intronic ($N_{CT}=4$, $N_{CSA}=18$, $N_{CV}=18$) (**Figure2**).

Lead SNPs of global and regional CT, CSA and CV, observed in the current GWAS, were previously associated with biometric (height), neurologic (Parkinson's disease, Corticobasal degeneration, Alzheimer's disease biomarkers), psychiatric (Neuroticism, Schizophrenia) and cognitive traits as well as with educational attainment and total intracranial volume (TIV) .
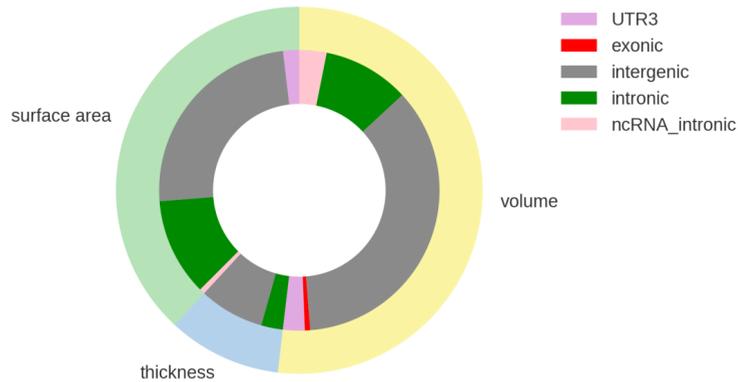
FIGURE 2: Proportion of functional annotation categories for global and regional cortical thickness, surface area and volume assigned by ANNOVAR.

GENE MAPPING

Additionally to positional annotation we used several strategies to link top SNPs to specific genes. First, we used FUMA for eQTL analysis, based on GTEx database, and for chromatin interaction mapping. Second, for brain eQTL analysis we used the AMP AD database. Last, based on SNP's p-values within genes we performed MAGMA analysis to compute gene-based p values. MAGMA gene-based association analyses revealed 2 and 12 genes significantly associated ($p < 5.87e^{-8}$) with global and regional CT respectively. 9 and 47 genes associated with global and regional CSA, and also 9 and 47 genes associated with global and regional CV. For global CSA and CV, 7 of the 9 significant genes overlapped, but there was no overlap with global CT. For regional CSA and CV we found 28 genes overlapping within the same 13 cortical regions. The same genes were significant for CT and CV of the superior parietal cortex and the insula but there were no common genes within a cortical region for CT and CSA. **Figure 3** shows the number of overlapping unique genes between the different mapping strategies for CSA, CT and CV.

PATHWAY ANALYSIS

MAGMA gene set analyses identified 7 pathways for CT, 3 pathways for CSA and 9 pathways for CV. Among them are the Gene Ontology (GO) gene sets 'hindbrain morphogenesis' for CT of middle temporal cortex, 'forebrain generation of neurons' for precentral CSA, and 'central nervous system neuron development' for transverse temporal CV. However, after Bonferroni correction only one significant ($p < 1.02 \times 10^{-7}$) pathway remained: 'regulation of catabolic process' for CT of inferior temporal cortex.
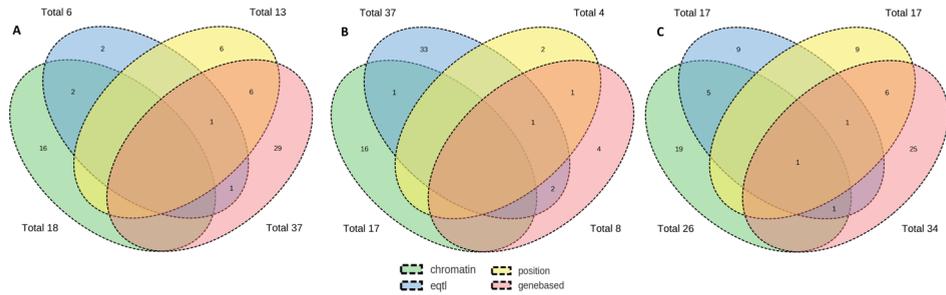
FIGURE 3: Number of overlapping genes between FUMA eQTL mapping, FUMA chromatin interaction mapping, ANNOVAR chromosome positional mapping and MAGMA gene based analysis for all cortical regions combined for cortical surface area **(A)**, thickness **(B)** and volume **(C)**.

HERITABILITY

Heritabiliy estimates (h^2) of global CT were 0.64 (se=0.12; p=3×10^{-7}) in ASPS-Fam and 0.45 (se=0.08; p=2.5×10^{-7}) in RS. For CSA, h^2 was 0.84 (se=0.12; p=2.63×10^{-11}) in ASPS-Fam and 0.33 (se=0.08, p=1×10^{-4}) in RS, and for CV, h^2 was 0.80 (se=0.11; p=$1.10\times10-9$) in ASPS-Fam and 0.32 (se=0.08; p=$1x10^{-4}$) in RS. There was a large range in heritability estimates of regional CT, CSA and CV (Table S25).

Heritability based on common SNPs as estimated with LDSR was 0.25 (SE=0.03) for global CT, 0.29 (SE=0.04) for global CSA and 0.30 (SE=0.03) for global CV. LDSR heritability estimates of regional CT, CSA and CV are presented in Table S25 and **Figure 4**. For regional analyses, heritability ranged from 0.05 to 0.18 for CT, from 0.07 to 0.36 for CSA and from 0.06 to 0.32 for CV. Superior temporal cortex (h^2_{CT}=0.18, h^2_{CSA}=0.30, h^2_{CV}=0.26), precuneus (h^2_{CT}=0.16, h^2_{CSA}=0.29, h^2_{CV}=0.28) and pericalcarine (h^2_{CT}=0.15, h^2_{CSA}=0.36, h^2_{CV}=0.32) are among the most genetically determined regions.

The results of partitioned heritability analyses for global and regional CT, CSA and CV with functional annotation and additionally with cell-type specific annotation are presented in Tables S26 and S27. For global CT we found enrichment for super-enhancers (enrichment=2.47, p=$3.79x10^{-6}$), introns (enrichment=1.71, p=$4.13x10^{-6}$) and histone marks (enrichment=3.09, p=$1.50x10^{-5}$). Repressors (enrichment=0.41, p=$1.06x10^{-7}$) and histone marks (enrichment=1.87, p=$5.94x10^{-6}$) were enriched for global CSA, and introns (enrichment=1.72, p=$2.77x10^{-7}$), super-enhancers (enrichment= 2.50-2.55, p=$2.81<10^{-6}$ - $4.33x10^{-7}$) and repressors (enrichment=0.49, p=$6.87x10^{-7}$). For regional CSA and CV the highest enrichment scores (>18) were observed for conserved regions.

FIGURE 4: Regional heritability estimates based on common SNPs
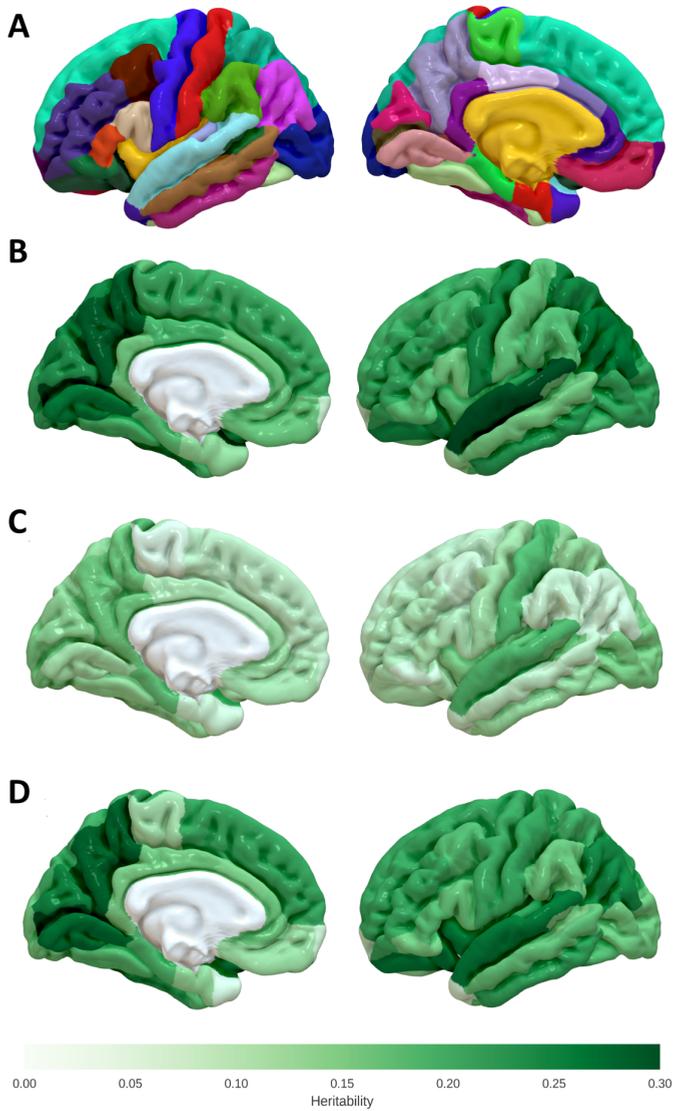
Regional heritability estimates based on common variants, calculated using LD score regression, for **(B)** cortical surface area, **(C)** cortical thickness and **(D)** cortical volume. Panel **(A)** shows the Freesurfer segmentation.

GENETIC CORRELATION

We found high genetic correlation (r_g) between global CSA and global CV (r_g=0.81, p=1.2×10^{-186}) and between global CT and global CV (r_g=0.46, p=1.4×10^{-14}), but not

between global CT and global CSA (r_g= -0.02, p=0.82). While genetic correlation between CSA and CV was strong (r_g >0.7) in most of the regions (Table S28 and **Figure 5**), it was generally weak between CSA and CT with r_g<0.3, and ranged from 0.09 to 0.69 between CT and CV. The postcentral and lingual cortex were those regions with the highest genetic correlations between CT and CV ($r_{g,\ postcentral}$= 0.69, p=2.80×10^{-33}; $r_{g,\ lingual}$= 0.59, p=7.12×10^{-22}) and between CT and CSA ($r_{g,\ postcentral}$= 0.43, p=1.60×10^{-5}; $r_{g,\ lingual}$= 0.46, p=9.82×10^{-6}).

Genetic correlation was also calculated between all 34 regions for CT. We found more significantly correlated regions for CT than for CSA and CV. For CT we observe correlation clusters within the lobes and also for cingulate regions, while for both, CSA and CV, only regions of the occipital lobe cluster together. There were also a few significant distant genetic correlations beyond the regions within the same lobe.

We computed genetic correlation between CT, CSA and CV, and biometric, neurological and psychiatric, and cerebral structural traits. We found significant genetic correlation (p<3.74×10^{-5}) between global CT, CSA and CV and TIV ($r_{g,\ CT}$=0.32, p_{CT}=4.96×10^{-6}; $r_{g,\ CSA}$=0.95, p_{CSA}=7.71×10^{-57}; $r_{g,\ CV}$=0.95, p_{CV}=4.81×10^{-70}). For global CSA and global CV we also found correlations with height ($r_{g,\ CSA}$=0.22, p_{CSA}=1.10×10^{-9}, $r_{g,\ CV}$=0.21, p_{CV}=1.06×10^{-10}), educational attainment ($r_{g,\ CSA}$=0.21, p_{CSA}=1.66×10^{-9}, $r_{g,\ CV}$=0.25, p_{CV}=1.39×10^{-16}) and cognitive performance ($r_{g,\ CSA}$=0.17, p_{CSA}=6.49×10^{-6}, $r_{g,\ CV}$=0.20, p_{CV}=4.62×10^{-10}). TIV was also correlated with CT of regions in frontal (parsopercularis, precentral), temporal (middle temporal, superior temporal) and parietal lobe (postcentral, precuneus). No significant associations were seen for neuropsychiatric traits.

FIGURE 5: Genetic correlation between CT, CSA and CV within cortical regions
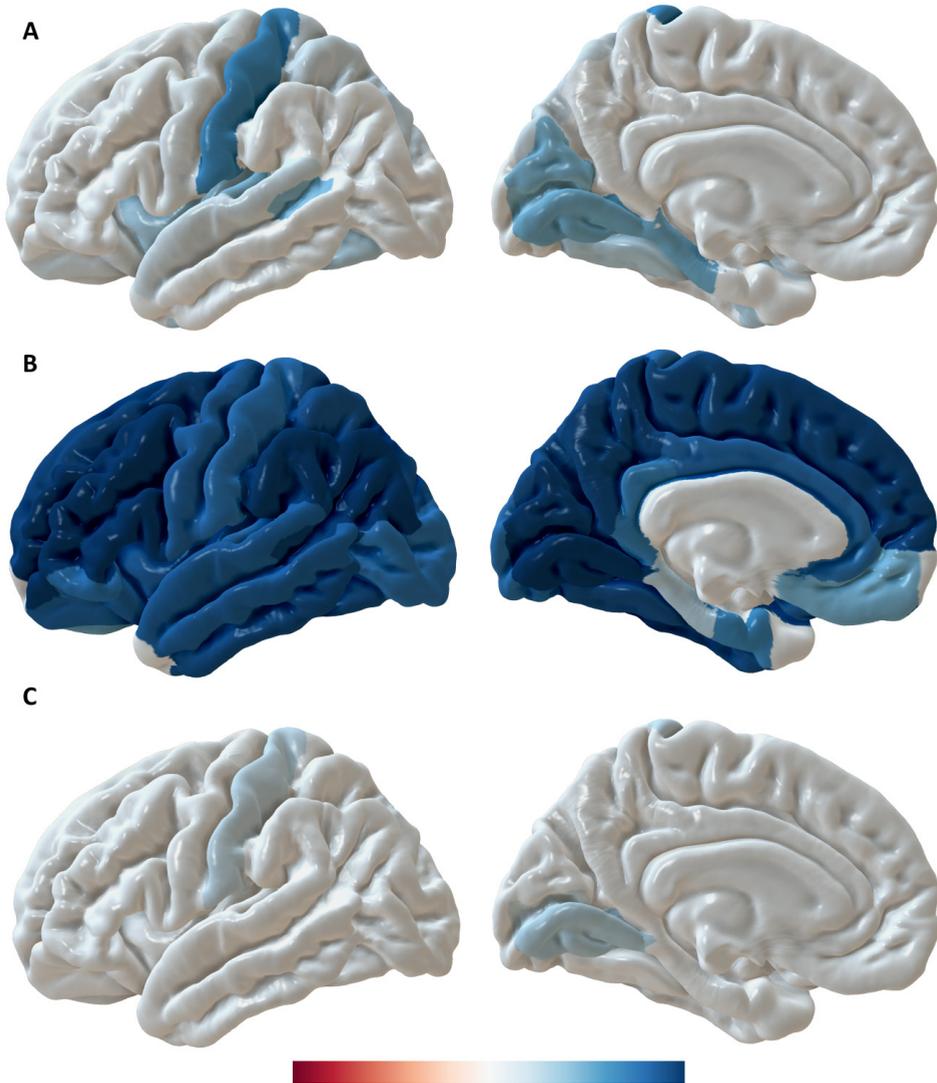
Genetic correlation between cortical thickness and volume **(A),** cortical volume and surface area **(B)**, and cortical thickness and surface area **(C)** within cortical regions.

