Human Brain Mapping

What is RSS?

Early View (Articles online in advance of print)

Published Online: 10 Dec 2007

Research Article

Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA†

Jingyu Liu [1 2 *], Godfrey Pearlson [3 4], Andreas Windemuth [5], Gualberto Ruano [5], Nora I. Perrone-Bizzozero [6], Vince Calhoun [1 2 3 4 6]

[1]The Mind Research Network, Albuquerque, New Mexico

[2]Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, New Mexico

[3]Olin Neuropsychiatry Research Center, Institute of Living, Hartford, Connecticut

[4]Department of Psychiatry, Yale University, New Haven, Connecticut

[5]Genomas, Inc., 67 Jefferson Street, Hartford, Connecticut

[6]Department of Neurosciences, University of New Mexico School of Medicine, Albuquerque, New Mexico

email: Jingyu Liu (jliu@themindinstitute.org)

[*]Correspondence to Jingyu Liu, The Mind Research Network, 1101 Yale Boulevard, NE, Albuquerque, NM

† The work was performed at Institute of Living, Hartford, Connecticut.

Funded by:

- National Institutes of Health; Grant Number: 1 R01 EB 000840, 1 R01 EB 005846, 2 RO1 MH43775, 5 RO1 MH52886
- MIND Institute

KEYWORDS

functional magnetic resonance imaging • single nucleotide polymorphisms • component analysis • genotypes • phenotypes • multiple modalities • Physiogenomics

ABSTRACT

There is current interest in understanding genetic influences on both healthy and disordered brain function. We assessed brain function with functional magnetic resonance imaging (fMRI) data collected during an auditory oddball task - detecting an infrequent sound within a series of frequent sounds. Then, task-related imaging findings were utilized as potential intermediate phenotypes (endophenotypes) to investigate genomic factors derived from a single nucleotide polymorphism (SNP) array. Our target is the linkage of these genomic factors to normal/abnormal brain functionality. We explored parallel independent component analysis (paraICA) as a new method for analyzing multimodal data. The method was aimed to identify simultaneously independent components of each modality and the relationships between them. When 43 healthy controls and 20 schizophrenia patients, all Caucasian, were studied, we found a correlation of 0.38 between one fMRI component and one SNP component. This fMRI component consisted mainly of parietal lobe activations. The relevant SNP component was contributed to significantly by 10 SNPs located in genes, including those coding for the nicotinic α-7cholinergic receptor, aromatic amino acid decarboxylase, disrupted in schizophrenia 1, among others. Both fMRI and SNP components showed significant differences in loading parameters between the schizophrenia and control groups ($P = 0.0006$ for the fMRI component; $P = 0.001$ for the SNP component). In summary, we constructed a framework to identify interactions between brain functional and genetic information; our findings provide a proof-of-

concept that genomic SNP factors can be investigated by using endophenotypic imaging findings in a multivariate format. Hum Brain Mapp, 2007. © 2007 Wiley-Liss, Inc.

Received: 4 January 2007; Revised: 27 July 2007; Accepted: 12 September 2007

DIGITAL OBJECT IDENTIFIER (DOI)

10.1002/hbm.20508 About DOI

ADDITIONAL MATERIAL

This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/1065-9471/suppmat .

ARTICLE TEXT

INTRODUCTION

Brain imaging techniques such as positron emission tomography, magnetic resonance imaging, magnetoencephalography, and electroencephalography are being applied increasingly to study the structure and function of the human brain in health and disease. More recently, the combination of imaging data (which may represent candidate endophenotypes or disease vulnerability markers) with genetic information appears to be a valuable way to study particular subset of polymorphisms which may have functional consequences. This approach is especially important in the study of schizophrenia, where even identical twins have only about 50% concordance for the disease [Lee et al.,[2005]]. Hence, it is likely that both genetic and endophenotypic information will be needed to comprehensively understand the disorder. Recent results in schizophrenia suggest that disturbed working memory and associated prefrontal cortex activation may be two genetically influenced markers for vulnerability to the illness [Brahmbhatt et al.,[2006]]. However, relatively little is known about the genetic determinants of these dysfunctions.

Most genetic linkage studies focus on particular gene(s) or single nucleotide polymorphism(s) (SNPs) of interest and examine the relationship between a genotype and a phenotype univariately. Using this approach, a number of genes have been associated with increased vulnerability for schizophrenia [Carter,[2006]; Harrison and Owen,[2003]]. For example, the association of the G72/G30 locus with schizophrenia and bipolar disorder has been reported in several studies [Detera-Wadleigh and McMahon,[2006]; Goldberg et al.,[2006b]]. Other genes or SNPs frequently studied in schizophrenia include catechol-*O*-methyltransferase (COMT) [Ho et al.,[2005]; Numata et al.,[2007]; Ohnishi et al.,[2006]], brain-derived neurotrophic factor (BDNF) [Bath and Lee,[2006]; Ho et al.,[2006]; Numata et al.,[2007]], and disrupted in schizophrenia 1 (DISC1) [Derosse et al.,[2007]; Porteous et al.,[2006]]. Furthermore, recent studies demonstrate that specific polymorphisms in COMT, BDNF, and DISC1, among others, are associated with subtle but consistent alterations in the patterns of brain activation, cognitive function, and clinical symptoms in patients [Meyer-Lindenberg and Weinberger, [2006]; Roffman et al.,[2006]].

There is substantial evidence that schizophrenia is likely to be a complex genetic disorder, with multiple risk genes of individually weak effect [Owen et al.,[2005]]. Therefore, a multivariate approach, capable of extracting hidden cross-information from a larger number of genes, has the potential to uncover influences of multiple genetic factors on the functioning of both healthy and disordered brains. From the genomic perspective, more recent work has focused on screening large numbers of SNPs and defining association in terms of ensembles of multiple SNPs. This physiogenomic approach [Ruano et al.,[2005b] has led to important discoveries of genetic markers in the fields of cardiology [Derosse et al.,[2007]; Ruano, et al.,[2006a]; Ruano, et al.,[2005a]], nutrition [Ruano et al.,[2006b]], and psychiatry [Ruano et al., [2006c]].

In this paper we propose a novel physiogenomic approach to simultaneously analyze brain images and genetic information from both schizophrenia patients and healthy controls. An association of 367 linearly weighted SNP genotypes, as one independent genetic factor, is assumed to influence to a certain degree one independent brain function - functionally connected brain regions with activations - extracted from fMRI data. The brain function here is considered as a potential endophenotype, and the goal is to find the relationship between the genetic association and brain function.

