Neurobiology of Aging 36 (2015) S151–S158

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

Imaging genetics

Genome-wide interaction analysis reveals replicated epistatic effects on brain structure

Derrek P. Hibar [a], Jason L. Stein [a], Neda Jahanshad [a], Omid Kohannim [a], Xue Hua [a], Arthur W. Toga [a], Katie L. McMahon [b], Greig I. de Zubicaray [c], Nicholas G. Martin [d], Margaret J. Wright [d], the Alzheimer's Disease Neuroimaging Initiative [1], Michael W. Weiner [e,f,g,h], Paul M. Thompson [a,*]

[a] Imaging Genetics Center, Institute for Neuroimaging and Informatics, University of Southern California, Los Angeles, CA, USA
[b] Centre for Magnetic Resonance, School of Psychology, University of Queensland, Brisbane, Queensland, Australia
[c] Functional Magnetic Resonance Imaging Laboratory, School of Psychology, University of Queensland, Brisbane, Queensland, Australia
[d] Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia
[e] Department of Radiology, UC San Francisco, San Francisco, CA, USA
[f] Department of Medicine, UC San Francisco, San Francisco, CA, USA
[g] Department of Psychiatry, UC San Francisco, San Francisco, CA, USA
[h] Department of Veterans Affairs Medical Center, San Francisco, CA, USA

ARTICLE INFO

Article history:
Received 2 May 2013
Received in revised form 10 February 2014
Accepted 16 February 2014
Available online 27 August 2014

Keywords:
Epistasis
Interaction
Genome-wide
GWAS
GWIA
Sure independence screening
Tensor-based morphometry

ABSTRACT

The discovery of several genes that affect the risk for Alzheimer's disease ignited a worldwide search for single-nucleotide polymorphisms (SNPs), common genetic variants that affect the brain. Genome-wide search of all possible SNP-SNP interactions is challenging and rarely attempted because of the complexity of conducting approximately 10^{11} pairwise statistical tests. However, recent advances in machine learning, for example, iterative sure independence screening, make it possible to analyze data sets with vastly more predictors than observations. Using an implementation of the sure independence screening algorithm (called EPISIS), we performed a genome-wide interaction analysis testing all possible SNP-SNP interactions affecting regional brain volumes measured on magnetic resonance imaging and mapped using tensor-based morphometry. We identified a significant SNP-SNP interaction between rs1345203 and rs1213205 that explains 1.9% of the variance in temporal lobe volume. We mapped the whole brain, voxelwise effects of the interaction in the Alzheimer's Disease Neuroimaging Initiative data set and separately in an independent replication data set of healthy twins (Queensland Twin Imaging). Each additional loading in the interaction effect was associated with approximately 5% greater brain regional brain volume (a protective effect) in both Alzheimer's Disease Neuroimaging Initiative and Queensland Twin Imaging samples.

1. Introduction

Many common brain disorders such as Alzheimer's disease (AD), schizophrenia, and bipolar disorder are more prevalent in family members of those affected than in the population as a whole (Lichtenstein et al., 2009; Pedersen et al., 2004). If disease risk increases in relatives of patients, it is possible to use family studies to estimate the overall proportion of disease risk attributable to common or rare transmitted variants in our DNA; this is the concept of heritability (Neale and Cardon, 1992).

However, identifying the specific DNA variants associated with increased disease risk is an incredibly complex task. There are over 3 billion base pairs in our DNA, and over 10 million of these are known to have variations that are somewhat prevalent (>1%) in the population (1000 Genomes Project Consortium et al., 2010). Each of these variants may have a unique and often unknown role to play in the biology of the human body although the vast majority likely has no role at all. Similarly, for

* Corresponding author at: Imaging Genetics Center, Institute for Neuroimaging and Informatics, Keck School of Medicine of USC, University of Southern California, 2001 N. Soto Street, SSB1-102, Los Angeles, CA 90032, USA. Tel.: +1 323 442 7246; fax: +1 323 442 0137.
E-mail address: pthomp@usc.edu (P.M. Thompson).

[1] Many investigators within the Alzheimer's Disease Neuroimaging Initiative (ADNI) contributed to the design and implementation of ADNI and/or provided data, but most of them did not participate in analysis or writing of this report. A complete list of ADNI investigators may be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

0197-4580/$ – see front matter © 2015 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neurobiolaging.2014.02.033

many brain disorders we have an incomplete understanding of the underlying etiology. Commonly measured clinical scores are used for diagnosis, but, in some cases, neuroimaging measures may offer better biomarkers of disease progression and severity (Braskie et al., 2013; Jack et al., 2004).

The field of neuroimaging genetics uses neuroimaging biomarkers as proxies for disease (also called endophenotypes; Gottesman and Gould, 2003) to identify specific genetic variants that affect quantitative measures of the brain structure or function. One goal of imaging genetics is to identify common genetic variants that affect the brain, positively or negatively, and then understand if and how any of those variations are associated with increased risk for developing a specific brain disease. Conversely, it is possible to use neuroimaging to identify the effects of AD risk genes whose function is not yet well understood (Braskie et al., 2011). For example, a common variant in the *CLU* gene confers a heightened risk for AD (by 10%–20%) in a large sector of the population, although the mechanism is not known. Neuroimaging of carriers of this variant revealed widespread reduction in the brains' fiber integrity around 50 years before the disease is typically diagnosed. Similarly, the *TREM2* gene harbors rarer variants that elevate AD by a still greater factor, and neuroimaging has recently established that carriers of the adverse variant lose brain tissue faster (Rajagopalan et al., 2013).