DISCUSSION

In our genome-wide association study of up to 22822 individuals of CT, CSA and CV we identified 161 genome-wide significant associations on 19 chromosomes. These associations were found for the whole cortex and for 26 cortical regions. Heritability differed between regions and estimation methods, but was generally higher for CSA and CV than for CT.

Globally and regionally, we found strong genetic correlation between CSA and CV, and weak genetic correlation between CSA and CT. The reason for these differential associations is unclear. However, clinically silent neurodegenerative processes which lead to cortical thinning rather than to a reduction in CSA might have been present in a substantial portion of our study participants which ranged in age from 20 to 90 years and may have distorted the correlations between CT and CSA. For CT, we have seen genetic correlation between regions within all lobes, while for CSA and CV we found genetic correlation mainly between regions of the occipital lobe.

It is beyond the scope of our study to discuss each of the 161 genome-wide significant associations revealed by current analyses. However, it is evident that the wealth of information which is prepared in 35 tables creates an important resource for future investigations towards development, connectivity, function and pathology of the brain. Nonetheless six genetic regions are of particular interest. These include several loci associate with multiple brain regions at low p-values.

Among the other genetic regions that were identified in the current investigation, some were previously linked to neurological and psychiatric disorders, cognitive functioning, educational attainment, cortical development and cerebral structural traits. This indicates biological and clinical relevance of our results.

In general, as expected, we did not find pattern of strong genetic correlation between long distance cortical regions, only between that one which are close to each other. Additionally, we found that regions in occipital lobe have a stronger genetic correlation of volume, surface area and thickness between each other, compare to other lobes, which might be explained by the processes during the development of human cortex. Moreover, we showed that volume of cortical regions has a complex pattern of genetic correlation with surface area and thickness, depending on region.

In conclusion, our findings indicate that the genetic architecture of the cerebral cortex varies by region. These data create a basis for targeted functional studies to identify the biological mechanisms behind these genetic signals.

REFERENCES

1. Thambisetty, M. *et al.* Longitudinal changes in cortical thickness associated with normal aging. *NeuroImage* **52**, 1215-1223, doi:10.1016/j.neuroimage.2010.04.258 (2010).

2. Storsve, A. B. *et al.* 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**, 8488-8498, doi:10.1523/JNEUROSCI.0391-14.2014 (2014).

3. Lemaitre, H. *et al.* Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume? *Neurobiology of aging* **33**, 617 e611-619, doi:10.1016/j.neurobiolaging.2010.07.013 (2012).

4. Hogstrom, L. J., Westlye, L. T., Walhovd, K. B. & Fjell, A. M. The structure of the cerebral cortex across adult life: age-related patterns of surface area, thickness, and gyrification. *Cerebral cortex* **23**, 2521-2530, doi:10.1093/cercor/bhs231 (2013).

5. Brans, R. G. *et al.* Brain plasticity and intellectual ability are influenced by shared genes. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **30**, 5519-5524, doi:10.1523/JNEUROSCI.5841-09.2010 (2010).

6. Fjell, A. M. *et al.* Development and aging of cortical thickness correspond to genetic organization patterns. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 15462-15467, doi:10.1073/pnas.1508831112 (2015).

7. Cox, S. R. *et al.* Brain cortical characteristics of lifetime cognitive ageing. *Brain structure & function* **223**, 509-518, doi:10.1007/s00429-017-1505-0 (2018).

8. Lee, S., Habeck, C., Razlighi, Q., Salthouse, T. & Stern, Y. Selective association between cortical thickness and reference abilities in normal aging. *NeuroImage* **142**, 293-300, doi:10.1016/j.neuroimage.2016.06.041 (2016).

9. Vuoksimaa, E. *et al.* The Genetic Association Between Neocortical Volume and General Cognitive Ability Is Driven by Global Surface Area Rather Than Thickness. *Cerebral cortex* **25**, 2127-2137, doi:10.1093/cercor/bhu018 (2015).

10. Vuoksimaa, E. *et al.* Is bigger always better? The importance of cortical configuration with respect to cognitive ability. *NeuroImage* **129**, 356-366, doi:10.1016/j.neuroimage.2016.01.049 (2016).

11. Dickerson, B. C. *et al.* Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of aging* **30**, 432-440, doi:10.1016/j.neurobiolaging.2007.07.022 (2009).

12. Lerch, J. P. *et al.* Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. *Cerebral cortex* **15**, 995-1001, doi:10.1093/cercor/bhh200 (2005).

13. Querbes, O. *et al.* Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain : a journal of neurology* **132**, 2036-2047, doi:10.1093/brain/awp105 (2009).

14. Lerch, J. P. *et al.* Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiology of aging* **29**, 23-30, doi:10.1016/j.neurobiolaging.2006.09.013 (2008).

15. Schwarz, C. G. *et al.* A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *NeuroImage. Clinical* **11**, 802-812, doi:10.1016/j.nicl.2016.05.017 (2016).

16. Kalin, A. M. *et al.* Subcortical Shape Changes, Hippocampal Atrophy and Cortical Thinning in Future Alzheimer's Disease Patients. *Frontiers in aging neuroscience* **9**, 38, doi:10.3389/fnagi.2017.00038 (2017).

17. Blanc, F. *et al.* Cortical Thickness in Dementia with Lewy Bodies and Alzheimer's Disease: A Comparison of Prodromal and Dementia Stages. *PloS one* **10**, e0127396, doi:10.1371/journal.pone.0127396 (2015).

18. Uribe, C. *et al.* Patterns of cortical thinning in nondemented Parkinson's disease patients. *Movement disorders : official journal of the Movement Disorder Society* **31**, 699-708, doi:10.1002/mds.26590 (2016).

19. Hwang, K. S. *et al.* Mapping cortical atrophy in Parkinson's disease patients with dementia. *Journal of Parkinson's disease* **3**, 69-76, doi:10.3233/JPD-120151 (2013).

20. Mak, E. *et al.* Baseline and longitudinal grey matter changes in newly diagnosed Parkinson's disease: ICICLE-PD study. *Brain : a journal of neurology* **138**, 2974-2986, doi:10.1093/brain/awv211 (2015).

21. Jubault, T. *et al.* Patterns of cortical thickness and surface area in early Parkinson's disease. *NeuroImage* **55**, 462-467, doi:10.1016/j.neuroimage.2010.12.043 (2011).

22. Gerrits, N. J. *et al.* Cortical Thickness, Surface Area and Subcortical Volume Differentially Contribute to Cognitive Heterogeneity in Parkinson's Disease. *PloS one* **11**, e0148852, doi:10.1371/journal.pone.0148852 (2016).

23. Narayana, P. A. *et al.* Regional cortical thickness in relapsing remitting multiple sclerosis: A multi-center study. *NeuroImage. Clinical* **2**, 120-131, doi:10.1016/j.nicl.2012.11.009 (2012).

24. Sailer, M. *et al.* Focal thinning of the cerebral cortex in multiple sclerosis. *Brain : a journal of neurology* **126**, 1734-1744, doi:10.1093/brain/awg175 (2003).

25. Calabrese, M. *et al.* Widespread cortical thinning characterizes patients with MS with mild cognitive impairment. *Neurology* **74**, 321-328, doi:10.1212/WNL.0b013e3181cbcd03 (2010).

26. Steenwijk, M. D. *et al.* Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain : a journal of neurology* **139**, 115-126, doi:10.1093/brain/awv337 (2016).

27. Gaist, D. *et al.* Migraine with visual aura associated with thicker visual cortex. *Brain : a journal of neurology*, doi:10.1093/brain/awx382 (2018).

28. DaSilva, A. F., Granziera, C., Snyder, J. & Hadjikhani, N. Thickening in the somatosensory cortex of patients with migraine. *Neurology* **69**, 1990-1995, doi:10.1212/01.wnl.0000291618.32247.2d (2007).

29. Messina, R. *et al.* Cortical abnormalities in patients with migraine: a surface-based analysis. *Radiology* **268**, 170-180, doi:10.1148/radiol.13122004 (2013).

30. Rimol, L. M. *et al.* Cortical volume, surface area, and thickness in schizophrenia and bipolar disorder. *Biological psychiatry* **71**, 552-560, doi:10.1016/j.biopsych.2011.11.026 (2012).

31. Rimol, L. M. *et al.* Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. *Biological psychiatry* **68**, 41-50, doi:10.1016/j.biopsych.2010.03.036 (2010).

32. Hedman, A. M. *et al.* Heritability of cortical thickness changes over time in twin pairs discordant for schizophrenia. *Schizophrenia research* **173**, 192-199, doi:10.1016/j.schres.2015.06.021 (2016).

33. van Haren, N. E. *et al.* Changes in cortical thickness during the course of illness in schizophrenia. *Archives of general psychiatry* **68**, 871-880, doi:10.1001/archgenpsychiatry.2011.88 (2011).

34. Thompson, P. M. *et al.* Mapping adolescent brain change reveals dynamic wave of accelerated gray matter loss in very early-onset schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 11650-11655, doi:10.1073/pnas.201243998 (2001).

35. Hibar, D. P. *et al.* Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Molecular psychiatry* **23**, 932-942, doi:10.1038/mp.2017.73 (2018).

36. Abe, C. *et al.* Cortical thickness, volume and surface area in patients with bipolar disorder types I and II. *Journal of psychiatry & neuroscience : JPN* **41**, 240-250 (2016).

37. Hanford, L. C., Nazarov, A., Hall, G. B. & Sassi, R. B. Cortical thickness in bipolar disorder: a systematic review. *Bipolar disorders* **18**, 4-18, doi:10.1111/bdi.12362 (2016).

38. Niu, M. *et al.* Common and Specific Abnormalities in Cortical Thickness in Patients with Major Depressive and Bipolar Disorders. *EBioMedicine* **16**, 162-171, doi:10.1016/j.ebiom.2017.01.010 (2017).

39. Ecker, C. *et al.* Brain surface anatomy in adults with autism: the relationship between surface area, cortical thickness, and autistic symptoms. *JAMA psychiatry* **70**, 59-70, doi:10.1001/jamapsychiatry.2013.265 (2013).

40. van Rooij, D. *et al.* Cortical and Subcortical Brain Morphometry Differences Between Patients With Autism Spectrum Disorder and Healthy Individuals Across the Lifespan: Results From the ENIGMA ASD Working Group. *The American journal of psychiatry* **175**, 359-369, doi:10.1176/appi.ajp.2017.17010100 (2018).

41. Zielinski, B. A. *et al.* Longitudinal changes in cortical thickness in autism and typical development. *Brain : a journal of neurology* **137**, 1799-1812, doi:10.1093/brain/awu083 (2014).

42. Schmaal, L. *et al.* Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Molecular psychiatry* **22**, 900-909, doi:10.1038/mp.2016.60 (2017).

43. Han, K. M. *et al.* Cortical thickness, cortical and subcortical volume, and white matter integrity in patients with their first episode of major depression. *Journal of affective disorders* **155**, 42-48, doi:10.1016/j.jad.2013.10.021 (2014).

44. Rakic, P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends in neurosciences* **18**, 383-388 (1995).

45. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nature reviews. Neuroscience* **10**, 724-735, doi:10.1038/nrn2719 (2009).

46. Mountcastle, V. B. The columnar organization of the neocortex. *Brain : a journal of neurology* **120 (Pt 4)**, 701-722 (1997).

47. Rimol, L. M. *et al.* Cortical thickness is influenced by regionally specific genetic factors. *Biological psychiatry* **67**, 493-499, doi:10.1016/j.biopsych.2009.09.032 (2010).

48. Chen, C. H. *et al.* Hierarchical genetic organization of human cortical surface area. *Science* **335**, 1634-1636, doi:10.1126/science.1215330 (2012).

49. Eyler, L. T. *et al.* Genetic and environmental contributions to regional cortical surface area in humans: a magnetic resonance imaging twin study. *Cerebral cortex* **21**, 2313-2321, doi:10.1093/cercor/bhr013 (2011).

50. Schmitt, J. E. *et al.* Identification of genetically mediated cortical networks: a multivariate study of pediatric twins and siblings. *Cerebral cortex* **18**, 1737-1747, doi:10.1093/cercor/bhm211 (2008).

51. Panizzon, M. S. *et al.* Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral cortex*

19, 2728-2735, doi:10.1093/cercor/bhp026 (2009).

52. Winkler, A. M. *et al.* Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage* **53**, 1135-1146, doi:10.1016/j.neuroimage.2009.12.028 (2010).

53. Blokland, G. A., de Zubicaray, G. I., McMahon, K. L. & Wright, M. J. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin research and human genetics : the official journal of the International Society for Twin Studies* **15**, 351-371, doi:10.1017/thg.2012.11 (2012).

54. Ma, X. *et al.* Regional Cortical Surface Area in Adolescents: A Preliminary MRI Twin Study of Genetic and Environmental Contributions. *Behavior genetics* **46**, 205-216, doi:10.1007/s10519-015-9755-1 (2016).

55. Kremen, W. S. *et al.* Genetic and environmental influences on the size of specific brain regions in midlife: the VETSA MRI study. *NeuroImage* **49**, 1213-1223, doi:10.1016/j.neuroimage.2009.09.043 (2010).

56. Chouinard-Decorte, F. *et al.* Heritable changes in regional cortical thickness with age. *Brain imaging and behavior* **8**, 208-216, doi:10.1007/s11682-014-9296-x (2014).

57. Rogers, J. *et al.* On the genetic architecture of cortical folding and brain volume in primates. *NeuroImage* **53**, 1103-1108, doi:10.1016/j.neuroimage.2010.02.020 (2010).

58. Joshi, A. A. *et al.* The contribution of genes to cortical thickness and volume. *Neuroreport* **22**, 101-105, doi:10.1097/WNR.0b013e3283424c84 (2011).

59. Lenroot, R. K. *et al.* Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. *Human brain mapping* **30**, 163-174, doi:10.1002/hbm.20494 (2009).

60. Wen, W. *et al.* Distinct Genetic Influences on Cortical and Subcortical Brain Structures. *Scientific reports* **6**, 32760, doi:10.1038/srep32760 (2016).

61. Lu, L. *et al.* Common and rare variants of the THBS1 gene associated with the risk for autism. *Psychiatric genetics* **24**, 235-240, doi:10.1097/YPG.0000000000000054 (2014).

62. Park, H. J., Kim, S. K., Kim, J. W., Kang, W. S. & Chung, J. H. Association of thrombospondin 1 gene with schizophrenia in Korean population. *Molecular biology reports* **39**, 6875-6880, doi:10.1007/s11033-012-1513-3 (2012).