Independent component analysis (ICA) is a statistical and computational technique for recovering hidden independent factors or components underlying sets of random variables, measurements, or signals. An advantage of ICA is its capability to reveal factors without the need for specific prior knowledge of the properties of these factors. ICA has been applied to the analysis of functional magnetic resonance image (fMRI) data in order to discover hidden components presenting (in the case of spatial ICA) brain activation in certain brain regions [Calhoun and Adali,[2006]; Calhoun et al., [2001a],[c],[2006]; McKeown and Sejnowski,[1998]; McKeown et al.,[1998]]. ICA can also be used to analyze spatial patterns from activation maps computed using a general linear model based upon paradigm timing [Calhoun et al.,[2006]]. Similarly, ICA is also a reasonable approach to discover the unknown but important information stored in genetic arrays. Researchers have utilized ICA [Lee and Batzoglou,[2003]; Liebermeister,[2002]] or principal component analysis [Horne and Camp,[2004]; Lin and Altman,[2004]] to identify or to select genes. Candidate gene association studies often utilize

SNP data. Studies selecting optimal SNP sets to capture intragenic genetic variation or tag-specific haplotypes have employed principal component analysis [Horne and Camp,[2004]; Lin and Altman,[2004]]. Dawy et al. proposed an ICA-based algorithm to map SNPs to a certain phenotype, assuming that SNP expressions affecting a given phenotype are independent sources transformed by a linear mixing process [Dawy et al.,[2005]].

We present an approach for revealing relationships between brain function and SNP groupings, i.e., to find a combination of SNPs related to a functional brain network. This approach involved solving three problems: revealing a set of specific independent brain functions, identifying independent SNP associations, and finding the relationship between SNP associations and brain functions. We develop an approach, called parallel ICA, with constraints applied directly to two modalities. This method allows independent components from two modalities to be identified simultaneously, while a term relating the two modalities is emphasized. We apply this approach to fMRI data collected during an auditory oddball task and to SNP data. Both data types were collected from each of 63 subjects, including 20 patients with schizophrenia and 43 healthy controls.

THEORY

Introduction to ICA

The basic ICA model shown in Eq. (1) defines a generative model for the observed data, which are typically given as a large database of samples. The observed variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed non-Gaussian (or only one Gaussian) and mutually independent and they are called independent components of the observed data. In Eq. (1), X is an observation matrix that can be composed of measurements such as speech signals, MRI images, or SNP genotypes. S contains the independent components, which consists of unknown sources such as multiple speakers' voices, brain activation networks, or genetic associations for various phenotypes. A is a linear mixing matrix, relating the sources to the contaminated measurements. W is an unmixing matrix. If W equals the inverse of A, then the Z, the estimated component matrix, is equivalent to S, the source matrix. Therefore, the essence of ICA is to find W so that Z is as close as possible to the true independent components contained in S. There are many ICA algorithms based on different independence criteria. Among them, the Infomax algorithm attempts to find the W matrix through maximizing an entropy function as defined in Eq. (2) [Bell and Sejnowski,[1995]; Cardoso,[1997]]. Before applying maximization, principle component analysis is used to reduce the dimensionality of the observation matrix X down to the same dimension as the component matrix. Hereafter, principle component analysis is always implemented before proceeding ICA.

$$X = A \cdot S; \qquad Z = W \cdot X; \qquad \text{If } W = A^{-1}, \quad \text{then } Z = S; \tag{1}$$

$$\max\{H(Y)\} = -E[\ln f_y(Y)]; \qquad Y = \frac{1}{1+e^{-U}}, \quad U = W \cdot X + W_0 \tag{2}$$

where $f_y(Y)$ is the probability density function of Y. E is the expected value. H is the entropy function.

Parallel ICA Structure

The purpose of parallel ICA proposed here is to discover independent components from two modalities, in addition to the relationship between them, as illustrated in Figure 1. S_1 represents the independent components estimated from data X_1 via an unmixing matrix W_1; S_2 represents independent components estimated from data X_2 via an unmixing matrix W_2. At the same time, a term specific to the relationship between S_1 and S_2 is built into the W_1 and W_2 matrices. Each unmixing matrix for both modalities is thus computed using information derived from both modalities. The relation between the two modalities can be specified by the user under a variety of rationales, and it can be correlated between the W matrices, other similarity measures, or any other relationship with reasonable interpretation.

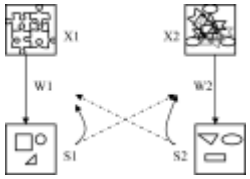

Figure 1. Parallel ICA structure.
[Normal View 18K | Magnified View 41K]

Three problems need to be solved simultaneously in parallel ICA. Two of them relate to maximizing the independence between components for the two modalities, respectively. The third is the determination of the relationship between the two modalities. For example, in this study we try to find the correlation between the column vector of A_1 from one modality, and

the column vector of A_2 from the other modality, as described in Eq. (3) (our choice of this relation is explained in Method section). For simplicity of explanation, only one component from each side is constrained.

$$\mathrm{Corr}(A_{1i}, A_{2j}) = \frac{\mathrm{Cov}(A_{1i}, A_{2j})}{\mathrm{Std}(A_{1i}) \cdot \mathrm{Std}(A_{2j})}; \quad A_1 = W_1^{-1} \tag{3}$$

where Corr is the correlation function; Cov is the covariance function. Std is the standard deviation function. *i* and *j* are indices of components.

Parallel ICA Optimization

Our algorithm is based upon the Infomax algorithm; hence, maximization of the mutual entropy is used to maximize the independence between components. The relationship between modalities is determined by adding a term which maximizes the squared correlation. The final maximization is shown in Eq. (4),

$$\begin{aligned}
&\max\left\{H(Y_1) + H(Y_2) + \mathrm{Corr}(A_1, A_2)^2\right\} \\
&= \left\{-E[\ln f_y(Y_1)] - E[\ln f_y(Y_2)] + \frac{\mathrm{Cov}(A_{1i}, A_{2j})^2}{\mathrm{var}(A_{1i}) \cdot \mathrm{var}(A_{2j})}\right\}; \\
&Y_1 = \frac{1}{1+e^{-U_1}}, \quad U_1 = W_1 \cdot X_1 + W_{10}, \quad A_1 = W_1^{-1}; \\
&Y_2 = \frac{1}{1+e^{-U_2}}, \quad U_2 = W_2 \cdot X_2 + W_{20}, \quad A_2 = W_2^{-1};
\end{aligned} \tag{4}$$

where *i* and *j* indicate the constrained components selected during every maximization iteration. These two indices can vary along the maximization process. Thus, the algorithm is adaptive to the continuing updated components.

