frontiers in
HUMAN NEUROSCIENCE

METHODS ARTICLE
published: 21 February 2012
doi: 10.3389/fnhum.2012.00021

An ICA with reference approach in identification of genetic variation and associated brain networks

Jingyu Liu[1,6]*, Mohammad M. Ghassemi[1], Andrew M. Michael[1], David Boutte[1], William Wells[2], Nora Perrone-Bizzozero[3], Fabio Macciardi[4], Daniel H. Mathalon[5], Judith M. Ford[5], Steven G. Potkin[4], Jessica A. Turner[1] and Vince D. Calhoun[1,3,6]

[1] The Mind Research Network, Albuquerque, NM, USA
[2] Department of Radiology, Harvard Medical School, Boston, MA, USA
[3] Department of Neurosciences, University of New Mexico School of Medicine, Albuquerque, NM, USA
[4] Department of Psychiatry and Human Behavior, University of California, Irvine, CA, USA
[5] Veterans Affairs Medical Center, San Francisco, CA, USA
[6] Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA

Edited by:
Kenneth Hugdahl, University of Bergen, Norway

Reviewed by:
Tamer Demiralp, Istanbul University, Turkey
Henrik Walter, Charite Universitätsmedizin, Germany

***Correspondence:**
Jingyu Liu, The Mind Research Network, 1101 Yale Boulevard, Albuquerque, NM 87131, USA.
e-mail: jliu@mrn.org

To address the statistical challenges associated with genome-wide association studies, we present an independent component analysis (ICA) with reference approach to target a specific genetic variation and associated brain networks. First, a small set of single nucleotide polymorphisms (SNPs) are empirically chosen to reflect a feature of interest and these SNPs are used as a reference when applying ICA to a full genomic SNP array. After extracting the genetic component maximally representing the characteristics of the reference, we test its association with brain networks in functional magnetic resonance imaging (fMRI) data. The method was evaluated on both real and simulated datasets. Simulation demonstrates that ICA with reference can extract a specific genetic factor, even when the variance accounted for by such a factor is so small that a regular ICA fails. Our real data application from 48 schizophrenia patients (SZs) and 40 healthy controls (HCs) include 300K SNPs and fMRI images in an auditory oddball task. Using SNPs with allelic frequency difference in two groups as a reference, we extracted a genetic component that maximally differentiates patients from controls ($p < 4 \times 10^{-17}$), and discovered a brain functional network that was significantly associated with this genetic component ($p < 1 \times 10^{-4}$). The regions in the functional network mainly locate in the thalamus, anterior and posterior cingulate gyri. The contributing SNPs in the genetic factor mainly fall into two clusters centered at chromosome 7q21 and chromosome 5q35. The findings from the schizophrenia application are in concordance with previous knowledge about brain regions and gene function. All together, the results suggest that the ICA with reference can be particularly useful to explore the whole genome to find a specific factor of interest and further study its effect on brain.

Keywords: independent component analysis with reference, genome-wide association study, brain network, schizophrenia, single nucleotide polymorphisms, functional magnetic resonance imaging

INTRODUCTION

Considerable research suggests that complex genetic factors contribute to the etiology of mental diseases including Alzheimer's, Parkinson's and schizophrenia (Serretti et al., 2007; Allen et al., 2008; Simon-Sanchez et al., 2009). For instance, schizophrenia is highly heritable, with heritability of 0.6–0.8 (Rao et al., 1981) and the concordance between identical twins is about 50% (Cannon et al., 1998), suggesting roles for both genetic and environmental influences. Linkage studies on family samples and association studies on population samples both identify multiple genetic variants associated with the disease (Hovatta et al., 1998; Levinson et al., 2000; Fan et al., 2006; Purcell et al., 2009). One way to test the genetic risk is to perform a focused study on a selection of specific genes or chromosome loci, which are hypothesized to relate to the disorder-based on *a priori* knowledge of the molecular and cellular functions. While useful, this approach may overlook genetic elements that have not yet been studied but may actually play an important role in a given disorder. This limitation, combined with the known genetic complexity of many diseases, provides strong motivation for performing a broad genome-wide association study (GWAS) on a large number of single nucleotide polymorphisms (SNPs).

GWAS methods enable the study of genes that have not been studied or that have been understudied for a particular disease (Hirschhorn and Daly, 2005) and thus, hold considerable potential for identifying genes associated with complex genetic illnesses (Hindorff et al., 2009). Unfortunately, GWAS is not without its own shortcomings. The main challenge is a statistical power limitation, which is further exacerbated if only a limited number of subjects are available. To balance a small sample size and a large volume of SNP data, a set-based approach has been proposed that, instead of testing the association of each genetic marker one by