63. Christopherson, K. S. *et al.* Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. *Cell* **120**, 421-433, doi:10.1016/j.cell.2004.12.020 (2005).

64. Blake, S. M. *et al.* Thrombospondin-1 binds to ApoER2 and VLDL receptor and functions in postnatal neuronal migration. *The EMBO journal* **27**, 3069-3080, doi:10.1038/emboj.2008.223 (2008).

65. Ikram, M. A. *et al.* Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nature genetics* **44**, 539-544, doi:10.1038/ng.2245 (2012).

66. Adams, H. H. *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nature neuroscience* **19**, 1569-1582, doi:10.1038/nn.4398 (2016).

67. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun* **9**, 2098, doi:10.1038/s41467-018-04362-x (2018).

68. Luciano, M. *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing

neuroticism. *Nature genetics* **50**, 6-11, doi:10.1038/s41588-017-0013-8 (2018).

69.   Smith, D. J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Molecular psychiatry* **21**, 749-757, doi:10.1038/mp.2016.49 (2016).

70.   Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nature genetics* **49**, 1511-1516, doi:10.1038/ng.3955 (2017).

71.   Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics* **48**, 709-717, doi:10.1038/ng.3570 (2016).

72.   International Parkinson Disease Genomics, C. *et al.* Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641-649, doi:10.1016/S0140-6736(10)62345-8 (2011).

73.   Simon-Sanchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature genetics* **41**, 1308-1312, doi:10.1038/ng.487 (2009).

74.   Pankratz, N. *et al.* Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Annals of neurology* **71**, 370-384, doi:10.1002/ana.22687 (2012).

75.   Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224-229, doi:10.1038/nature14101 (2015).

76.   Pountney, D. L., Raftery, M. J., Chegini, F., Blumbergs, P. C. & Gai, W. P. NSF, Unc-18-1, dynamin-1 and HSP90 are inclusion body components in neuronal intranuclear inclusion disease identified by anti-SUMO-1-immunocapture. *Acta neuropathologica* **116**, 603-614, doi:10.1007/s00401-008-0437-4 (2008).

77.   Muller, M. B. & Wurst, W. Getting closer to affective disorders: the role of CRH receptor systems. *Trends in molecular medicine* **10**, 409-415, doi:10.1016/j.molmed.2004.06.007 (2004).

78.   Desikan, R. S. *et al.* Genetic overlap between Alzheimer's disease and Parkinson's disease at the MAPT locus. *Molecular psychiatry* **20**, 1588-1595, doi:10.1038/mp.2015.6 (2015).

79.   Spillantini, M. G. & Goedert, M. Tau pathology and neurodegeneration. *The Lancet. Neurology* **12**, 609-622, doi:10.1016/S1474-4422(13)70090-5 (2013).

80.   Grinberg, I. *et al.* Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nature genetics* **36**, 1053-1055, doi:10.1038/ng1420 (2004).

81.   Umemura, K. *et al.* Autotaxin expression is enhanced in frontal cortex of Alzheimer-type dementia patients. *Neuroscience letters* **400**, 97-100, doi:10.1016/j.neulet.2006.02.008 (2006).

82.   Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nature reviews. Genetics* **14**, 507-515, doi:10.1038/nrg3457 (2013).

83.   Chen, C. H. *et al.* Genetic influences on cortical regionalization in the human brain. *Neuron* **72**, 537-544, doi:10.1016/j.neuron.2011.08.021 (2011).

84.   Psaty, B. M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circulation. Cardiovascular genetics* **2**, 73-80, doi:10.1161/CIRCGENETICS.108.829747 (2009).

85.   Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**, e1001779, doi:10.1371/journal.pmed.1001779 (2015).

86.   Bearden, C. E. & Thompson, P. M. Emerging Global Initiatives in Neurogenetics: The Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA) Consortium. *Neuron* **94**, 232-236, doi:10.1016/j.neuron.2017.03.033 (2017).

87.   Clarke, L. *et al.* The 1000 Genomes Project: data management and community access. *Nature methods* **9**, 459-462, doi:10.1038/nmeth.1974 (2012).

88.   Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968-980, doi:10.1016/j.neuroimage.2006.01.021 (2006).

89.   Fischl, B. *et al.* Automatically parcellating the human cerebral cortex. *Cerebral cortex* **14**, 11-22 (2004).

90.   Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nature protocols* **9**, 1192-1212, doi:10.1038/nprot.2014.071 (2014).

91.   Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).

92.   Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

93.   Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164, doi:10.1093/nar/gkq603 (2010).

94.   Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature communications* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).

95.   Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337, doi:10.1093/bioinformatics/btq419 (2010).

96.   de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).

97.   MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research* **45**, D896-D901, doi:10.1093/nar/gkw1133 (2017).

98.   Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Current Alzheimer research* **9**, 628-645 (2012).

99.   Bennett, D. A. *et al.* Overview and findings from the rush Memory and Aging Project. *Current Alzheimer research* **9**, 646-663 (2012).

100.   Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *American journal of human genetics* **62**, 1198-1211, doi:10.1086/301844 (1998).

101.   Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).

102.   Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* **42**, 565-569, doi:10.1038/ng.608 (2010).

103.   Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).

104.   Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature genetics* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).

CHAPTER 5.3

Voxel-wise genome
wide-association study
of gray matter

ABSTRACT

Objective

The human brain has an intricate structure that is partly determined by genetics. While neuroimaging can discern millions of morphometric features, genome-wide association studies have only focused on a limited number of aggregate measures.

Methods

The structure of the cortical and subcortical grey matter was characterized using voxel-based morphometry, a technique that generates a volume for each of the 1.5 million voxels. Subsequently, genome-wide association studies were performed for all these voxels using a newly developed software algorithm in over 18.000 individuals from 12 cohorts within the CHARGE consortium and UK Biobank participants. The genome-wide brain-wide significance threshold was determined at $p = 5 \times 10^{-13}$.

Results

The association analyses between 9 million genetic variants and 1.5 million brain voxels took 17 hours to compute, which would require over 2 years using conventional software. We found 1.270.907 genome-wide brain-wide significant associations between genetic variants and grey matter voxels, distributed between 55 brain regions with minimum p-value$=2*10^{-87}$. These associations came from 930 variants that mapped to 29 independent loci. We found 62.920 unique voxels significantly associated at least with one variant.

Conclusions

The genome-wide interrogation of high-resolution neuroimaging data is a powerful approach to identify genetic determinants of brain structure, even at sample sizes considerably smaller than studies of aggregate measures This is just a first attempt for using high-dimensional neuroimaging features for genetic association analysis, which showed that such measures are very promising endophenotypes and can reveal us more about complex brain genetic architecture.

INTRODUCTION

The structure of the brain is at the foundation of unravelling the neural basis of human behavior. Differences in its anatomical configuration can explain how the brain affects various functions, including learning ability,[1] cognitive performance,[2] and even political orientation.[3] Structural abnormalities are also a common feature of neurological and psychiatric diseases,[4-6] and provide insight into their pathophysiology.

Among the structural features that are genetically regulated and have implications for cortical function is the distribution of gray matter across the cortex. This varies widely across normal individuals, with developmental waves of gray matter gain and loss subsiding by adulthood, and complex deficit patterns observed in Alzheimer's disease, schizophrenia, and healthy subjects at genetic risk for these disorders.

Neuroimaging studies, mainly using magnetic resonance imaging (MRI), primarily focused on structural measures that were aggregated across the whole brain or particular regions of interest. Advances in image acquisition and downstream processing, however, have enabled the quantification of brain structure in a way that captures more of its complexity on high-dimensional level. One such approach is voxel-based morphometry (VBM)[7], which calculates the concentration of brain tissue at the most fine-grained level of an image, namely the voxel. Studies employing VBM have been able to map structural brain differences that are more localized and not necessary follow anatomically defined brain regions.

Genetics plays an important role in determining brain structure and twin studies have found a high degree of heritability across the brain.[8] This was shown for volumes of large regions of interest, but also for the refined brain maps using VBM.[9] Furthermore, the amount of shared heritability between different cortical regions decreases with distance indicating that there is regional specificity in their genetic makeup. While this is known, gene discovery, which is now commonly done using genome-wide association studies (GWAS), has focused on the aggregate measures of regions of interest, mostly volumes. Data sharing and computational issues made it imposable, until recently, to perform GWAS on high-dimensional neuroimaging phenotypes.

Here we used a newly developed software tool[10] to perform a comprehensive genome-wide and brain-wide scan in over 18.000 individuals from 12 cohorts. VBM was used to calculate grey matter densities of 1.5 million voxels and subsequently perform GWAS for each of these.

METHODS

IMAGE PROCESSING

In cohorts voxel-based morphometry was performed with an optimized protocol[11] using the FSL software.[12] All grey matter density maps were non-linearly registered to a the ICBM MNI152 template (Montreal Neurological Institute). The MNI152 standard-space T1-weighted average structural template has a 1x1x1 mm^3 voxel resolution and was derived from 152 structural images, which have been warped and averaged into the common MNI152 co-ordinate system after high-dimensional nonlinear registration. To avoid effects of the registration step on the grey matter we implemented a spatial modulation procedure by multiplying voxel densities with the Jacobian determinants estimated during spatial normalization. Finally, images were smoothed using an isotropic Gaussian kernel of 3mm (FWHM 8mm).

VOXEL-WISE GENOME-WIDE ASSOCIATION ANALYSIS

We investigated the association between all voxels inside gray matter mask and all SNPs which passed the quality control. Regression model included age and sex as covariates. All cohorts, first, computed high-dimensional partial derivatives and encoded their phenotypes and genotypes according to HASE framework protocol (https://github.com/roshchupkin/hase/wiki/HD-Meta-analysis-GWAS). After exchanging these data with central side, the final step of pooled meta-analysis was performed. Due to enormous amount of results data to store, we set the threshold to save only SNP-voxel association with p-value below 5*10**-8. The analysis was performed on cluster using 100 cores with 64 RAM each.

PERMUTATIONS

Based on voxel-based permutations, we identified 100.000 independent tests for VBM analysis[13], which yield to $5*10^{-8}/10^5 = 5*10^{-13}$ threshold for voxel-wide genome-wide significance.

ANNOTATION

Annotation of genome-wide significant variants was performed using the ANNOVAR software package[14]. We used the clumping function in plink (0.2 LD threshold, 500Kb distance) to identify the most significant SNP in each linkage disequilibrium (LD) block.

Additionally, we searched the NHGRI-EBI Catalog of published GWAS[15] for previous SNP-trait associations of genome-wide significant variants.

RESULTS

Using the strict genome-wide voxel-wide p-value threshold for significance we found 1.270.907 genome-wide brain-wide significant associations between genetic variants and grey matter voxels, distributed between 55 brain regions with minimum p-value$=2*10^{-87}$ (**Table 1**). These associations came from 930 variants that mapped to 29 independent loci. We found 62.920 unique voxels significantly associated at least with one variant. As you can see from **Table 1** many loci associated with voxels which below to more than one brain region. The association map for the top two variant is shown on **Figure 1.** From 29 independent loci nine are located in intronic regions and one in exome.

The distribution of significant voxels between different brain regions is in **Supplementary Table 1**. The most of all significant voxels were found in precentral, postcentral gyruses and superior parietal gyrus. We found significant associations for all seven subcortical structures. However, as you can see from **Supplementary Table 1,** the signal can vary quite substantial between left and right hemisphere.
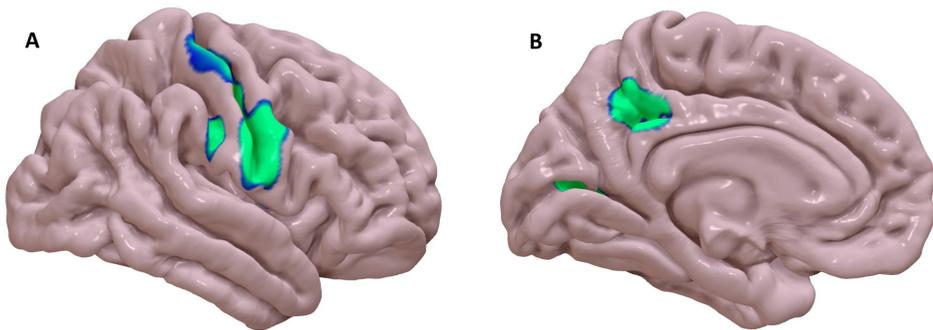

FIGURE 1: Brain map of significant voxel-based association. **A –** association map of rs1080066; **B –** association map of rs76341705;

TABLE 1: Voxel-wide genome-wide significant loci.

RSID	CHR	bp	A1	A2	Annotation	type	MAF	Min p-value	Number regions	Number of voxels
rs1080066	15	39634222	G	A	C15orf54(dist=87179),THBS1(dist=239058)	intergenic	0.91	2.05E-87	12	39390
rs76341705	14	59628679	A	G	LINC01500(dist=144633),DAAM1(dist=26702)	intergenic	0.87	2.36E-29	13	11538
rs945270	14	56200473	G	C	KTN1(dist=49171),RPL13AP3(dist=32490)	intergenic	0.57	1.90E-22	4	3337
rs62002283	15	39595347	G	A	C15orf54(dist=48304),THBS1(dist=277933)	intergenic	0.88	1.75E-21	5	833
rs17179798	12	5314508	A	G	KCNA5(dist=158554),LOC101929584(dist=286654)	intergenic	0.78	4.60E-21	5	656
rs3845778	2	37133127	T	C	STRN	intronic	0.52	3.77E-19	3	1297
rs11614730	12	79925802	A	G	MIR5692B(dist=73362),PAWR(dist=59943)	intergenic	0.31	4.78E-18	2	42
rs62057122	17	43901001	T	C	CRHR1,MGC57346-CRHR1	intronic	0.79	1.02E-17	5	287
rs11070185	15	39625618	C	T	C15orf54(dist=78575),THBS1(dist=247662)	intergenic	0.21	1.53E-17	5	933
rs1822105	15	39634303	C	T	C15orf54(dist=87260),THBS1(dist=238977)	intergenic	0.58	3.31E-17	4	1124
rs7776727	7	96199681	T	C	C7orf76(dist=66846),LOC100506136(dist=51288)	intergenic	0.32	7.66E-17	4	705
rs17698176	17	44819595	G	T	NSF	intronic	0.78	1.15E-16	4	531
rs1887163	11	92011168	C	G	DISC1FP1(dist=1362948),FAT3(dist=74094)	intergenic	0.41	1.74E-16	2	627
rs199448	17	44809001	G	A	NSF	intronic	0.78	8.64E-16	4	507
rs113935260	14	59473028	C	T	LINC01500	ncRNA intronic	0.92	5.71E-15	3	220
rs4817506	21	34348569	A	G	C21orf62(dist=162516),OLIG2(dist=49647)	intergenic	0.50	7.27E-15	3	42
rs73123652	12	65874956	C	T	LOC100507065	ncRNA intronic	0.88	1.10E-14	1	115
rs4887907	16	77466321	T	G	ADAMTS18	intronic	0.41	1.59E-14	2	41
rs13107325	4	103188709	T	C	SLC39A8	exonic	0.93	1.59E-14	3	42
rs56023709	16	87229344	C	A	LOC440390(dist=131747),LOC101928708(dist=16377)	intergenic	0.39	1.82E-14	2	112
rs888278	2	65999623	T	G	SPRED2(dist=339967),MIR4778(dist=585758)	intergenic	0.36	2.39E-14	2	16
rs747286	12	5276878	T	C	KCNA5(dist=120924),LOC101929584(dist=66284)	intergenic	0.56	2.42E-14	2	69
rs4774220	15	55962861	A	G	PRTG	intronic	0.75	3.43E-14	3	38
rs160459	14	59074136	C	A	KIAA0586(dist=58587),NONE(dist=NONE)	intergenic	0.54	4.23E-14	1	123
rs17616633	16	51451683	C	T	SALL1(dist=266500),LINC01571(dist=344747)	intergenic	0.43	5.26E-14	1	38
rs4809942	20	52463522	A	T	ZNF217(dist=263886),SUMO1P1(dist=27518)	intergenic	0.48	8.98E-14	1	15
rs181288	12	5317850	C	T	KCNA5(dist=161896),LOC101929584(dist=25312)	intergenic	0.31	1.30E-13	3	18
rs2619062	12	78288811	C	T	NAV3	intronic	0.57	1.75E-13	1	8
rs73080976	3	50187326	A	G	RBM5(dist=30929),SEMA3F(dist=5522)	intergenic	0.92	3.30E-13	1	4

DISCUSSION

Using a comprehensive approach to test associations between millions of genetic variants and millions of grey matter densities, we identified 29 loci influencing brain structure. Many of found loci were previously reported to be associated with other traits and disorders: Parkinson, obesity, intelligence, neuroticism, panic, blood pressure, autism, and schizophrenia. Our results highlight the feasibility and power of such studies and pave the way for larger efforts in the future.