The three terms in maximization function [Eq. (4)] have different characteristics; to maximize two entropies equally, we simply maximize the first two terms in parallel with two learning rates, using the natural gradient maximization [Amari, [1998]; Yang and Amari,[1997]]. The third term is optimized using the steepest descent method, and the step size is calculated at each iteration on the selected two components. Thus, we arrive at the following update rules:

$$\begin{aligned}
&\text{For the first term: } \Delta W_1 = \frac{\partial H_1}{\partial W_1} = \lambda_1 \cdot \left[I + (1 - 2Y_1)U_1^T\right]. \\
&\text{For the second term: } \Delta W_2 = \frac{\partial H_2}{\partial W_2} = \lambda_2 \cdot \left[I + (1 - 2Y_2)U_2^T\right]. \\
&\text{For the third term: } \Delta A_{1i} = \frac{\partial \mathrm{Corr}(A_{1i}, A_{2j})^2}{\partial A_{1i}} \\
&\quad = \lambda_{c1} \cdot \eta_1 \cdot \frac{2\mathrm{Corr}(A_{1i}, A_{2j})}{\mathrm{Std}(A_{2j})\mathrm{Std}(A_{1i})} \\
&\qquad \times \left\{(A_{2j} - \overline{A_{2j}}) + \frac{\mathrm{Cov}\,(A_{1i}, A_{2j})\,(\overline{A_{1i}} - A_{1i})}{\mathrm{Var}\,(A_{1i})}\right\}; \\
&\Delta A_{2j} = \frac{\partial \mathrm{Corr}\,(A_{1i}, A_{2j})^2}{\partial A_{2i}} \\
&\quad = \lambda_{c2} \cdot \eta_2 \cdot \frac{2\mathrm{Corr}\,(A_{2j}, A_{1i})}{\mathrm{Std}(A_{2j})\mathrm{Std}(A_{1i})} \\
&\qquad \times \left\{(A_{1i} - \overline{A_{1i}}) + \frac{\mathrm{Cov}\,(A_{1i}, A_{2j})\,(\overline{A_{2j}} - A_{1j})}{\mathrm{Var}(A_{2j})}\right\}
\end{aligned} \tag{5}$$

where the λs are the learning rates for Data 1, Data 2, and correlation terms, and the η is the step size calculated at each step according to Wolfe conditions [Nocedal and Wright,[1999]]. The learning rates, determining the emphasizing weight put onto each term during the maximization, play important roles in the convergence and balance.

Overfitting Issue

Any data-driven approach possibly encounters a problem with overfitting due to too many parameters or too strong overlearning, which can lead to false discovery. We use additional techniques to avoid overfitting that may be caused by two possible reasons in our algorithm: over-emphasizing correlations and overestimating component number.

To avoid an overemphasized correlation, we adaptively adjust the learning rate of the correlation term [λ_c in Eq. (5)] in the maximization function. By monitoring the entropy term $H(.)$ online, we can, to some degree, assess the overall effect of connection term on the total maximization function. Based on the level in which the $H(.)$ term is altered, we change the λ adaptively to balance the two aspects in the maximization function.

Estimating the correct number of components is still considered a challenge in the field of blind source separation. The Akaike information criterion (AIC), an information theoretic approach for determining data dimensionality [Akaike,[1974]], is a reasonable approach widely used in different areas of research. However, it tends to overestimate component number. To eliminate overestimated component influence, we use modified AIC method proposed by Li et al., [Li et al.,[2007]] for fMRI data, through a subsampling scheme to obtain a set of effectively i.i.d. samples to compensate the spatial smoothness in fMRI images. Owing to no ready-to-use method for the SNP data, we first use AIC method to estimate components' number, and then, reduce the component number cautiously and empirically to reach a consistence level among different runs. The principle we applied in the reduction process is similar to a leave-one-out cross-evaluation, where all samples except one are used in different runs and a consistent result is arrived using reasonable parameters.

Simulations presented later will give a clear picture of the algorithm performance against overfitting under vulnerable conditions.

SUBJECTS AND DATA SPECIFICATION

In this study, two types of data were collected from 63 participants, including 20 patients with schizophrenia and 43 healthy controls. All of them provided written, informed, IRB-approved consent at Hartford Hospital. fMRI data were used to understand brain functions and SNP data were used to find genetic influences.

Participants

Participants were recruited via advertisement, presentations at local universities and clinics, and by word-of-mouth. Prior to inclusion in the study, healthy participants were screened to ensure they were free from DSMIV Axis I or Axis II psychopathology [assessed using the SCID [Spitzer et al.,[1996]] and also interviewed to determine that there was no history of psychosis in any first-degree relatives. Patients met criteria for schizophrenia in the DSM-IV on the basis of the Structured Clinical Interview for DSM IV [First et al.,[1995]] and review of the case file. All selected subjects were White/non-Hispanic. The patients were 39 ± 10 years old, ranging from 20 to 54, and the healthy controls were 48 ± 16 years old ranging from 21 to 83. There were 21 female and 22 male participants in the 43 healthy controls, and 3 female and 17 male Schizophrenia (SZ) patients. To include as many subjects as possible in our study, we decided to use all subjects, instead of the relatively balanced subset, while assessing the effect of age, gender, and other factors on our measurements. Full scale IQ scores for the groups evaluated from the National Adult Reading Test (NART) [Blair and Spreen,[1989]] were 110.9 ± 6.6 for controls and 104.9 ± 10.1 for patients (only 40 controls and 19 patients had NART scores recorded). Positive and Negative Syndrome Scale (PANSS) scores for 16 SZ patients (the other four patients' PANSS scores were missing) were also observed, with PANSS total scores of 67.6 ± 30.0, positive symptom scores of 15.4 ± 4.1, and negative symptom scores of 14.5 ± 6.7. The 17 SZ patients who provided medication information were taking 10 different antipsychotic in variable doses.

fMRI Data Collection

FMRI data were collected during performance of an auditory oddball task [Kiehl and Liddle,[2003]], which consists of detecting an infrequent sound within a series of frequent sounds. Auditory stimuli were presented to each participant by a computer stimulus presentation system via earphones. The standard stimulus was a 500-Hz tone, the target stimulus was a 1,000-Hz tone, and the novel stimuli consisted of nonrepeating random digital noises (e.g., tone sweeps, whistles). A full description of the task design is available [Kiehl et al.,[2005]]. The participants were instructed to respond as quickly and accurately as possible with their right index finger on a keypad every time they heard the target stimulus.

Scans were acquired at the Olin Neuropsychiatry Research Center at the Institute of Living on a Siemens Allegra 3 T dedicated head MRI scanner equipped with 40 mT/m gradients and a standard quadrature head coil. The functional scans were acquired using gradient-echo echo-planar-imaging with the following parameters (repeat time = 1.50 s), echo time = 27 ms, field of view = 24 cm, acquisition matrix = 64 × 64, flip angle = 70°, voxel size = 3.75 × 3.75 × 4 mm^3, slice thickness = 4 mm, gap = 1 mm, 29 slices, ascending acquisition).

fMRI Data Preprocessing

Six "dummy" scans were performed at the beginning to allow for longitudinal equilibrium, after which the paradigm was automatically triggered to start by the scanner. Data were preprocessed using the software package SPM2 (http://www.fil.ion.ucl.ac.uk/spm/). Images were realigned using INRIalign - a motion correction algorithm unbiased by local signal changes [Freire and Mangin,[2001]]. Data were spatially normalized into the standard Montreal Neurological Institute space [Friston et al.,[1995]], resliced to 3 mm^3, and spatially smoothed with a 10-mm^3 Gaussian kernel. Data for each participant were analyzed by a multiple regression incorporating regressors for the novel, target, and standard and

their temporal derivatives plus an intercept term. The target-related contrast images were used in this study for parallel ICA. To balance the size of one fMRI image and one SNP image, containing the SNP genotypes from one subject, we used a mask based upon one-sample *t*-test against zero activation to select meaningful voxels, and downsampled the images by a factor of 2. The resultant images with a size of 7,060 voxels were the input from fMRI modality to parallel ICA.