Until recently, neuroimaging genetics studies have tended to focus on candidate genes such as brain-derived neurotrophic factor (Bueller et al., 2006) and catechol-O-methyltransferase (Egan et al., 2001). Biffi et al. (2010) looked at AD candidate genes APOE, CR1, and PICALM and found that each gene has significant effects on neuroimaging biomarkers like hippocampal volume. Candidate gene studies examine small subsets of gene changes chosen from the millions of variants in our DNA based on prior hypotheses about underlying disease pathways. However, many candidate gene studies have a mixed history of replication (see Supplementary Tables 7 and 8 in Stein et al., 2012). For many candidate genes in psychiatry, although not so much in the dementia field, there is some level of controversy or uncertainty as to whether the effects are robust; very large consortia, such as the Psychiatric Genomics Consortium (Ripke et al., 2011) and the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) Consortium (Hibar et al., 2013; Jahanshad et al., 2013; Stein et al., 2012) have been set up to verify genetic effects with unprecedented power. In contrast, genome-wide association studies (GWAS), which systematically screen millions of common variants in our DNA, called single-nucleotide polymorphisms (SNPs), have recently found a large number of replicated associations of genetic polymorphisms with disease, often using a hypotheses free screen of the genome (Harold et al., 2009). For example, Stein et al. (2012) performed a GWAS of mean hippocampal volume, total brain volume, and intracranial volume in 10,372 subjects for the ENIGMA Consortium. Stein et al. (2012) identified 2 genome-wide significant SNPs related to hippocampal volume rs7294919 (located in chromosome 12q24.22) and intracranial volume rs10784502 (located in chromosome 12q14.3). The results were independently replicated in another large GWAS by the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (Bis et al., 2012). The convergent results from the ENIGMA and Cohorts for Heart and Aging Research in Genomic Epidemiology consortia provide evidence and replication for real genetic effects on hippocampal and intracranial volumes that are consistent worldwide.

However, many of the reported findings from GWA studies have small effect sizes and explain only a small proportion of the variance estimated to be because of purely genetic factors. In the ENIGMA study of hippocampal volume, Stein et al. (2012) showed that mean hippocampal volume was 64%–72% heritable, but their most significant SNP explained only 0.265% percent of the total observed variance in hippocampal volume. Similarly, height is very highly heritable (around 80%; Macgregor et al., 2006; Silventoinen et al., 2003), and a large GWAS of height in 183,727 subjects identified 180 significant SNPs that collectively explain 10% of the observed variance in height (Lango Allen et al., 2010). These findings have led to speculation about the source of the missing heritability; the proportion of variance in a trait that we know is influenced by genetics, but that is undetectable, so far, in the common genetic variants examined to date in GWA studies. Potential sources of the missing heritability might be caused by nonadditive effects like dominance and SNP-SNP interactions (called epistasis; Carlborg and Haley, 2004) and gene-by-environment interactions (Visscher et al., 2008), and rare genetic variants (Manolio et al., 2009). It is also possible that deeper sequencing of the genome will identify causal loci with greater effects, as GWAS often genotypes only a subset of the common variants in the genome. Whole-exome sequencing and whole-genome sequencing, for example, are already underway for the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort. Although interaction testing holds promise, depending on the influence of the underlying interaction current statistical approaches can be underpowered (Marchini et al., 2005). Further still, some estimates show that interactions in regions outside of the highly polymorphic human leukocyte antigen region in the genome might not significantly improve our understanding of the problem of missing heritability (Clayton, 2009). In this article, we will focus our analysis on SNP-SNP epistatic interactions. These are not well studied and some of the computational reasons and challenges are explained below, along with a proposed solution.

Some prior studies have examined epistatic effects of SNPs on brain structure (Pezawas et al., 2008; Tan et al., 2007; Wang et al., 2009). Chiang et al. (2012), tested for SNP effects on diffusion imaging measures, and aggregated all SNPs with correlated effects into a network. The concept here is different and aims to assess gene pairs that influence each other's effects on the brain. None of these prior studies has considered genome-wide genotype data; the closest conceptually related study tested interaction effects for preselected SNPs in genes and pathways already known to be related to AD (Meda et al., 2013). Any approach based on preselecting a pair of genes will overlook a vast search space of potential interactions among SNPs in the genome that have no obvious prior connection. In an interaction model, a predictor variable in the model does not have to be significant to result in a significant interaction. This is another way of saying that dropping nonsignificant SNPs from the SNP-SNP interaction search will miss some important interactions (Cordell, 2009). Given this, prior hypotheses focusing only on SNPs that have the largest known individual effects may also overlook large epistatic interaction effects. Intriguingly, power estimates for detecting interactive effects for certain models of the genetic contribution to complex traits are comparable with those for single SNP tests (Marchini et al., 2005). The inclusion of interaction terms was shown to boost the power to detect main effects in models of type 1 diabetes (Cordell et al., 2001). Here, we examine the genome-wide, SNP-SNP "interactome" to test genetic associations with a quantitative biomarker of AD (temporal lobe volume) in the publicly available ADNI data set. We further examine the whole-brain effects of interaction pairs in statistical parametric maps generated with tensor-based morphometry (TBM); we also replicate our tests in an independent, nonoverlapping data set of young healthy twins from the Queensland Twin Imaging (QTIM) study (de Zubicaray et al., 2008).