Brain genetic consortia CHARGE and ENIGMA found genome-wide significant associations when studying the volume of specific brain regions, such as the hippocampal formation and putamen.[16,17]. However, more fine-grained phenotypes might be better for gene identification, since the brain is composed of many structures with a different genetic background. Regions of interest, such as the hippocampus, putamen or thalamus, themselves consist of multiple sub regions or nucleus that have specialized functions. It can be reasonably argued that functionally distinct components may also have a partly distinct genetic background. Neuroimaging measures that can more accurately depict these might thus represent better phenotypes for genetic studies. Our voxel-wide genome-wide study provides further support for this notion. Stein *et al.* previously reported a voxel-wise GWAS study.[18] There, the authors showed that such studies are feasible by performing association analyses between 32,000 voxels and 450,000 genetic variants. That study was done in a small sample of 740 individuals as a proof-of-principle, but found no variants withstanding their multiple testing correction. Here, we aimed to do gene discovery in over 18.000 individuals, using more genetic variants and voxels, and identified 29 loci at a much stricter significance threshold.

High-dimensional association approach can also be applied to other neuroimaging measures: cortical thickness, pial and white matter surface areas, local gyrification patterns, the shape of subcortical structures, and white matter microstructure as measured with diffusion imaging, are among some of the most prominent. Such studies, however, do not need to be restricted to neuroimaging and genetics per se, but are also of particular interest to other fields. It can be applied to MRI or other type of images of different organ in combination with different type of omics data. Increasing number of studies have linked transcriptomics, metabolomics or methylation data to human diseases. Systematic cross-investigations between such large-scale datasets remain to be performed, and could potentially yield a trove of novel insights into disease.

In conclusion, we perform a comprehensive genome-wide and brain-wide scan in identified the genetic variants associated with different brain structures. Our results demonstrate that

such high-dimensional neuroimaging analysis is currently feasible and can localize genetic effects to specific sub-regions as opposed to using aggregate measures. Studying the relation between genetics and brain anatomy, at the higher resolution, would likely benefit researches aimed to advance our understanding of the brain and of the underlying neurobiology of complex disorders.

REFERENCES

1. Zatorre, R. J., Fields, R. D. & Johansen-Berg, H. Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature neuroscience* **15**, 528-536 (2012).

2. Raz, N. Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. (2000).

3. Kanai, R., Feilden, T., Firth, C. & Rees, G. Political orientations are correlated with brain structure in young adults. *Current biology* **21**, 677-680 (2011).

4. Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L. & Greicius, M. D. Neurodegenerative diseases target large-scale human brain networks. *Neuron* **62**, 42-52 (2009).

5. Soares, J. C. & Mann, J. J. The anatomy of mood disorders—review of structural neuroimaging studies. *Biological psychiatry* **41**, 86-106 (1997).

6. McDonald, C. *et al.* Regional volume deviations of brain structure in schizophrenia and psychotic bipolar disorder. *The British Journal of Psychiatry* **186**, 369-377 (2005).

7. Ashburner, J. & Friston, K. J. Voxel-based morphometry—the methods. *Neuroimage* **11**, 805-821 (2000).

8. Peper, J. S. *et al.* Genetic influences on human brain structure: a review of brain imaging studies in twins. *Human brain mapping* **28**, 464-473 (2007).

9. Thompson, P. M. *et al.* Genetic influences on brain structure. *Nature neuroscience* **4**, 1253-1258 (2001).

10. Roshchupkin, G. V. *et al.* HASE: Framework for efficient high-dimensional association analyses. *Sci Rep* **6**, 36076, doi:10.1038/srep36076 (2016).

11. Good, C. D. *et al.* in *Biomedical Imaging, 2002. 5th IEEE EMBS International Summer School on.* 16 pp. (IEEE).

12. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208-S219 (2004).

13. Roshchupkin, G. V. *et al.* Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology. *Neurobiol Aging* **48**, 204-211, doi:10.1016/j.neurobiolaging.2016.08.024 (2016).

14. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164, doi:10.1093/nar/gkq603 (2010).

15. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research* **45**, D896-D901, doi:10.1093/nar/gkw1133 (2017).

16. Bis, J. C. *et al.* Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* **44**, 545-551, doi:10.1038/ng.2237 (2012).

17. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* (2015).

18. Stein, J. L. *et al.* Voxelwise genome-wide association study (vGWAS). *neuroimage* **53**, 1160-1174 (2010).

SUPPLEMENTARY FILES

SUPPLEMENTARY TABLE 1: Number of unique significant voxels per different brain regions in left and right hemisphere.

Brain Region	Number of unique significant voxels	Region size (#voxels)
Postcentral gyrus right	10942	42995
Precentral gyrus left	9461	52486
Precentral gyrus right	9081	53659
Postcentral gyrus left	8587	46092
Superior parietal gyrus left	3481	63572
Superior parietal gyrus right	3040	67336
Cingulate gyrus posterior part left	2084	11796
Putamen left	1641	7438
Remainder of parietal lobe right	1618	66178
Cingulate gyrus posterior part right	1608	12439
Putamen right	1399	7064
Cuneus left	1156	14454
Middle frontal gyrus right	726	81880
Caudate nucleus right	725	6170
Lingual gyrus right	668	18495
Remainder of parietal lobe left	668	65794
Lingual gyrus left	635	18132
Insula right	611	22398
Cuneus right	542	13755
Gyri parahippocampalis et ambiens right	489	6944
Lateral remainder of occipital lobe left	449	64895
Gyri parahippocampalis et ambiens left	426	6823
Medial and inferior temporal gyri right	402	28066
Lateral remainder of occipital lobe right	307	66957
Posterior temporal lobe right	305	72577
Insula left	288	21930

SUPPLEMENTARY TABLE 1 CONTINUED.

Brain Region	Number of unique significant voxels	Region size (#voxels)
Lateral occipitotemporal gyrus (gyrus fusiformis) right	242	6478
Posterior temporal lobe left	220	70660
Pallidum (globus pallidus) left	195	1895
Pallidum (globus pallidus) right	156	1940
Hippocampus left	119	3183
Superior temporal gyrus, central part right	99	21840
Straight gyrus (gyrus rectus) right	53	6520
Amygdala left	40	2192
Nucleus accumbens left	37	504
Inferior frontal gyrus right	33	27123
Medial and inferior temporal gyri left	31	27036
Thalamus left	25	10524
Medial orbital gyrus left	23	9311
Straight gyrus (gyrus rectus) left	23	5306
Subgenual anterior cingulate gyrus right	20	2192
Lateral occipitotemporal gyrus (gyrus fusiformis) left	19	6430
Anterior orbital gyrus right	12	9787
Middle frontal gyrus left	11	80119
Superior frontal gyrus right	11	84342
Superior temporal gyrus, central part left	11	20954
Caudate nucleus left	4	6059
Lateral orbital gyrus right	4	5895
Medial orbital gyrus right	4	9243
Hippocampus right	3	3504
Subgenual anterior cingulate gyrus left	2	2095
Nucleus accumbens right	1	384
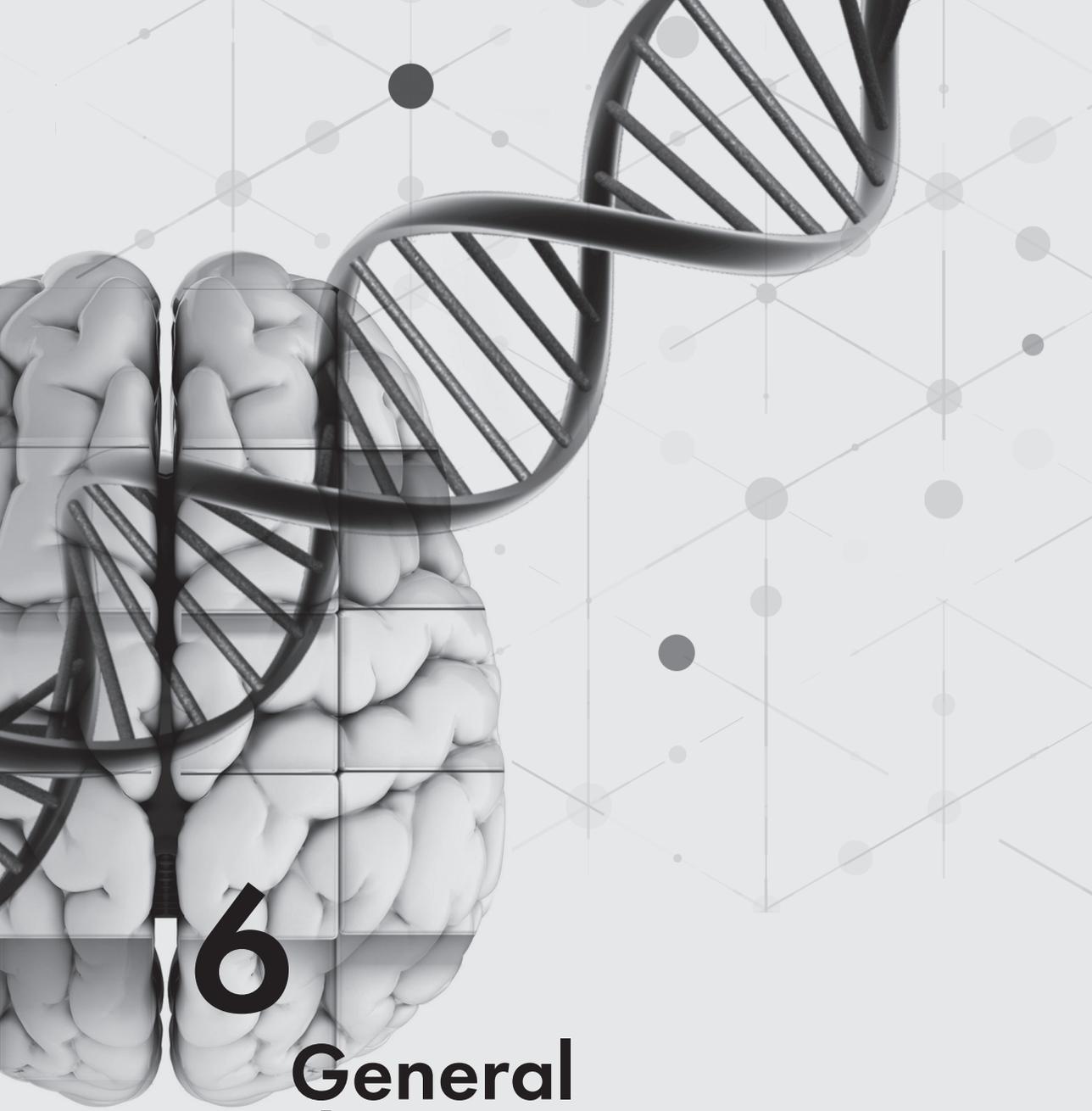
6

General
discussion

DISCUSSION

In this thesis, I showed that by combining data and knowledge from genetics and imaging it is possible to gain more inside into the complex genetic architecture of our brain. I also solved several methodological issues and developed methods for the integrated analysis of imaging and genetic data, which opens new ways for exploring the relation between these data. In this chapter, I summarize my contributions to the field and discuss them. Given that imaging genetics is still a novel area of research, many challenges still lie ahead. Therefore I also devoted a large part of this chapter to possible future research directions.

METHODOLOGICAL BARRIERS

Several chapters of the work presented in this dissertation (**4.1, 5.2, and 5.3**) became possible as a result of the original idea of voxel-wise genome-wide analysis. At the beginning of my PhD (March 2014), there were no options to perform such analyses due to computational requirements of thousand years for individual sites. Additionally, the sheer size of intermediate results was not suitable for multicenter settings where summary statistics would need to be exchanged for meta-analysis. I therefore developed algorithms and a framework to overcome both issues. **In chapter 4.1** I showed that the co-called **HASE** framework allows for reducing the computational time to just hours and limiting intermediate summary data from TBs to GBs. The framework was developed to be user-friendly for multicenter studies, and to employ it as part of GWAS pipelines in different analysis settings. In chapter 5.2 the framework was employed in a cortical brain regions GWAS study in combination with classical meta-analysis workflow, just to speed up a single side analysis. In chapter 5.3 the full functionality was used in a voxel-wise GWAS, and all findings in this chapter could not have been obtained without **HASE**. Having started as a tool to address a quite specific research problem, our **HASE** framework has now became a broadly used instrument for analysis genetics, gene expression and methylation data in many ongoing projects. We have published the framework online as open source software, and expect that many projects are currently using it and that the number of applications for which it will be used will grow significantly. We are planning to maintain this framework and add additional functionality to support the large scale exploration of omics data in both single- and multicenter studies.