SNP Data Collection and Preprocessing

A blood sample was obtained for each subject and DNA extracted. Genotyping was performed using the Illumina BeadArray™ platform and the GoldenGate™ assay [Fan et al.,[2003]; Oliphant et al.,[2002]]. The PG Array of Genomas was used, which contains a SNP array consisting of 384 SNPs from 222 genes from six physiological systems: neurobiology, cardiovascular system, inflammation, metabolism, cholesterol biochemistry, and cell proliferation. The following pathways were represented: insulin resistance, glucose metabolism, energy homeostasis, adiposity, apolipoproteins and receptors, fatty acid and cholesterol metabolism, lipases, receptors, cell signaling and transcriptional regulation, growth factors, drug metabolism, blood pressure, vascular signaling, endothelial dysfunction, coagulation and fibrinolysis, vascular inflammation, cytokines, neurotransmitter axes (serotonin, dopamine cholinergic, histamine, glutamate), and behavior (satiety). The PG Array is a product of Genomas, Inc. and its detailed composition has been published as a patent application. Genotyping analysis software, GenCall, was used to cluster the resultant intensities from the genotyping microarray into three clusters: AA, AB, and BB without assuming dominant or recessive inheritance. On the basis of the GenCall score, a number between 0 and 1 indicating how close to the center of the cluster a sample lies, we set up a threshold to select only reliable genotype results. SNPs with a GenCall score of 0.25 or higher were selected resulting in 367 SNPs. Genotypes are inherently categorical and can be represented as discrete numbers, e.g., 1 for one type of homozygous, 0 for heterozygous, and -1 for the other type of homozygous. Negative and positive signs are not important in our test, since we look at variation of genotypes, and signs can be switched by the mixing matrix based on effects on the phenotype.

METHOD

Now, we apply the parallel ICA onto the two modalities described earlier, with the goal of identifying functional brain networks, SNP associations, and their relationship. Components extracted from fMRI can be interpreted as networks of brain regions that express functional changes in different subjects to different degrees. Components extracted from SNP data are distinct linear combinations of SNPs that may affect certain generic functionalities or phenotypes [Dawy et al., [2005]; Lin and Altman,[2004]]. These components are expressed to different degrees in different subjects. The loading parameters for each component reflect the component's influence/expression on subjects [Calhoun et al.,[2001b]]. Relationships between the two modalities can be incorporated as well. For example, if a component extracted from fMRI data is functionally related to a component extracted from SNP data - in other words, an association of SNPs has functional consequences revealed in a specific fMRI brain network - then we expect that the influence/expression pattern of these two components in all participants is correlated. Hereafter, the loading parameters are used in the paper to address subjects' component patterns.

To demonstrate the algorithm, we implemented the method on a dataset including schizophrenia patients and healthy controls. Our goals were to first identify connections between fMRI and SNP and then to investigate group differences. In other words, after we found the linked components between brain functions and genetic associations, we tested for a significant difference between schizophrenia and healthy groups in the loading parameters of the components.

Parallel ICA for fMRI and SNP Data

We represented fMRI data collected from the participants as a set of spatially independent voxels which are linearly mixed [Calhoun et al.,[2004],[2006]). Hence, the *X* and *S* matrices in Eq. (1) are constructed as follows:

$$X_f = [x_{f1}, x_{f2}, x_{f3}, \ldots, x_{fn}]^T; \quad S_f = [s_{f1}, s_{f2}, s_{f3}, \ldots, s_{fm}]^T;$$
$$S_f = W_f \cdot X_f; \qquad (6)$$
$$A_f = W_y^{-1}; \qquad A_f = [a_{f1}, a_{f2}, \ldots, a_{fm}];$$

where *n* is the number of participants and *m* is the number of components. x_{fi} and s_{fi} are both vectors of voxel values in brain images. The A_f matrix is a participant-by-component ($n \times m$) mixing matrix. Each column of A_f shows information about how a component appears in every participant. The *p*th column of A_f matrix, for example, contains a loading parameter/influence for the *p*th component for each of the *n* participants.

We defined a genetic independent component as a specific SNP association, i.e., a group of SNPs with various degrees of contribution, which partially determines a specific phenotype or endophenotype. This association can be modeled as a linear combination of SNP genotypes [Dawy et al.,[2005]; Lee and Batzoglou,[2003]],

$$s = \beta_1 \cdot \mathrm{snp}_1 + \beta_2 \cdot \mathrm{snp}_2 + \cdots + \beta_{n1} \cdot \mathrm{snp}_n;$$

where snp is a genotype at a given locus and β is a weight contributed from a SNP to the genetic association. Beside the independent component, the weight itself is also of interest, implying the influence factor and type, i.e., inhibitory or excitory to a phenotype. With the assumption that each component has a independent distribution pattern in 367 SNPs, we constructed the SNP data matrix, *X*, in a participant-by-SNP direction. The mixing process is presented in Eq. (7),

$$X_s = [x_{s1}, x_{s2}, x_{s3}, \ldots, x_{sn}]^T, \quad S_s = [s_{s1}, s_{s2}, s_{s3}, \ldots, s_{sm}]^T;$$
$$S_s = W_s \cdot X_s; \qquad (7)$$
$$A_s = W_s^{-1}; \quad A_s = [a_{s1}, a_{s2}, a_{s3}, \ldots, a_{sn}]^T$$

where *n* is the number of participants and *m* is the number of components. x_{si} is a vector of 367 SNP genotypes for one participant. s_{si} is a vector of 367 SNP weights for one genetic component. A_s is the matrix of the loading parameters, presenting the influence of each SNP component on participants.

Under current data structure, the relationship between the influences/presences of brain function and genetic component is calculated as the correlation between the columns of the fMRI A_f matrix and the SNP A_s matrix. Thus, we have the correlation term defined as in Eq. (3), and the maximization function defined as in Eq. (5), where Data 1 is the fMRI and Data 2 is the SNP. The procedure of parallel ICA is illustrated in Figure 2.

Figure 2. Parallel ICA procedure.
[Normal View 8K | Magnified View 21K]

The algorithm proceeds as follows:

1 Two analyses for fMRI and SNP, respectively, are initialized with specified learning rates λ_f, λ_s, λ_{cf}, and λ_{cs}.

2 Two *W* matrices, if necessary, are updated based on their own entropy terms.

3 The stop criterion is assessed for both modalities. If both processes satisfy the criterion, the whole parallel ICA process stops. If only one process satisfies the criterion, then the iteration for this modality stops and the corresponding *W* matrix is finalized.

4 One component from each modality is selected to be the related components according to the highest correlation

5 Two *W* matrices, if necessary, are updated based on the correlation terms. Afterward, the process goes back to step 2.