In **chapter 4.2** I adressed another fundamental problem in the genetics field, pleiotropy, which relates to the fact that genes may associate with more than one phenotype. Currently, around 43% of genes have been reported in GWAS catalogs to associate with more than one phenotype[1], suggesting that the pleiotropy might be rather a common phenomenon, than something rare. There are several approaches to investigate the genetic correlation between phenotypes. However, identifying a single pleiotropy variant has always proven to

be challenging[2]. Existing methods to study pleiotropy are lacking or suffer from a bias toward the most statistically powered phenotype. In this thesis I developed a solution to address this issue, taking a combinatorial approach, based on sum ranks of SNPs statistics. This approach directly incorporates information from GWAS summaries and tests the pleiotropy hypothesis. Compared to others tests our approach is robust to extreme p-values and only finds significant pleiotropy when a signal of association is present for all traits. Moreover, the method is able to identify single loci pleiotropy, even when there is no significant genetic correlation. This property made it possible to find important genetic variants which affect several traits and disorders (**chapter 4.2, 5.1, 5.3**). An important finding in this thesis, and for the imaging genetics field, is that using this method we found significant pleiotropic loci between subcortical structures and schizophrenia (**chapter 4.2**). In a previous proof-of-concept study[3], using other techniques, no evidence was found for such a relationship, neither at a high level nor for single genetic markers. This raised considerable debate in the research community about the importance of imaging GWAS discoveries for understanding disease etiology since no genetic overlap was found. Importantly, the latest subcortical GWAS[4] with larger sample size also identified significant SNPs which associated with schizophrenia. Our method was able to obtain this result based on summary statistics on a smaller dataset. Additionally, we found variants not yet significant for each trait independently. Finally, when applying this method for new GWAS results of the commissure tract (**chapter 5.1**), we found that regardless of the underlying mechanism through which genes affect the anterior commissure, there seems to be pleiotropy with various neurodegenerative diseases (Alzheimer 's, Parkinson, frontotemporal dementia). The results reported in this thesis provide only a few examples, but most probably pleiotropy between neuroimaging biomarkers and neurodegenerative diseases is more common. Taken together this demonstrates the potential of the imaging genetics field to significantly contribute to a better understanding of complex disease genetic architecture.

An additional advantage of our method is that it is not limited to study pleiotropy between two phenotypes, but it is also possible to apply it to several traits. We demonstrated this by investigating pleiotropy between five psychiatric disorders in **chapter 4.2**. Clinical observations show varying degrees of symptoms overlap between these diseases[5], suggesting the existence of a (partially) shared genetic architecture. Additionally, the strong genetic correlation between pairs of these disorders has been shown using GWAS summary statistics[6]. However, to discover specific loci remained challenging. Using our method we were able to find pleiotropic genes, not only between pairs, like *RERE, CACNA1C* for bipolar/schizophrenia but also between three or four diseases, like *ITIH3, SFMBT1* genes. Our findings can help with better understanding of complex etiology of psychiatric diseases and provide candidate genes for drug development.

The existence of a strong genetic correlation between many phenotypes and disorders sug-

gests that multi-phenotype pleiotropy is a common phenomenon and as a result the method we developed has the potential to be applied widely.

HIGH-DIMENSIONAL ENDOPHENOTYPES

Several chapters of this thesis describe the analysis of high-dimensional (HD) imaging phenotypes, which can potentially serve as better endophenotypes for genetic discoveries than using simple image-derived measures. While big data analytics approaches have facilitated scientific discoveries in fields such as genomics and imaging, cross-investigations between multiple big datasets has remained impractical. The goal of such analyses would be that by increasing data dimensionality it would be possible to more precisely study the complex relation between omics and neuroimaging data. Besides the biological relevance of such research, in this thesis we aimed to overcome fears of using HD imaging phenotypes in genetic studies. In **chapters 2.1, 5.2,** even with quite a complex pipeline for data processing and in a multicenter settings with different scanners, field strength, and acquisition protocols, we demonstrated that it is possible to develop a framework for HD data harmonization, and successfully performed imaging genetics analyses with HD imaging phenotypes. Applying this methodology in the context of population and family-based studies we showed that local gray matter density, represented on voxel level, is significantly heritable and varies widely. The clusters of highly heritable voxels are located in subcortical as well in cortical regions, which overlap with regions of high reproducibility. However, there are also regions with high reproducibility which are not heritable at all. These results implie that the voxels heritability follows some pattern due to the complexity of brain genetic architecture, not measurement bias or error. To support interpretation and visualization of the results of our work we made a special online portal: http://imagene.nl/heritability. Such online portals and tools in omics and neuroimaging analysis are becoming increasingly adopted[6-8] to provide the way for the research community to explore the results and data.

For subcortical shapes we observed the same complex patterns of genetic effects in **chapter 2.2**. Moreover, our analysis revealed that the genetic architecture of subcortical shapes goes beyond just volumetric measures and serves as a complimentary endophenotype. In our voxel based genetic analyses we showed that using the most heritable voxel can reduce the signal to noise ratio and gain more power in association analysis, compare to just gross measurements. For the shapes we demonstrated that exclusion of such HD endophenotype from the analysis would lead to missing a substantial part of information about the genetic architecture of subcortical structures.

In **chapter 3.1** we demonstrated that although detecting significant genetic effects on the individual voxel requires large sample sizes, using such three-dimensional association maps

in combination with gene expression information may help to gain additional insight on how genes affect brain structure. We found that VBM association patterns overlap with some of the previously identified genes (CLU, SLC24A4, and MEF2C). We also showed that VBM analysis combined with expression data could provide evidence for new candidate genes in genetic loci, where the causal gene has not been strongly established by biological experiments[9]. In **chapter 3.2,** I showed that such high-dimensional imaging phenotypes can not only be successfully applied to investigate the relation of brain anatomy at the voxel level with genetic data, but also with cognitive measures. Several clusters of voxels were found to be significantly associated with global cognition. Interestingly, each of these clusters was located within multiple anatomic regions, confirming that complex functions of human cognition are not accurately represented by arbitrarily defined anatomical brain regions.

There has been considerable, and very reasonable, criticisms related to the issue of decreasing statistical power by increasing the number of tests when studying HD phenotypes[10]. Therefore imaging genetics studies have typically been limited to one or a few phenotypes/regions of interest (ROI) based on a prior knowledge[11-13]. However, in **chapter 5.2** we showed that such criticism must be taken with care and sufficient understanding of the origin of the HD phenotype. For gray matter density at the voxel level, we pointed out two important reasons why a HD approach may makes more sense. First, segmentation of biologically defined brain structures may aggregate too much functional information; therefore the more localized genetic effect on substructure may vanish when using an ROI, leading to less statistical power. Second, the effect may be distributed between several brain regions, and in such a case the analysis on preselected structures would provide uncomplete information about genetic influence. Furthermore, our results showed that, due to HD nature of voxels, the genetic effect size on the single voxel is much bigger compared to classical non-HD phenotypes. Thus, the increased significance of the association readily overcomes the stricter correction threshold that needs to be applied. The top found genetic variant in our study had a p-value of 10^{-87} which is far beyond any thresholds and explains about 3% of voxel density variation.

FUTURE DIRECTIONS

The methodology and research results described in this thesis only address a small part of the questions and challenges which lie ahead for the imaging genetics field. Neuroscience and genomics are rapidly developing and an enormous amount of data and knowledge is accumulating. This opens unique opportunities for scientists to address questions which have challenged human conscience since ancient times. In this section I will discuss possible future directions, opportunities, and how I think different problems can be approached.

NEW IMAGING MEASURES

In imaging genetics, independent improvement in both these fields is needed to accelerate the progress in our understanding of the genetic architecture of the human brain and its relation to various complex diseases. Neuroimaging can provide a rich variety of phenotypes for genetic analysis. However, most of the imaging genetics research which has been done so far has focused on structural data such as volume, cortical and shape thickness, surface area or voxels. While these are very relevant descriptors of brain anatomy, our brain is a much more complex system and a richer set of features will be required. White matter is composed of bundles, which connect various gray matter areas of the brain to each other and carry nerve impulses between neurons. Changes in white matter has been shown to relate to various neurodegenerative[14,15] and psychiatric diseases[16], stroke[17], and cognitive performance[18]. Evidently, the disruptions or anomalies in connections between brain regions play an important role in overall wellbeing. Therefore, genetic influences on these white matter connectivity measures are important to identify. Genetic studies of such structural connectivity using diffusion MRI commonly fall into three categories[19]: diffusion properties of white matter based on fractional anisotropy (FA) or mean diffusivity (MD); 3D model of tracts or fibers extracted using tractography[20]; a network of brain connections, shown as connectivity matrix of graphs. We have only just started imaging genetics analyses to explore such imaging phenotypes, but already quite promising results have been shown[19]. Moving towards HD endophenotypes will be a next logical step, because by aggregating connectivity measurements into some global parameters, most probably, we are losing even more biologically relevant information. However, connectivity research is still quite far from structural imaging genetics research due to several aspects. First of all, the sample sizes are even smaller than in structural analysis, where by itself it is far away from the millions of subjects in e.g. GWAS studies for height or BMI[21]. The need for more data in imaging genetics is an acute problem and I will describe it later. The more important issues for connectivity analysis are improving a standardization and harmonization of data acquisition and processing[22].

Ensuring that image-derived measurements are reliable and homogeneous is essential in multi-center imaging genetics studies, and especially challenging if we want to apply methods for HD connectivity endophenotypes. While there are many groups worldwide working on these issues, and considerable improvements have been obtained during the last years[23], still a lot of work has to be done in this direction.

Resting-state functional MRI is another imaging technique that can be used to evaluate regional interactions/connectivity that has shown promising results in imaging genetic studies[24,25]. This domain is even less explored than MRI diffusion data and suffers from the same challenges in the standardization of image acquisition and processing. The harmonization

is particularly challenging for resting state fMRI due to the wide variability of acquisition protocols and scanner platforms; this leads to site-to-site variance in quality, resolution and temporal signal-to-noise ratio[25].

NEW IMAGING METHODS

Improving image data harmonization would allow exploring an HD cross-domain analysis of structural, functional, diffusion and omics data. Many methods were already proposed in neuroimaging field for such combined analysis, like join independent component analysis (ICA)[26], canonic correlation analysis (CCA)[27], partial least squares (PLS)[27] and deep learning[28]. Adding omics as an extra domain should help with defining new imaging endophenotypes which are more relevant for genetic analyses of brain morphology, function and brain disease. By defining new imaging phenotypes I do not only mean selection (for instance the most heritable phenotypes), but also in some way to learn a specific low dimensional representation of HD measurements. This data reduction step can be performed in several ways, and deep learning approaches seem very promising in that respect. A number of possible ways are discussed below.

A first option, is to define the most relevant imaging biomarkers for a specific disorder using neuroimaging methods and then study the genetics of these specific phenotypes. Given a disorder with a strong genetic component, especially when associated with a relatively homogeneous pathological process, learning a disease-specific representation may be closer to the biological function than an expert-defined region of interest[29]. In the example on **Figure 1A,** a convolutional neural network (CNN), trained for classification, learns which parts of the brain are important for distinguishing between healthy and patients with Alzheimer disease. CNNs may not only be able to learn the complex feature interaction in disease, but can also be viewed as a method towards dimensionality reduction[30]. Feature maps or last fully connected layers might be used as a low dimensional representation of the endophenotypes. Another way to achieve the same is using autoencoders[30]. The neural network can learn the most relevant measurements for outcome information and project it into low dimensional latent space. Such techniques are much more promising than classical principal component analysis (PCA) or ICA, because it extracts phenotypes which are more relevant to a specific problem. In the example above, the deep hidden layers can be interpreted as low dimensional representations of the brain regions affected by Alzheimer Disease and can be used in follow up analysis as biomarkers. Such methods were already successfully applied outside the neuro field, for example, in studies of facial shape[31]. **A second option**, is to directly use images to assess the genetic architecture. For example, as shown on **Figure 1B,** CNN can take complex HD data as input without any prior assumptions about the relation between input measurements. Thus, we can supply the network with imaging data from several domains (T1, DTI, FLAIR etc). While

it is still difficult in deep learning to go back and check which input parameters are the most important for a specific task, this issue attracts more and more attention, especially in medical image analysis field, and a few quite promising techniques were developed recently[32,33].
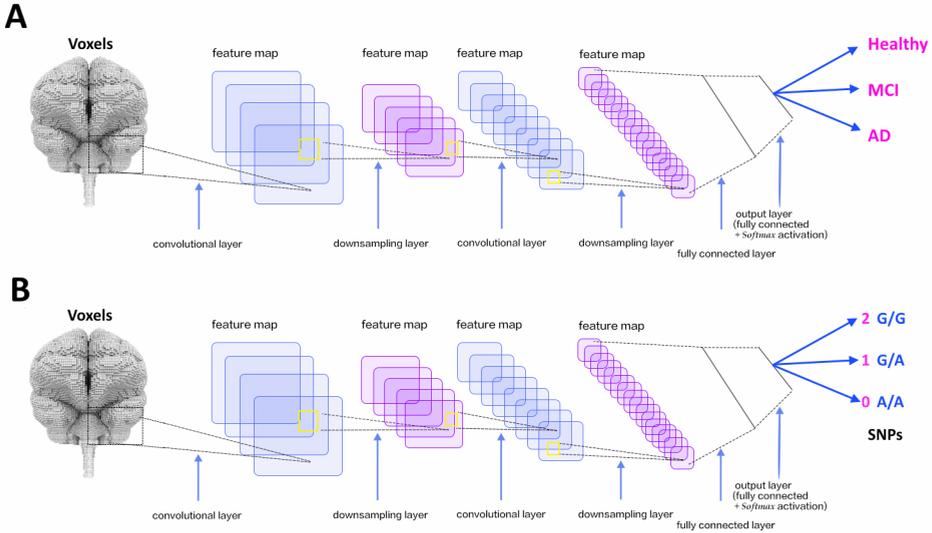

FIGURE 1: Deep learning application in imaging genetics

Finally, a third approach would be to define imaging endophenotypes based on genetics. Heritability estimates and genetic correlation between neuroimaging measurements are a useful instrument for defining more homogeneous endophenotype for genetic analysis. While higher heritability does not necessarily imply higher discoverability, it still can be used for feature selection or for increasing the genetic signal to noise ratio, as we showed in our voxel-based experiments[34]. Genetic correlation is a more powerful function, which allows combining measurements with similar genetic architecture and thus increasing statistical power. It would be interesting to incorporate genetic correlation directly into computational methods, for example, into PCA[35] or ICA. We would thus define HD imaging features that are the most genetically homogeneous or genetically independent. In such scenario it is possible that these new endophenotypes will serve as better imaging biomarkers for prediction/classification of diseases with strong genetic predisposition.

OMICS

During the past decade, genome-wide association studies have uncovered thousands of genetic variants that influence risk for complex human traits and diseases. While these findings

have dramatically increased our knowledge about human genetics, we are still quite far from the complete understanding of genetic architecture and underlying biology of these complex phenotypes. First, in contrast to Mendelian diseases, more than 90% of detected in GWAS variants are located in non-protein coding regions of the genome, and many even far away from the nearest known gene[36]. Therefore delineating the causal genetic variants and biological mechanisms underlying the observed statistical associations with disease risk requires additional, post-GWAS analysis. This still remains very challenging and such functional studies are currently lacking. Second, the most important loci from GWASs have small effect sizes and the significant hits, together, only explain a modest fraction of the predicted trait heritability. In the beginning, this was quite unexpected and even led researchers to call such phenomenon as "missing heritability"[37] . Since the first time it was mentioned, this issue has partly been resolved by taking into account large effect size of rare variants or by combining all common variants well below the threshold for statistical significance[38]. However, the mystery of "missing heritability" still remains largely unsolved[39].

As mentioned above, the vast majority of GWASs hits have been found in noncoding genome regions. This raises two main questions: (i) what are the **molecular functions** of the causal variants; (ii) which **genes** are affected. The noncoding location suggests that these variants affect gene expression other than in a changing protein sequence way, for example by the effect on transcription, splicing, mRNA stability etc. Thus**,** studying **epigenetics**, the gene regulatory mechanisms that do not involve changes in underlying DNA, are becoming more important. The epigenetic mechanisms are thought to be under both genetic and environmental influence, and at least partially heritable[40]. Epigenetic regulation is especially important for the human brain, where expression of the genes is extremely dynamic and changes during the lifespan[41]. Any disruption in such complex and precise machinery may have an effect on brain and related complex disorders development. One of such regulatory mechanism is **DNA methylation**, when methyl groups are added to the DNA molecule. Methylation can change the activity of a DNA segment without changing the sequence. When located in a gene promoter, DNA methylation typically acts to repress gene transcription. Several studies suggest that changes in DNA methylation may have a role in the onset of Alzheimer's Disease[42] or in the developmental of schizophrenia[43]. Another type of epigenetic regulation are **histone modifications**. These are chemical modifications of the DNA-binding histone proteins and influence transcription as well as other DNA processes. For example, trimethylation of histone H3 lysine 4 (H3K4me3) is associated with promoter regions, but trimethylation of histone H3 lysine 27 (H3K27me3) is associated with inactive promoters of protein-coding genes. Recently, owing to the effort of several consortia (such as ENCODE project[44], Road Epigenome project[45], FANTOM[46]) a large amount of data about human epigenetics and genome annotation have become available. These databases contain information about epi-

genetics for dozens of different type of cells and tissues. However, studying them individually is inefficient and requires novel algorithms for combining databases for molecular functional analysis[47,48]. Several unsupervised and supervised approaches already exist for integrating epigenetics features for inference of genetic variants functions[49,50]. Developing such analytical tools and new methods will become an essential part of the post-GWAS era[51]. **To answer the second question**, i.e. identifying which gene is affected by the causal variant, gene expression databases are available for dozens of cell and tissue types[52]. Thus screening for the association between gene expression level and loci, expression quantitative trait loci (**eQTL**) analysis, can be easily performed. However, because gene expression is tissue-specific, there are considerable challenges in the analysis of these data. Most studies include only blood cell types, as the most available tissue in patient-participants. To collect other tissue types can be problematic, especially in an invasive research setting. Tissue-specific databases, such as the Allen Human Brain Atlas, provide rich additional information on the spatial location of gene expression in the brain. This specific database is quite limited in the number of samples, but already allows to perform more advanced analyses[53-55].

All omics domains can be linked to each other (**Figure 2**) in a similar way as the different neuroimaging modalities described above. A next important step would be to develop cross-domain methodologies. Combining epigenetics, gene expression, genotype and (GWASs) data will have a significant impact on predicting susceptibility genes and discovering loci with biological functions[56]. It is most probable that many potentially important findings have to date not been found from omics data simply because the computational methods and algorithms that would highlight them are not developed yet or because of the lack of application of machine learning and deep learning techniques in the omics field.

OMNIGENIC

The GWAS results for complex traits and diseases are generally interpreted in a paradigm in which genetic variances are driven by an accumulation of weak effects on the key genes and regulatory pathways. This is leading by Fisher's famous '**infinitesimal model**', which he proposed in 1918 to explain continuous variations in human phenotypes: if many genes affect a trait, then the random sampling of alleles at each gene produces a continuous, normally distributed phenotype in the population. For example, recent results for height and BMI GWAS[57] demonstrate that there are hundreds or even thousands of genetic variants with very weak effects contributing to these phenotypes. Intuitively, we might expect that all these associations clustered towards some specific gene functional pathways that drive the trait etiology. However, it seems that this is not the case. For instance, for schizophrenia, 71%–100% of 1MB windows in the genome contribute to the heritability of this disease[58]. This implies that in every 1MB of DNA present genetic(s) variants associated with schizophrenia and therefore causal

genes are distributed among different (all) pathways. These observations gave rise for the idea of **omnigenic**[59], that gene regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes and that most heritability can be explained by effects on genes outside core pathways. This model summarizes all limitations of current research paradigm for causal variant discovery, namely the GWAS strategy with only **one single variant test** at a time, assuming the **additive genetic architecture**. While it is the most straightforward (and has been quite a successful way) to perform genetic association analysis, it has limitations and flaws, which in the era of big data became more obvious. The omnigenic paradigm proposes that with increasing sample size for GWAS at some point, since all genes somehow are interconnected in complex networks inside the cell, all genes will be associated with phenotype or disease. It also suggests that attention should be shifted from single variant analysis to more complex multivariate, gene interaction (**epistasis**) modeling. All current methods for heritability analysis assume an only additive model and high estimation of heritability leads many researchers to a wrong conclusion that *epistasis effect does not significantly contribute to genetic variance*. In fact, simulations have shown [60] that no matter which underlying genetic architecture is true: additive, dominant or epistasis, the inference of heritability assuming an additive model would still be high. Therefore, high additive heritability estimates do not say anything about epistasis contribution to total genetic variance. In model organisms, such as yeast, the interaction between two or three genes is quite a common phenomenon [61]. Moreover, the non-additive genetic architecture might also partly explain the "missing heritability"[62]. Actually, gene-gene or gene-environmental interaction models were used in the "early phase" of GWAS, but it quite fast became obvious that such analysis requires enormous computational resources and statistical power. Since that time computational facilities and methodologies improved significantly. Machine learning techniques, such as random forest or deep learning can model **millions connections and interactions** between input features, while the algorithms are trained within a short period of time. However, the application of deep learning for omics data requires the adaptation of convolutional layers from CNN (**Figure 1**) to the omics domain. In the omics field we are dealing with just one dimension, compared to 2D, 3D or even 4D structures in imaging field. Therefore it might sound as an easier task. In reality, omics data have a much more complex interaction structure alone this single dimension and classical convolution would not work. Proposing omics specific convolutional layers (networks) will be a quite challenging but promising area of research. It can be done based on an existing model, for example, from nature language processing applications, or by developing completely new approaches, for instance using linkage disequilibrium for convolution kernels size and epigenetic annotations as extra input channels. Graph convolutional neural networks (GCNN)[63] is an additional direction which is worth to explore for testing the omnigenic hypothesis. This type of network can be trained on input data represented as a graph, which fits perfectly to

model complex gene interactions inside different cell types, as proposed by the omnigenic paradigm. To the best of my knowledge, GCNN have not been applied to omics data so far. Additionally, using a machine learning techniques for interaction modeling it will be possible to redefine the way how **heritability** and **genetic correlation** are computed, which should definitely reduce the gap in "missing heritability".
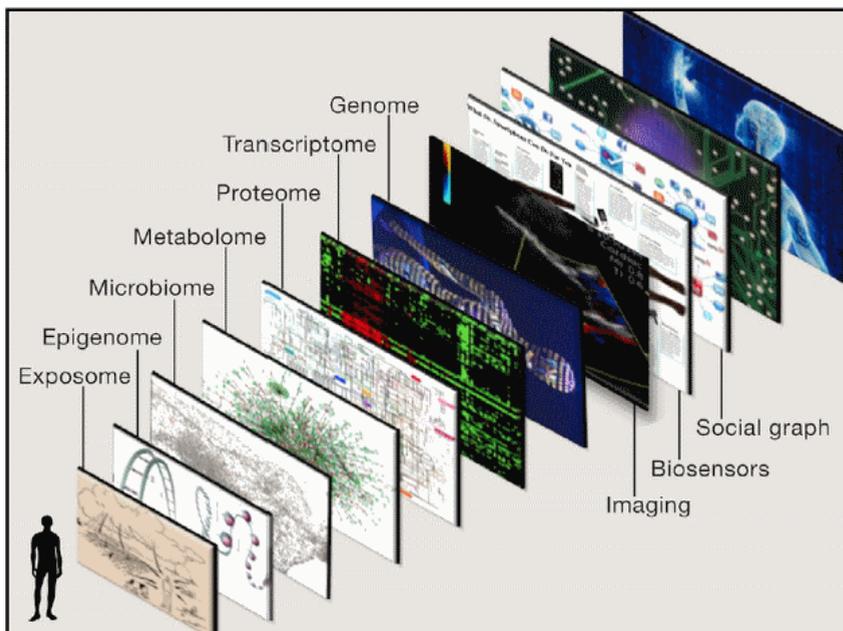

FIGURE 2: Omics data complexity[64]

CONCLUSIONS

In this thesis, I developed method for the integrated analysis of imaging and genetic data, and applied these to better understand the relation between genetics and brain structure and diseases. Specifically, **(i)** I investigated the heritability of high-dimensional MRI brain based phenotypes such as voxels and subcortical shapes; and **(ii)** developed a framework for HD omics analysis, which we successfully applied and discovered which genetic variants are associated with different brain voxels and cortical regions; **(iii)** I developed a method which allows to utilize GWAS summary statistics for pleiotropic loci discovery.

While we are only at the early stage of development of imaging genetics field itself, at the same time we are facing the post-GWAS era with an enormous amount of data available through national or world-wide biobanks. In coming years data might not be the main prob-

lem anymore and analysis, which was previously possible to carry out only on the multi-center level and which took a lot of time, will be possible to perform within a day(s). This will not only accelerate scientific finding, but will also allow challenging the old dogmas with novel ideas, such as described above omnigenic model or cross-domain analysis. On this basis, the scientific success will heavily rely on three components: **computational facilities, analytical tools** and **ability to generate such novel ideas**. Imaging genetics has a huge potential for addressing novel research questions to advance our knowledge of the brain and of the underlying neurobiology of complex disorders.

I strongly believe that these expanded efforts in imaging genetics will have a dramatic clinical impact in future. The field will improve our understanding of disease development, both leading to novel targets for therapy, and new tools for early detection and improved prognosis. This will, when translated towards clinical practice, have an enormous influence in every day healthcare.

REFERENCES

1. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research* **45**, D896-D901, doi:10.1093/nar/gkw1133 (2017).

2. Angelica, M. D. & Fong, Y. Pleiotropy in complex traits: challenges and strategies. *October* **141**, 520-529, doi:10.1016/j.surg.2006.10.010.Use (2013).

3. Franke, B. *et al.* Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nature Neuroscience*, doi:10.1038/nn.4228 (2016).

4. Satizabal, C. L. *et al.* Genetic Architecture of Subcortical Brain Structures in Over 40,000 Individuals Worldwide. *bioRxiv*, 173831 (2017).

5. Smoller, J. W. *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-1379, doi:10.1016/S0140-6736(12)62129-1.Identification (2013).

6. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279, doi:10.1093/bioinformatics/btw613 (2017).

7. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).

8. Gorgolewski, K. J. *et al.* NeuroVault.org: A repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain. *NeuroImage* **124**, 1242-1244, doi:10.1016/j.neuroimage.2015.04.016 (2016).

9. Steinberg, S. *et al.* Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nature Genetics* **47**, 445-447, doi:10.1038/ng.3246 (2015).

10. Medland, S. E., Jahanshad, N., Neale, B. M. & Thompson, P. M. Whole-genome analyses of whole-brain data : working within an expanded search space. *Nature Publishing Group* **17**, 791-800, doi:10.1038/nn.3718 (2014).

11. Bis, J. C. *et al.* Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nature Genetics* **44**, 545-551, doi:10.1038/ng.2237 (2012).

12. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **8**, doi:10.1038/nature14101 (2015).

13. Adams, H. H. H. *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nature Neuroscience* **id**, doi:10.1038/nn.4398 (2016).

14. Canu, E. *et al.* White matter microstructural damage in Alzheimer's disease at different ages of onset. *Neurobiol Aging* **34**, 2331-2340, doi:10.1016/j.neurobiolaging.2013.03.026 (2013).

15. Bozzali, M. *et al.* White matter damage in Alzheimer's disease assessed in vivo using diffusion tensor magnetic resonance imaging. *Journal of neurology, neurosurgery, and psychiatry* **72**, 742-746 (2002).

16. Hamoda, H. M. *et al.* Abnormalities in thalamo-cortical connections in patients with first-episode schizophrenia: a two-tensor tractography study. *Brain Imaging Behav*, doi:10.1007/s11682-018-9862-8 (2018).

17. Evans, T. E. *et al.* White Matter Microstructure Improves Stroke Risk Prediction in the General Population. *Stroke; a journal of cerebral circulation*, 1-8, doi:10.1161/STROKEAHA.116.014651 (2016).

18. Cremers, L. G. M. *et al.* Altered tract-specific white matter microstructure is related to poorer cognitive performance: The Rotterdam Study. *Neurobiology of Aging* **39**, 108-117, doi:10.1016/j.neurobiolaging.2015.11.021 (2016).

19. Thompson, P. M., Ge, T., Glahn, D. C., Jahanshad, N. & Nichols, T. E. Genetics of the connectome. *NeuroImage* **80**, 475-488, doi:10.1016/j.neuroimage.2013.05.013 (2013).

20. de Groot, M. *et al.* Improving alignment in Tract-based spatial statistics: evaluation and optimization of image registration. *NeuroImage* **76**, 400-411, doi:10.1016/j.neuroimage.2013.03.015 (2013).

21. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in~ 700,000 individuals of European ancestry. *bioRxiv*, 274654 (2018).

22. Petrov, D. *et al.* in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 515-522 (Springer).

23. Thompson, P. M. *et al.* The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging and behavior* **8**, 153-182, doi:10.1007/s11682-013-9269-5 (2014).

24. Xu, J. *et al.* Heritability of the Effective Connectivity in the Resting-State Default Mode Network. *Cereb Cortex* **27**, 5626-5634, doi:10.1093/cercor/bhw332 (2017).

25. Adhikari, B. M. *et al.* Heritability estimates on resting state fMRI data using ENIGMA analysis pipeline. *arXiv preprint arXiv:1709.08578* (2017).

26. Groves, A. R., Beckmann, C. F., Smith, S. M. & Woolrich, M. W. Linked independent component analysis for multimodal data fusion. *NeuroImage* **54**, 2198-2217, doi:10.1016/j.neuroimage.2010.09.073 (2011).