To avoid false discoveries resulting from overfitting, a leave-one-out evaluation is implemented to test the fidelity of discoveries. Because of the limited subject numbers, 62 out of 63 subjects were analyzed during 63 runs by parallel ICA with the same parameter setup, and each run includes one different subject. The consistency among the 63 repetitions was evaluated.

Two-Group Comparison

Since the participants were composed of patients and healthy controls, a two-group comparison exploring between-group differences informed us whether the components were schizophrenia-relevant (SZ-relevant refers to any direct or indirect connection to SZ). As described earlier, each column of *A* matrix reveals the presence pattern of one component in all participants. A two-sample *t*-test is conducted on the fMRI-SNP correlated column vectors of A_s and A_f, with consideration of other possible factors, such as gender, IQ, and schizophrenia symptom severity. We examined each factor's contribution to the linked components' loading parameters using linear regression.

SIMULATION

To evaluate the performance of the parallel ICA algorithm, we simulated two datasets with the same dimensionalities as the fMRI and SNP data, respectively. Eight source signals with different distributions (one Gaussian) were included for each dataset separately (an example of simulated sources were plotted in Fig. 3), as well as two random mixing matrices. One source from the fMRI data and one source from the SNP data were correlated by making one column of the fMRI mixing matrix similar to one column of the SNP mixing matrix to a certain degree. Random Gaussian noise was superimposed into the mixed source data afterwards, as explained in Eq. (8). The accuracy of recovering both the true sources and the correlation was assessed under different true connection conditions.

$$\begin{aligned}\text{Data } 1 &= A'(43-\text{by}-43)\cdot S'(43-\text{by}-8000)\\ &\quad +\text{noise}(43-\text{by}-8000);\\ \text{Data } 2 &= A''(43-\text{by}-43)\cdot S''(43-\text{by}-367)\\ &\quad +\text{noise}(43-\text{by}-367);\end{aligned}$$

8

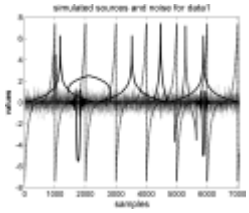

Figure 3. Simulated source singles from Data 1 as well as the noise superimposed.
[Normal View 28K | Magnified View 83K]

Since in reality the component number is usually unknown and an estimated component number plays an important role in algorithm performance, we evaluated our algorithm under different estimated component numbers.

RESULTS

In this section we first present the simulation results followed by the results uncovered from 63 subjects' fMRI and SNP data.

Simulation Results

The accuracy used hereafter is defined as the correlation between the true source(s) and the estimated component(s). An example of true related sources and corresponding extracted components is shown in Figure 4, when the true correlation is 1.0 and the component number is set to 8. We can tell whether the estimated components represent the true sources very well with the exception of scale. In order to see the improvement of parallel ICA performance, we compared it with standard ICA, applied to these two datasets separately. The results from both methods under different conditions are presented in Table I.

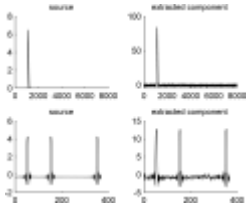

Figure 4. True related sources (left) and corresponding components (right).
[Normal View 8K | Magnified View 22K]

Table I. Simulation results from parallel ICA and standard ICA

True correlation	1.00	0.80	0.60	0.40	0.20
Parallel ICA					
Extracted correlation	0.97 ± 0.01	0.79 ± 0.02	0.60 ± 0.06	0.42 ± 0.07	0.20 ± 0.03
Accuracy of SNP	0.90 ± 0.00	0.90 ± 0.00	0.89 ± 0.00	0.99 ± 0.01	X*
Accuracy of fMRI	0.99 ± 0.00	0.99 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00
Standard ICA					
Extracted correlation	0.95 ± 0.01	0.77 ± 0.03	0.55 ± 0.04	0.37 ± 0.03	0.20 ± 0.03
Accuracy of SNP	0.90 ± 0.00	0.90 ± 0.00	0.89 ± 0.00	0.99 ± 0.00	X*
Accuracy of fMRI	0.99 ± 0.00	0.99 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00

X* is not the right component.

The ability of parallel ICA to determine the optimal result varies under different conditions as shown in Table I. The correlation determined is slightly lower than the true correlation when the latter is high. When the true correlation is low

(when it can be considered as noise effect), parallel ICA does not extract a false connection. Comparing this approach to standard ICA, we can clearly see the improvement of parallel ICA in terms of correlation discovery, when such a correlation exists.

We also investigated the performance of the algorithm when using an incorrect component number. While eight true components are included in each dataset, an over/under estimated number is applied in the parallel ICA, and the corresponding results are listed in Table II. The simulation shows that an overestimated component makes the algorithm more vulnerable to overfitting with lower accuracy of the component and higher correlation, and a moderately underestimated component number does not markedly influence the correlated particular results, illustrated in Table II. Thus, we would rather use an underestimated component number than an overestimated component number.

Table II. Simulation results from parallel ICA using different component number

Component numbers	2/2	4/4	6/6	8/8	10/10	12/12	14/14
Extracted correlation	0.44 ± 0.05	0.48 ± 0.05	0.48 ± 0.04	0.50 ± 0.04	0.57 ± 0.05	0.60 ± 0.05	0.58 ± 0.07
Accuracy for the SNP source	0.95 ± 0.00	0.91 ± 0.04	0.91 ± 0.05	0.89 ± 0.04	0.84 ± 0.04	0.81 ± 0.05	0.55 ± 0.34
Accuracy for the fMRI source	0.97 ± 0.03	0.99 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.68 ± 0.45

True correlation: 0.5; True component number: 8/8.

Results From 63 Subjects' fMRI and SNP Data

The modified AIC estimate [Li et al.,[2007]] of the number of components for the 63 subjects' reduced fMRI data are 5. Twelve components are estimated for the SNP data by the regular AIC. Then, the numbers are further reduced to 7, based on the principle mentioned in Overfitting Issue section. Results include components for the fMRI and SNP data, and their loading parameters.

Results from parallel ICA

Parallel ICA revealed a correlation of 0.38 between one fMRI component and one genetic component. For explanation/display purposes, significant SNPs in this linked SNP component and high activation regions of the linked fMRI component are only presented. The fMRI component, after converting to *Z* scores, is thresholded at $|Z| > 2.5$. Similarly, SNPs with contribution weights above 2.5 are selected and listed, representing the genetic component. The chosen fMRI component is plotted in Figure 5 ($|Z| > 2.5$) and active regions are listed in Table III.

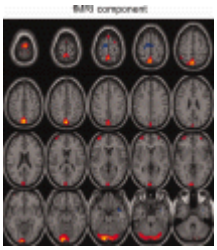

Figure 5. The linked fMRI component discovered by parallel ICA. [Normal View 42K | Magnified View 134K]

Table III. Talairach label of regions of interest

Area	Brodmann area	*L*/*R* volume (mL)	*L*/*R* random effects: max *Z* (*x*, *y*, *z*)
Positive			
Precuneus	7 19	0.8/0.1	4.8(0,- 70,50)/3.4(6, -76,51)
Lingual gyrus	18 17	0.7/0.2	4.7(-12, -85, -13)/3.9(6, -85, -13)
Cuneus	17 18 19	0.5/0.0	4.7(-12, -96, -3)/NA
Fusiform gyrus	18 19	0.2/0.2	5.1(-18, -91, -13)/3.7(24, -85, -18)

Superior parietal lobule	7	0.2/0.1	4.5(-6, -70,56)/2.7(6, -70,56)
Postcentral gyrus	5 7	0.2/0.1	4.2(-6, -46,71)/3.0(6, -40,71)
Inferior occipital gyrus	17 18	0.1/0.0	5.0(-12, -91, -8)/3.3(30, -85, -13)
Negative			
Superior frontal gyrus	6	0.1/0.0	2.8(-12, -11,64)/NA
Medial frontal gyrus	6	0.1/0.0	2.5(-12, -6,58)/NA
Superior temporal gyrus	38	0.0/0.1	NA/2.7(36,7, -28)

Significant SNPs in the selected genetic component are listed with their information in Table IV. Cross correlations among these SNPs in terms of genotype patterns on all participants were also calculated. Among them, rs3087454 and rs1355920, both in the α7 nicotinic cholinergic receptor (CHRNA7), are strongly correlated with a correlation coefficient of 0.53. The rs821616 in the chromosome 1 DISC1 gene and rs4765623 in the chromosome 12 scavenger receptor class B, member 1 (SCARB1) gene are also highly correlated with a correlation of 0.47. Presently, it is not clear why these two SNPs may be in linkage disequilibrium and further studies will be needed to explain this association. The recent finding that the same SNPs are not correlated in patients with diabetes suggests that the link between these SNPs may be specific to schizophrenia (A. Windemuth, unpublished observation). Finally, we observed correlations of 0.28 between rs1355920 and rs885834 and between rs4765623 and rs2071521.

Table IV. Significant SNPs and their information extracted by parallel ICA

SNP	*Z* score	Gene
rs1466163	-4.08	AADC: aromatic **L**-amino acid decarboxylase
rs2429511	3.97	ADRA2A: α-2A adrenergic receptor gene
rs3087454	-3.09	CHRNA7: cholinergic receptor, nicotinic, α7
rs821616	2.96	DISC1: disrupted in schizophrenia 1
rs885834	-2.78	CHAT: choline acetyltransferase
rs1355920	-2.77	CHRNA7: cholinergic receptor, nicotinic, α7
rs4765623	2.73	SCARB1: scavenger receptor class B, member 1
rs4784642	-2.71	GNAO1: guanine nucleotide binding protein (G protein), α activating activity polypeptide O
rs2071521	2.58	APOC3: apolipoprotein C-III
rs7520974	2.55	CHRM3: cholinergic receptor, muscarinic 3

Out of 63 runs on the leave-one-out evaluation data, the connection between these fMRI (see Fig. 5) and SNP components is 0.37 ± 0.07.

Patients versus controls comparison

The loading parameters of the selected SNP/fMRI component were also studied for the patient group versus control group difference. Figure 6 shows both components' loading parameters in all participants, with 20 SZ first and followed by 43 healthy controls. A two-sample *t*-test was also computed on the loading parameters: the fMRI component showed group differences with $P = 0.0006$ and the SNP component with $P = 0.001$. Potential factors contributing to this genetic/phenotypic pattern include gender, IQ, age, and schizophrenic severity. Percentages of variance explained by these factors are listed in Table V, after linear regression. A significant between-group difference exists, providing evidence that this SNP/fMRI component is schizophrenia-relevant.

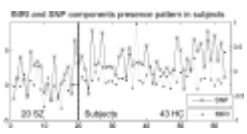

Figure 6. fMRI and SNP loading parameters.
[Normal View 21K | Magnified View 58K]

Table V. Percentage of variance explained by potential factors

r^2	SNP (%)	fMRI (%)
IQ	0.16	0.29
Gender	0.65	1.02
Age	2.14	0.04
PANSS (total)	7.56[a]	32.15[a]
PANSS (positive)	20.96[a]	9.00[a]
PANSS (negative)	19.19[a]	44.92[a]

[a] Only 16 SZ patients.

ANALYSIS AND DISCUSSION

Correlation

With leave-one-out cross-evaluation, we can reasonably say that a relationship exists between fMRI and genetic data as revealed by parallel ICA. Therefore, the connection between brain function and genes is presented as a pair of fMRI and genetic components, whose correlation value was 0.38. The fMRI component revealed brain regions whose functions are explained in the following section, and the genetic component is associated with genes previously known to influence specific brain functions or behaviors. The potential interplay of these two components is interpreted as a possible genetic influence on brain function. However, because of limited SNP types and subjects, the results are preliminary, based on the data we had, but suggest the direction of future studies.

fMRI Component

The selected fMRI component consists of many regional activations. However, the regions with higher *Z* scores listed in Table III are of great interest.

The largest portion of this component is located in precuneus. In a review paper, Cavanna et al. summarized functional subdivision of precuneus into an anterior region, involved in self-centered mental imagery strategies, and a posterior region, subserving successful episodic memory retrieval [Cavanna and Trimble,[2006]]. Gray matter volumes were reduced in all parietal subregions in schizophrenia patients compared to healthy controls (consistent with prior studies, e.g. [Frederikse, et al.,[2000]] and the volume alterations in schizophrenia may support the notion that a regional posterior parietal deficit is critical for the manifestation of overt psychotic symptoms [Zhou et al.,[2007]]. The second region identified was lingual gyrus, where a volumetric alternation in SZ patients has been noted in several studies [Gaser et al.,[1999]]; in one of these, the predictive power of parietal activation was supported only in combination with lingual gyrus activity [Whalley et al.,[1999]].

A third region identified was the cuneus, which is important in memory retrieval [Addis et al.,[2004]; Cabeza et al.,[1997]; Cavanna and Trimble,[2006]]. The cuneus appears in several schizophrenic studies [Kircher et al.,[2003]; Neckelmann et al.,[2006]], but to our knowledge has not yet been intentionally investigated from a SZ relevant viewpoint.

Three deactivated regions constituted this fMRI component: superior frontal, medial frontal, and superior temporal gyri. The superior frontal gyrus [Goldberg et al.,[2006a]] is involved in self-awareness and executive functions. The medial frontal gyrus is associated with high-level executive functions and decision-related processes. Since our input fMRI data were contrast images (target stimulus) collected in the auditory oddball test. It is not surprising to see involvement of a portion of superior temporal gyrus, which includes auditory processing regions and has been implicated in schizophrenia, particularly in reference to auditory hallucinations [Barta et al.,[1990]; Pearlson et al.,[1996]].

Genetic Component

The related genetic component was contributed to by 10 SNPs located in nine genes: aromatic L-amino acid decarboxylase (AADC), α-2A adrenergic receptor gene (ADRA2A), CHRNA7, DISC1, choline acetyltransferase (CHAT), SCARB1, apolipoprotein C-III (APOC3), muscarinic cholinergic receptor, 3 (CHRM3). CHRNA7, DISC1, and CHAT are well-known schizophrenia susceptibility genes.