27. Sui, J., Adali, T., Yu, Q., Chen, J. & Calhoun, V. D. A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of neuroscience methods* **204**, 68-81, doi:10.1016/j.jneumeth.2011.10.031 (2012).

28. Plis, S. M. *et al.* Deep learning for neuroimaging: a validation study. *Front Neurosci* **8**, 229, doi:10.3389/fnins.2014.00229 (2014).

29. Fan, C. C. *et al.* Beyond heritability: Improving discoverability in imaging genetics. *Hum Mol Genet*, doi:10.1093/hmg/ddy082 (2018).

30. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504-507, doi:10.1126/science.1127647 (2006).

31. Claes, P. *et al.* Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat Genet* **50**, 414-423, doi:10.1038/s41588-018-0057-4 (2018).

32. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. *See https://arxiv. org/abs/1610.02391 v3* **7** (2016).

33. Oktay, O. *et al.* Attention U-Net: Learning Where to Look for the Pancreas. *arXiv preprint arXiv:1804.03999* (2018).

34. van der Lee, S. J. *et al.* Gray matter heritability in family-based and population-based studies using voxel-based morphometry. *Hum Brain Mapp* **38**, 2408-2423, doi:10.1002/hbm.23528 (2017).

35. Gutman, B. A. *et al.* in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. 1226-1230 (IEEE).

36. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science*

337, 1190-1195, doi:10.1126/science.1222794 (2012).

37. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753, doi:10.1038/nature08494 (2009).

38. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-569, doi:10.1038/ng.608 (2010).

39. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am J Hum Genet* **99**, 139-153, doi:10.1016/j.ajhg.2016.05.013 (2016).

40. Szyf, M. Nongenetic inheritance and transgenerational epigenetics. *Trends in molecular medicine* **21**, 134-144, doi:10.1016/j.molmed.2014.12.004 (2015).

41. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519-523, doi:10.1038/nature10524 (2011).

42. De Jager, P. L. *et al.* Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci* **17**, 1156-1163, doi:10.1038/nn.3786 (2014).

43. Jaffe, A. E. *et al.* Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat Neurosci* **19**, 40-47, doi:10.1038/nn.4181 (2016).

44. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

45. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

46. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461, doi:10.1038/nature12787 (2014).

47. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nature Protocols* **12**, 2478-2492, doi:10.1038/nprot.2017.124 (2017).

48. Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343, doi:10.1038/nature13835 (2015).

49. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).

50. Backenroth, D. *et al.* FUN-LDA: A Latent Dirichlet Allocation Model for Predicting Tissue-Specific Functional Effects of Noncoding Variation: Methods and Applications. *Am J Hum Genet* **102**, 920-942, doi:10.1016/j.ajhg.2018.03.026 (2018).

51. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**, 731-740, doi:10.1038/Nmeth.3901 (2016).

52. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660, doi:10.1126/science.1262110 (2015).

53. Roshchupkin, G. V. *et al.* Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology. *Neurobiol Aging* **48**, 204-211, doi:10.1016/j.neurobiolaging.2016.08.024 (2016).

54. Richiardi, J. Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 11-14 (2015).

55. Shin, J. *et al.* Cell-Specific Gene-Expression Profiles and Cortical Thickness in the Human Brain. *Cereb Cortex*, 1-11, doi:10.1093/cercor/bhx197 (2017).

56. Andrews, S. V. *et al.* Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder. *Nature Communications* **8**, doi:Artn 101110.1038/S41467-017-00868-Y (2017).

57. Yengo, L. *Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry*, 2018).

58. Loh, P. R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* **47**, 1385-1392, doi:10.1038/ng.3431 (2015).

59. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186, doi:10.1016/j.cell.2017.05.038 (2017).

60. Huang, W. & Mackay, T. F. C. The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis. *PLoS Genetics* **12**, doi:ARTN e100642110.1371/journal.pgen.1006421 (2016).

61. Kuzmin, E. *et al.* Systematic analysis of complex genetic interactions. *Science* **360**, doi:10.1126/science.aao1729 (2018).

62. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T. L. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234-237, doi:10.1038/nature11867 (2013).

63. Henaff, M., Bruna, J. & LeCun, Y. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163* (2015).

64. Topol, E. J. Individualized medicine from prewomb to tomb. *Cell* **157**, 241-253, doi:10.1016/j.cell.2014.02.012 (2014).

7

Summary

SUMMARY

Imaging genetics is still a quite novel area of research which attempts to discover how genetic factors affect brain structures and functions. In this thesis, using a various methodological approaches I showed how it can contribute to our understanding of the complex genetic architecture of the human brain.

Chapter 2 deals with two high-dimensional brain endophenotypes: voxels of gray matter density (**chapter 2.1**) and subcortical shapes (**chapter 2.2**). We found that both these endophenotypes are highly heritable and observed complex patterns of mosaic genetic effects. And in case of subcortical shapes the heritability estimates go beyond just volumetric measures.

In chapter 3, we showed that such high-dimensional phenotype can be successfully used not only to spatially localize relation between brain structures and genetic variants (**chapter 3.1**), but also with cognitive functions (**chapter 3.2**). In **chapter 3.1** we demonstrated that although detecting significant genetic effects on individual voxels requires large sample sizes, using three-dimensional association maps in combination with genes expression information can help to mine additional insight about how genes affect the brain structure.

Such cross-investigation approaches between high-dimensional phenotypes and omics data on multi-center level can provide deep insight into genetic architecture. In **chapter 4.1** we presented a framework (HASE) which (i) dramatically reduces computational time and **(ii)** allows performing meta-analysis for millions of phenotypes without exchanging of TBs of summary statistics. In **chapter 4.2** I presented a new method for pleiotropy analysis, the phenomena when one gene affects many outcomes. This method is based just on GWAS summary and able to detect pleiotropic genetic variants not only between two phenotypes, but also between several traits. We successfully applied it to investigate shared genetic architecture between subcortical brain structures and schizophrenia as well as between five major psychiatric disorders.

Finally, **chapter 5** presents several studies where we applied methods from chapter 4 within consortiums. In **chapter 5.1** we studied genetics of anterior commissure and found that there is a pleiotropy between various neurodegenerative diseases and underlying mechanism through which genes affect the anterior commissure. In **chapter 5.2** we performed genome-wide association studies for measures of surface area, cortical thickness and volume of 34 different cortical regions, in total more than 100 GWASes. Many contributed cohorts, with lack of computational facilities, used HASE framework to speed-up the analysis. This allowed

us to find 161 new genome-wide significant genetic variants which are associated with human brain cortex. And the last, in **chapter 5.3,** after processing more than 18.000 subjects with voxel-based morphometry pipeline and extracted gray matter voxels, we applied HASE framework for voxel-wise genome-wide association study (vGWAS), association between 1.5 million voxels and 9 million genetic variants (more than 13.500 billion regressions). Even with a stricter threshold of significance we were able to identify 29 independent loci associated with gray matter density in various regions of human brain.

NEDERLANDSE SAMENVATTING

Beeldvormingsgenetica is een vrij nieuw onderzoeksgebied waarin getracht wordt te ontdekken hoe hersenstructuren en -functies beïnvloed worden door genetische factoren. In dit proefschrift heb ik middels verschillende methodologieën aangetoond hoe dit onderzoeksgebied kan bijdragen aan ons begrip van de complexe genetische architectuur van het menselijk brein.

Hoofdstuk 2 behandelt twee hoog-dimensionale hersen-endofenotypen: voxels van grijze stof densiteit (**hoofdstuk 2.1**) en subcorticale vormen (**hoofdstuk 2.2**). We hebben gevonden dat beide endofenotypen zeer erfelijk zijn met complexe patronen van variërende genetische effecten, waarbij de erfelijkheid van de volumetrische maten wordt overstegen door die van de subcorticale vormen.

In **hoofdstuk 3** hebben we aangetoond dat een dergelijk hoog-dimensionaal fenotype niet alleen succesvol kan worden gebruikt om de relatie tussen hersenstructuren en genetische varianten te lokaliseren (**hoofdstuk 3.1**), maar ook tussen hersenstructuren en cognitieve functies (**hoofdstuk 3.2**). In **hoofdstuk 3.1** hebben we laten zien dat, hoewel er een grote steekproefgrootte nodig is voor het detecteren van significante genetische effecten op individuele voxels, drie-dimensionele associatiekaarten in combinatie met genexpressie-informatie kunnen helpen meer inzicht te verkrijgen in de manier waarop genen hersenstructuren beïnvloeden.

Zulke integrale onderzoekbenaderingen, waarin hoog-dimensionele fenotypen worden gekoppeld aan omics data in meerdere onderzoekscentra, kunnen een diepgaand inzicht geven in de genetische architectuur. In **hoofdstuk 4.1** hebben we een raamwerk (HASE) gepresenteerd, die (**i**) de rekentijd drastisch reduceert en (**ii**) ons de kans geeft om een meta-analyse te verrichten van miljoenen fenotypen zonder terabytes aan samenvattende statistieken uit te hoeven wisselen. In **hoofdstuk 4.2** heb ik een nieuwe methode gepresenteerd voor de analyse van pleiotropie, het fenomeen waarbij een gen meerdere uitkomsten beïnvloedt. Deze methode is alleen gebaseerd op de samenvattende statistieken van genoomwijde associatie-studies (GWAS) en is in staat om pleiotrope genetische varianten te detecteren tussen meer dan twee fenotypen. We hebben dit succesvol toegepast door de overlappende genetische architectuur te onderzoeken tussen zowel subcorticale hersenstructuren en schizofrenie als tussen vijf grote psychiatrische ziektes.

Tot slot worden in **hoofdstuk 5** studies besproken waarin we de methoden zoals beschreven in hoofdstuk 4 binnen consortia hebben toegepast. In **hoofdstuk 5.1** hebben we de gene-

tica van de commissura anterior bestudeerd, waarbij we hebben gevonden dat er pleiotropie lijkt te zijn waardoor genen zowel de commissura anterior als verschillende neurodegenera-tieve ziektes beïnvloeden. In **hoofdstuk 5.2** hebben we een GWAS uitgevoerd voor maten van de oppervlakte, de dikte en het volume van 34 verschillende hersenschors regio's, het-geen resulteerde in meer dan 100 GWAS. Veel bijdragende cohorten met beperkte compu-terfaciliteiten hebben het HASE raamwerk gebruikt om de analyse te versnellen. Dit heeft ons in staat gesteld om 161 nieuwe genoomwijde significante varianten te identificeren die zijn geassocieerd met de menselijke hersenschors. Tenslotte hebben we in **hoofdstuk 5.3** – na-dat we scans van meer dan 18.000 mensen hebben verwerkt middels voxel-gebaseerde mor-fometrie (VBM) en grijze stof voxels hebben verkregen – het HASE raamwerk gebruikt om een voxelwijde GWAS (vGWAS) te verrichten tussen 1,5 miljoen voxels en 9 miljoen genetische varianten, resulterend in meer dan 13.500 miljard regressies. Zelfs met een strengere drem-pelwaarde voor significantie waren we in staat om 29 onafhankelijke loci te identificeren die geassocieerd waren met grijze stof densiteit in verschillende regio's van het menselijke brein.

List of publications

JOURNAL PAPERS

1. **G. V Roshchupkin**, B. A. Gutman, et al. (2016). "Heritability of the shape of subcortical brain structures in the general population." Nature Communications **7**.

2. **G.V. Roshchupkin**, H. Adams, et al. (2016). "HASE: Framework for efficient high-dimensional association analyses." Scientific Reports **6**.

3. **G.V. Roshchupkin**, H. H. Adams, et al. (2016). "Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology." Neurobiology of Aging **48**: 204-211.

4. van der Lee, S. J.*, **G. V. Roshchupkin***, et al. (2017). "Gray matter heritability in family-based and population-based studies using voxel-based morphometry." Human Brain Mapping.

5. Ikram, M. A., M. W. Vernooij, **G. V. Roshchupkin**, et al. (2016). "Genetic susceptibility to multiple sclerosis: Brain structure and cognitive function in the general population." Multiple Sclerosis Journal.

6. Rigters, S. C., D. Bos, M. Metselaar, **G. V. Roshchupkin** et al. (2017). "Hearing Impairment Is Associated with Smaller Brain Volume in Aging." Frontiers in Aging Neuroscience **9**.

7. M A. Ikram, H.I. Zonneveld, **G.V. Roshchupkin** et al "Heritability and genome-wide associations studies of cerebral blood flow in the general population" The Journal of Cerebral Blood Flow & Metabolism

8. W. Huizinga, D.H.J. Poot, M.W. Vernooij, **G.V. Roshchupkin**, et al "A spatio-temporal reference model of the brain". NeuroImage

9. S.C. Rigters, L.G.M. Cremers, M.A. Ikram, M.P. van der Schroeff, M. de Groot, **G.V. Roshchupkin**, et al "White Matter Microstructure and Hearing Acuity in Older Adults: a population-based cross-sectional DTI study." Neurobiology of aging

10. Adams, H. H., H. Adams **... G. V. Roshchupkin ...** et al. (2016) "Partial derivatives meta-analysis: pooled analyses when individual participant data cannot be shared." bioRxiv

11. U. Mutlu, M.K. Ikram ,**G.V. Roshchupkin**, et al "Retinal neurodegeneration and voxel-based morphometry of brain MRI: the Rotterdam Study." Human Brain Mapping.

12. L.G.M. Cremers, L. Lahousse, M. de Groot, **G.V. Roshchupkin**, et al "Poor lung function is associated with reduced tract-specific white matter microstructure. The Rotterdam Study" submitted

13. H. I. Zonneveld*, **G. V. Roshchupkin*** et al "The neural substrate of cognition in general population." Submitted.

14. H.H. Adams*, A. Teumer*, D.P. Hibar*, **G. V. Roshchupkin*** et al "Genetic architecture of the human anterior commissure." In preparation.

15. **G.V. Roshchupkin*,** C. D. Langen* et al "Heritability of connectome of the brain in a population-based study." In preparation.