CHRNA7 is a member of a superfamily of ligand-gated ion channels that mediate fast cholinergic transmission at synapses. This gene is located in a chromosomal location involved in the genetic transmission of schizophrenia [De Luca et al.,[2004]; Freedman et al.,[2001]]. In our study, two SNPs in the CHRNA7 gene, rs3087454 and rs1355920, were found to correlate with patterns of brain activation during an auditory oddball task. These results are consistent with previous studies linking polymorphisms in CHRNA7 promoter region with sensory gating alterations in patients with SZ, as measured by the P50 inhibition in auditory evoked response [Freedman et al.,[1997]; Leonard et al.,[2002]]. The associations of CHRNA7 with schizophrenia vary in different ethnic groups. Studies on populations in France, Switzerland, and USA reported positive results [Freedman et al.,[2006]; Houy et al.,[2004]; Leonard et al.,[2002]; Stassen et al.,[2000]], and studies on population in China and additional populations in USA reported negative results [Fan et al.,[2006]; Gault et al.,[2003]]. Our analyses of Caucasian patients further support the idea that CHRNA7 is a candidate gene related to brain function in schizophrenia.

DISC1 was identified as a novel gene disrupted by translocation that segregated with schizophrenia in a Scottish family. It is also a key susceptibility factor for major mental illnesses [Derosse et al.,[2007]; Sawamura and Sawa,[2006]]. Several genetic studies have shown evidence of SNPs in DISC1 associated with schizophrenia, schizoaffective disorder, and bipolar disorder [Hodgkinson et al.,[2004]]. Among them, SNP rs821616 selected in our study was previously reported to be associated with schizophrenia (P = 0.004) in a family-based study [Callicott et al.,[2005]]. Specially, patients who were carriers of a common haplotype containing the minor allele at rs821616 had significantly lower ratings on paranoid delusions than noncarrier, presenting a significant association of DISC1 with lifetime severity of delusion in SZ [Derosse et al.,[2007]]. Therefore, it is reasonable that this SNP was selected as being relevant to brain function, particularly from a SZ and healthy control dataset.

CHAT participates in modulating wide-ranging cholinergic-dependent functions including cognitive performance, sleep, arousal, movement, and visual information processing. Compelling evidence has mounted implicating CHAT in schizophrenia [Holt et al.,[2005]; Karson et al.,[1996]]. A positive connection between three SNPs (rs1880676, rs3810950, and rs733722) locating in CHAT and SZ was reported in the study by Mancama et al. [[2007]] on Spain (Bosque) population. The three SNPs were not present in our SNP arrays, but we detected a different SNP, rs885834, located only 2,000 bases from the CHAT gene.

AADC is an enzyme implicated in two metabolic pathways, synthesizing important neurotransmitters, dopamine, norepinephrine, epinephrine, and serotonin. Evidence implicates AADC with schizophrenia [Ikemoto,[2002],[2004]; Ikemoto et al.,[2003]]. AADC may possibly act as a modulator of age at onset in males with schizophrenia [Borglum et al., [2001]]. Elevated AADC activity was observed in the brain of patients with psychosis [Reith et al.,[1994]]. The number of AADC-positive neurons was reduced in the striatum in schizophrenia compared to controls [Ikemoto et al.,[2003]]. On the other hand, there are contradictory reports of no association between AADC and SZ [Speight et al.,[2000]; Zhang et al., [2004]].

ADRA2A is a member of the G protein-coupled receptor superfamily. The receptor has a critical role in regulating neurotransmitter release from sympathetic nerves and from adrenergic neurons in the central nervous system. Possible associations of the ADRA2A with symptoms of attention-deficit/hyperactivity disorder have been studied and confirmed by many researchers [Deupree et al.,[2006]; Park et al.,[2005]; Schmitz et al.,[2006]; Wang et al.,[2006]]. However, no association between ADRA2A polymorphisms and schizophrenia has been found up to now [Clark et al.,[2007]].

Heterotrimeric guanine nucleotide-binding proteins (G proteins) integrate signals between receptors and effector proteins, important signal transducing molecules in cells. G proteins are functionally categorized into the inhibitory G proteins (Gi), the stimulatory G proteins (Gs), and other G proteins (Go). The Go is the most abundant G protein class expressed in brain, but its function is less known and maybe involved in mediating extracellular signal-regulated kinase activation by G protein-coupled receptors [Zhang et al.,[2003]]. The expression of GNAO1 was reported significantly decreased in individuals with schizophrenia compared to unaffected family controls [Vawter et al.,[2004]].

SCARB1 has affinity for high-density lipoproteins (HDLs) and mediates the selective uptake of cholesterol esters. Several studies have shown that SCARB1 protein is expressed in the human brain [Husemann and Silverstein,[2001]; Srivastava, [2003]], and contributes to selective uptake of HDL-associated vitamin E in the brain, to play an important role in the blood-brain barrier [Goti et al.,[2001]]. APOC3, a very low density lipoprotein protein, is also a constituent of HDL and triglyceride-rich lipoprotein, inhibiting lipoprotein lipase and hepatic lipase. CHRMS is a member of G protein-coupled receptor, whose function is defined by acetylcholine binding. Muscarinic receptors influence many effects of acetylcholine in the central and peripheral nervous systems. The CHRM3 controls smooth muscle contraction and its stimulation causes secretion of glandular tissue. The connection of these last three genes with mental disorders is not clear. However, the strong correlation of SCARB1 and DISC1 SNPs suggests that this gene may be associated with schizophrenia.

Schizophrenia-Relevant Components

We have discussed a pair of fMRI/SNP components that are functionally correlated. This suggests that the selected genes contribute directly or indirectly to or partially influence the related brain functions. In addition, we investigated a potential relationship between this pair of fMRI/SNP components and schizophrenia, considering possible factors such as gender,

IQ, and symptom severity. Because of incomplete medication status information, we did not include these in our analysis. All the factors except symptom severity (PANSS) scores can maximally explain 2% of total variance of loading parameters for both fMRI and genetic components. For SZ patients, the PANSS scores show a strong relationship with the fMRI/SNP components. Positive and negative symptom scores show different connection patterns for the fMRI and SNP components. Further analysis is needed to explore the potential relations of SZ severity with brain function and genetic associations, but generally speaking, the strong observed connections actually support a correlation between a subjective measurement of disease severity and an objective brain imaging determination.

From the *t*-test results ($P = 0.001$ for SNPs; $P = 0.0006$ for fMRI), we found a significant difference between two subject groups. After we explored the characteristics of fMRI and genetic components, the schizophrenia-relevant changes are consistent with known dysfunctions in this illness. For example, the abnormalities in parietal lobe (precuneus, superior parietal gyrus) and superior frontal cortex occur in regions commonly implicated as abnormal in schizophrenia. Furthermore, as discussed earlier, CHRNA7, DISC1, and CHAT are considered candidate genes for schizophrenia vulnerability and brain alterations. However, in order to confirm the connection between these genes and the function of specific brain regions, as well as their schizophrenia relevance, the same approach will need to be applied to a much larger group of subjects using more SNPs. We are now in the process of collecting data from additional subjects as well as analyzing these data with a whole genome SNP analysis.