16. **G.V. Roshchupkin** et al „Genome-wide association studies of 1.5 million features of

brain morphometry." <u>In preparation.</u>

17. E. Hofer*, **G.V. Roshchupkin*,** et al. "GWAS of cortical thickness, surface area and volume in cortical regions of interest." <u>Submitted.</u>

18. Claudia L Satizabal* , Hieab HH Adams* , Derrek P Hibar*, Charles C White*, Maria J Knol, Jason L Stein, Markus Scholz, Murali Sargurupremraj, Neda Jahanshad, **Gennady V Roshchupkin ...** Sarah E Medland**, Joshua M Shulman**, Paul M Thompson**, Sudha Seshadri**, M Arfan Ikram** "Genetic Architecture of Subcortical Brain Structures in Over 40000 Individuals Worldwide." <u>Submitted.</u>

19. Sven J. van der Lee*, Hieab H.H. Adams*, Najaf Amin, Lisa R. Yanek **... Gennady Roshchupkin ...** Meike W. Vernooij, M. Arfan Ikram, Charles S. Decarli "Novel genetic loci associated with brain lobar volumes" <u>In preparation.</u>

20. Katrina L. Grasby*, Neda Jahanshad* **... Gennady V Roshchupkin ...** Jason L. Stein**, Paul M. Thompson**, Sarah E. Medland** "The genetic architecture of the human cerebral cortex." <u>Submitted.</u>

21. Benjamin Grenier-Boley , Rebecca Sims , Dr. Joshua Bis , Adam Naj , Dr. Anne BOLAND , Maria Vronskaya , Mr. Sven van der Lee **... Gennady Roshchupkin ...** Kevin Mayo , John Growdon , Thomas "Meta-analysis of genetic association with diagnosed Alzheimer's disease identifies novel risk loci and implicates Abeta, Tau, immunity and lipid processing." <u>Submitted.</u>

22. **G.V. Roshchupkin** et al "Sum Ranking, simple but powerful method for detecting pleiotropic loci." <u>In preparation.</u>

23. J.Wang, M.Knol … W. Niessen, **G.V. Roshchupkin** "Grey Matter Age Prediction as a Biomarker for Risk of Dementia: Population-based Study." <u>In preparation.</u>

24. Natalia Vilor-Tejedor, M Arfan Ikram, **Gennady V. Roshchupkin** et al "Independent Multiple Factor Association Analysis for multiblock data in Imaging Genetics." <u>Submitted.</u>

25. Jean Shin, Shaojie Ma, Edith Hofer, Yash Patel, **Gennady V. Roshchupkin** … Sudha Seshadri, Tomas Paus for the neuroCHARGE Working Group "Planar cell polarity pathway and development of the human visual cortex." <u>Submitted.</u>

26. Verbruggen JG, Ikram MA, **Roshchupkin GV** … Adams HH "Asymptomatic intracranial meningiomas in the general population: spatial distribution and determinants". <u>In preparation.</u>

27. Wolter FJ, **Roshchupkin GV ...** Bos D "Carotid artery stenosis and imaging markers of neurodegeneration: an interhemispheric comparison in individuals with unilateral steno-occlusive disease" <u>In preparation</u>

CONFERENCES

1. **G.V. Roshchupkin**, S.J. van der Lee, H.H.H. Adams , C. Langen, M.W. Vernooij, C.M. van Duijn, W.J. Niessen, M.A. Ikram „Alzheimer Disease SNPs and their effect on brain morphology: The Rotterdam Study." *HBM 2015: Annual Meeting of the Organization for Human Brain Mapping, 2015*

2. **G.V. Roshchupkin**, H.H.H. Adams, S.J. van der Lee et al. Effect of Alzheimer Disease risk loci on brain morphology through modulation of gene expression." *HBM 2016: Annual Meeting of the Organization for Human Brain Mapping, 2016*

3. Adams, H. H., **G.V. Roshchupkin**, et al. (2016). " Anterior commissure: neuroanatomic and cognitive correlates in a population-based study." Alzheimer's & Dementia: The Journal of the Alzheimer's Association 12(7): P523

4. **G.V. Roshchupkin**., H. I. Zonneveld, et al. (2016). "Grey matter density in relation to cognitive function." Alzheimer's & Dementia: The Journal of the Alzheimer's Association **12**(7): P288. (oral presentation)

5. Adams, H. H., **G.V. Roshchupkin**, et al. (2016) "Genome-wide association studies of 1.5 million features of brain morphometry: the Rotterdam Study" CHARGE consortium meeting, Charlottesville, USA

6. **G.V. Roshchupkin**, Adams, H. H. et al. (2016) "HASE: Framework for efficient high-dimensional association analyses." CHARGE consortium meeting, Charlottesville, USA

7. **G.V. Roshchupkin**, Adams, H. H. et al. (2017) "Genome-wide association studies of the human anterior commissure followed by fine-mapping and expression analysis." CHARGE consortium meeting, New York, USA

8. Vanja Vlasov, Hieab H.H. Adams**, G.V. Roshchupkin,** et al. (2017). "Grey matter density in relation to cognitive function." AAIC 2017, Alzheimer's Association International Conference.

9. Q. Yang, **G.V. Roshchupkin**, et al "A fast, accurate two-step linear mixed model for genetic analysis applied to repeat MRI measurements", Pacific Symposium on Biocomputing (2018)

10. **G.V. Roshchupkin**, Adams, H. H. et al. (2018) "Sum Ranking, simple but powerful method for detecting pleiotropic loci." CHARGE consortium meeting, Rotterdam, Netherlands (oral presentation)

11. M.J. Knol, **G.V. Roshchupkin,** et al. (2018) "High-dimensional analysis of RNA expression with cortical thickness." CHARGE consortium meeting, Rotterdam, Netherlands

12. E. Hofer, **G.V. Roshchupkin,** et al. (2018) "GWAS of cortical thickness, surface area and volume in cortical regions of interest." CHARGE consortium meeting, Rotterdam, Netherlands

13. J. Shin, **G.V. Roshchupkin**, et al (2018) "Genetic variations associated with growth of the

human cerebral cortex: A principal-component approach" CHARGE consortium meeting, Rotterdam, Netherlands

14. **G.V. Roshchupkin**, et al (2018) "High-dimensional voxel-wise genotype-wide association study.", *HBM 2018: Annual Meeting of the Organization for Human Brain Mapping*

CONFERENCE PAPERS

1. C. D. Langen, **G.V. Roshchupkin** "Heritability of connectivity and disconnectivity of the brain in a population-based study." (ISBI 2017)

2. N Jahanshad, J Faskowitz, **G.V. Roshchupkin** (2015) "Multi-site meta-analysis of morphometry", 14th International Workshop on Data Mining in Bioinformatics

3. N Jahanshad, **G.V. Roshchupkin**, J Faskowitz et al Multi-site meta-analysis of image-wide genome-wide associations of morphometry. MICCAI Imaging Genetics Workshop, 2015

4. W. Huizinga, D.H.J. Poot, **G.V. Roshchupkin**, E.E. Bron, M.A. Ikram, M.W. Vernooij, D. Rueckert, W.J. Niessen and S. Klein, Modeling the brain morphology distribution in the general aging population, SPIE 2016

5. Q. Yang, **G.V. Roshchupkin**, et al "A fast, accurate two-step linear mixed model for genetic analysis applied to repeat MRI measurements", NIPS BigNeuro workshop (2018)
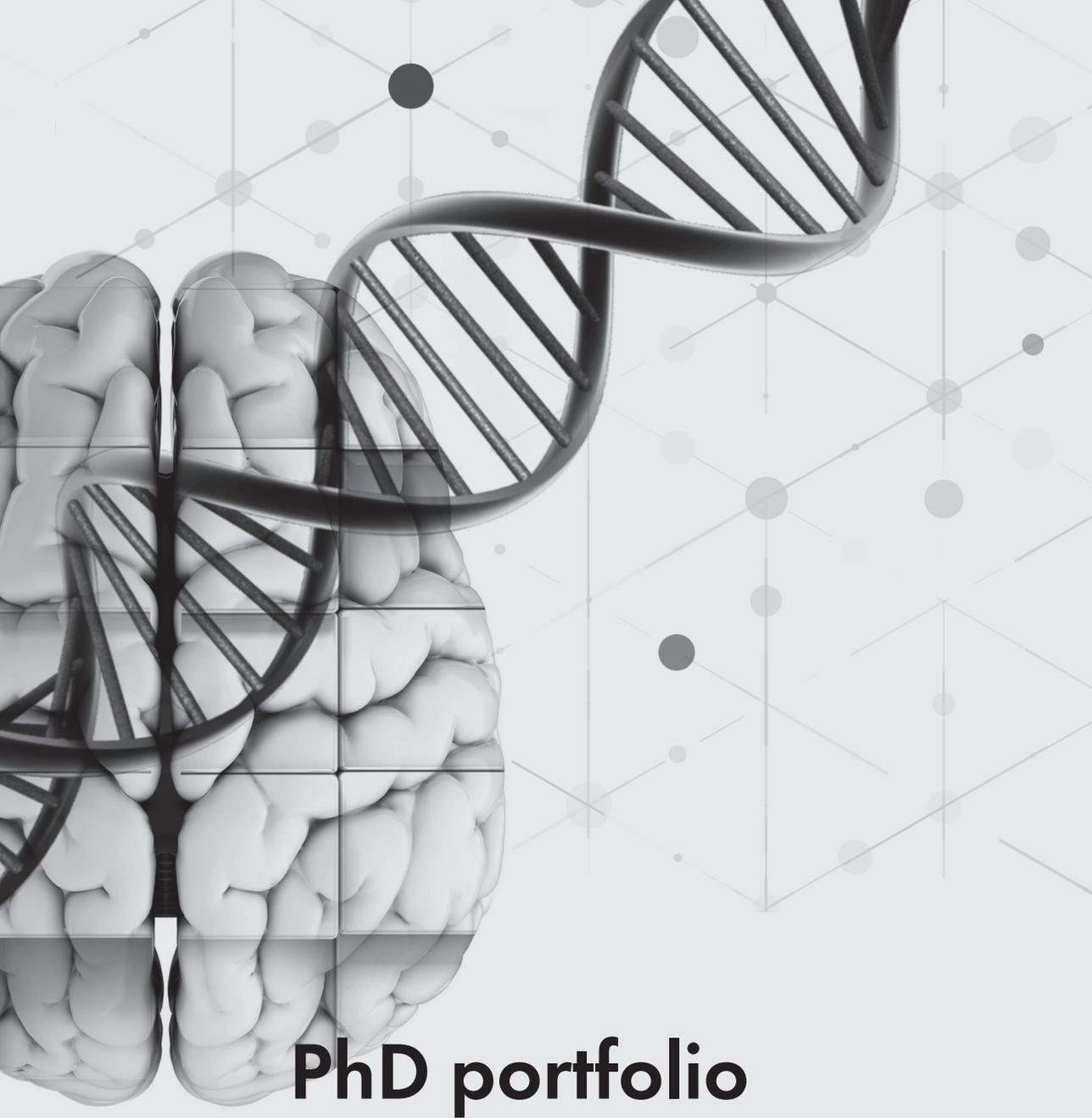
PhD portfolio

COURSES

Knowledge driven Image Segmentation	2014
Principles of genetic epidemiology	2014
Genome wide association analysis	2014
Scientific English Writing	2015
Advance Pattern Recognition	2015
A Programmer's Guide for Modern High-Performance Computing	2016
Research Integrity	2016
History of epidemiological ideas	2018
Microbiomics	2018
Summer school on deep learning for image analysis, Denmark	2014

INTERNATIONAL AND LOCAL RESEARCH MEETINGS

HD-READY: High Dimensional research in Alzheimer's Disease, Rotterdam, Netherlands, 2014

HD-READY: High Dimensional research in Alzheimer's Disease, Rotterdam, Netherlands, 2015

Full-HD: Full exploitation of High-Dimensionality in brain imaging Rotterdam, Netherlands, 2017

Full-HD: Full exploitation of High-Dimensionality in brain imaging, Stockholm, Sweden, 2017

BRIDGE: BRain Imaging, cognition, Dementia and next generation GEnomics, Greifswald, Gerrnany, 2017

BRIDGE: BRain Imaging, cognition, Dementia and next generation Genomics, Graz, Austria, 2018

Medical Informatics research lunch meeting (biweekly), 2014-2018

Neuro Image Analysis group meeting (weekly), 2014-2017

Neuro-epi group meeting (weekly), 2016-2018

TEACHING

Teaching Assistant

Image Processing course, Minor Medical Delta eduation, 2016, Rotterdam, Netherlands

Teaching Assistant

Practice of Epidemiologic Analysis, Erasmus Summer School 2017, Rotterdam, Netherlands

Invited lecturer

Cognomics Summer School 2018, Nijmegen, Netherlands

SUPERVISING STUDENTS

Supervision capstone thesis: Emine Taytas "Hypertension and brain morphometry: the heart-brain interaction".

Supervision master thesis: Ingrid Szilagyi "A Multivariate Genome Wide Association study of Subcortical Brain Structures: the Rotterdam Study".

Supervision master thesis: John Wang "Deep Learning approaches for multivariate imaging genetics analysis in the Rotterdam Study"

PEER REVIEWS OF SCIENTIFIC JOURNALS

Human Brain Mapping, Transactions on Medical Imaging, International Journal of Epidemiology, Frontiers in Neuroinformatics, Frontiers in Genetics

AWARDS

Medical Delta Young Scientist award (second place) , 2017
International CHARGE consortium Golder Tiger Contribution Awards, 2018

INTERNATIONAL RESEARCH VISITS

04/2017 – 05/2017   **Imaging Genetics Center, Keck School of Medicine, USC, Los Angeles, USA.**

GRANTS

- Full-HD (JPND 2016) – Co-applicant
- Erasmus Trust Fund (2017) – Research visit to another universirty abroad
- Alzheimer Nederland (2017) – Travel grant to another university
- Automated Detection of Intracranial Meningiomas and Aneurysms (MRace 2017) - Co-applicant
- Nvidia research GPU grant (2018)

About
the author

Gennady Vasilievich Roshchupkin was born in Sergiev Posad, Moscow region, Russia on 11th April 1988. He finished Physics and Mathematics high school in 2005 in his home town. While in school, he was always curious about science and participated in various local and national Olympiads (competitions) in mathematics, physics, computer science, chemistry and biology. After winning the diplomas in physics and mathematics at interregional Russian Olympiads, he was accepted to Moscow State University in Physics Department. He chose to study astronomy and finished the university in 2011 with specialization in stellar astronomy and astrophysics. After the university, he worked for 2.5 years as the senior research engineer in the field of network security and mathematical modeling in a startup company in Moscow. He participated in several projects, about machine learning and algorithm optimization for fast big data processing. In 2014 Gennady started his PhD at the department Medical Informatics, Erasmus MC Medical Center in the Netherlands in the Biomedical Imaging Group Rotterdam (BIGR) supervised by prof. Wiro Niessen and prof. Arfan Ikram (Epidemiology Department). For his thesis, he worked under Population imaging genetics (ImaGene) project developing new methods to analyze large datasets of MRI scans and omics data. In 2017, Gennady, supported by grants from Alzheimer Netherlands and Erasmus Trustfonds visited Imaging Imaging Genetics Center, Los Angeles, USA led by prof. Paul Thompson, where he spent several months working on new analytical solutions for high-dimensional imaging data analysis together with dr. Neda Jahanshad. During his PhD at Erasmus MC, Gennady was also a part of The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium and in 2018 he was awarded by CHARGE Golden Tiger Contribution Awards.

Since March 2018, Gennady has started to work as a joint postdoc between the departments of Epidemiology and Medical Informatics, Erasmus MC. He is planning to continue his scientific career and focus his research on complex omics data analysis.

.