CONCLUSION AND FUTURE WORK

In this paper, through parallel ICA, we built up a physiogenomic framework to combine fMRI data and genetic data to investigate connections between them. The results of this study suggest a valid relationship between specific regional brain functions and the selected genes. Brain regions included those precuneus, cuneus, and lingual gyrus, mainly involved in memory retrieval network. Some of these regions were previously implicated in schizophrenia and other psychiatric disorders. Genetic associations revealed the contributions of 10 SNPs (in 9 genes). Three of them, CHRNA7, DISC1, and CHAT have been previously reported to be closely associated with schizophrenia, and the other two have shown connections with brain dysfunction. Some have been linked to each other functionally, including SCARB1 and APOC3; yet others have not yet been studied. Moreover, this pair of SNP/FMRI component also showed a significant difference between the schizophrenia and control groups. Both components thus appear to be schizophrenia-relevant. However, these latter results must be considered preliminary and subject to replication in a larger SNP set with more subjects.

In summary, our study demonstrates a novel approach to analyzing multimodality, fMRI, and genetic data, in order to investigate connections between them. The new approach can assess multivariate genetic influence on endophenotypes, such as brain function related to mental disorders. As proof of principle, we tested 367 SNPs that were chosen initially for a metabolism study. The identification of CHRNA7 and DISC1 as associated genes validates our approach. Given that current technology can investigate over 500,000 SNPs, the analysis of such data will provide a much more comprehensive means to identify possible SNP/fMRI associations, and the proposed approach is well suited to perform such an analysis.

Acknowledgements

We thank the research staff at the Olin Neuropsychiatry Research Center who helped to collect and process the data. This research was supported by the National Institutes of Health, under grants 1 R01 EB 000840, 1 R01 EB 005846 (to VDC), and 2 RO1 MH43775 and 5 RO1 MH52886 (to GP) and a grant from the MIND Institute (NPB), as well as SBIR R43 grant MH075481 (to GR).

REFERENCES

Addis DR,McIntosh AR,Moscovitch M,Crawley AP,McAndrews MP (2004): Characterizing spatial and temporal features of autobiographical memory retrieval networks: A partial least squares approach. *Neuroimage* **23**: 1460-1471. Links

Akaike H (1974): A new look at statistical model identification. *IEEE Trans Automatic Control* **19**: 716-726. Links

Amari S (1998): Natural gradient works efficiently in learning. *Neural Comput* **10**: 251-276. Links

Barta PE,Pearlson GD,Powers RE,Richards SS,Tune LE (1990): Auditory hallucinations and smaller superior temporal gyral volume in schizophrenia. *Am J Psychiatry* **147**: 1457-1462. Links

Bath KG,Lee FS (2006): Variant BDNF (Val66Met) impact on brain structure and function. *Cogn Affect Behav Neurosci* **6**: 79-85. Links

Bell AJ,Sejnowski TJ (1995): An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* **7**: 1129-1159. Links

Blair JR,Spreen O. (1989): Predicting premorbid IQ: A revision of the National Adult Reading Test. *Clinical Neuropsychologist* **3**: 129-136. Links

Borglum AD,Hampson M,Kjeldsen TE,Muir W,Murray V,Ewald H,Mors O,Blackwood D,Kruse TA (2001): Dopa decarboxylase genotypes may influence age at onset of schizophrenia. *Mol Psychiatry* **6**: 712-717. Links

Brahmbhatt SB,Haut K,Csernansky JG,Barch DM (2006): Neural correlates of verbal and nonverbal working memory deficits in individuals with schizophrenia and their high-risk siblings. *Schizophr Res* **87**: 191-204. Links

Cabeza R,Mangels J,Nyberg L,Habib R,Houle S,McIntosh AR,Tulving E (1997): Brain regions differentially involved in remembering what and when: A PET study. *Neuron* **19**: 863-870. Links

Calhoun VD,Adali T (2006): Unmixing fMRI with independent component analysis. *IEEE Eng Med Biol Mag* **25**: 79-90. Links

Calhoun VD,Adali T,McGinty VB,Pekar JJ,Watson TD,Pearlson GD (2001a): fMRI activation in a visual-perception task: Network of areas detected using the general linear model and independent components analysis. *Neuroimage* **14**: 1080-1088. Links

Calhoun VD,Adali T,Pearlson GD,Pekar JJ (2001b): A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp* **14**: 140-151. Links

Calhoun VD,Adali T,Pearlson GD,Pekar JJ (2001c): Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum Brain Mapp* **13**: 43-53. Links

Calhoun VD,Adali T,Pekar JJ (2004): A method for comparing group fMRI data using independent component analysis: Application to visual, motor and visuomotor tasks. *Magn Reson Imaging* **22**: 1181-1191. Links

Calhoun VD,Adali T,Kiehl KA,Astur R,Pekar JJ,Pearlson GD (2006): A method for multitask fMRI data fusion applied to schizophrenia. *Hum Brain Mapp* **27**: 598-610. Links

Callicott JH,Straub RE,Pezawas L,Egan MF,Mattay VS,Hariri AR,Verchinski BA,Meyer-Lindenberg A,Balkissoon R,Kolachana B,Goldberg TE,Weinberger DR (2005): Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proc Natl Acad Sci USA* **102**: 8627-8632. Links

Cardoso JF (1997): Infomax and maximum likelihood for blind source separation. *IEEE Signal Process Lett* **4**: 112-114. Links

Carter CJ (2006): Schizophrenia susceptibility genes converge on interlinked pathways related to glutamatergic transmission and long-term potentiation, oxidative stress and oligodendrocyte viability. *Schizophr Res* **86**: 1-14. Links

Cavanna AE,Trimble MR (2006): The precuneus: A review of its functional anatomy and behavioural correlates. *Brain* **129**(Part 3): 564-583. Links

Clark DA,Mata I,Kerwin RW,Munro J,Arranz MJ (2007): No association between ADRA2A polymorphisms and schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* **144**: 341-343. Links

Dawy Z,Sarkis M,Hagenauer J,Mueller J (2005): A novel gene mapping algorithm based on independent component analysis. IEEE International Conference on Acoustics, Speech, and Signal Processings. 5: 381-384.

De Luca V,Wang H,Squassina A,Wong GW,Yeomans J,Kennedy JL (2004): Linkage of M5 muscarinic and α7-nicotinic receptor genes on 15q13 to schizophrenia. *Neuropsychobiology* **50**: 124-127. Links

Derosse P,Hodgkinson CA,Lencz T,Burdick KE,Kane JM,Goldman D,Malhotra AK (2007): Disrupted in Schizophrenia 1 genotype and positive symptoms in schizophrenia. *Biol Psychiatry* **61**: 1208-1210. Links

Detera-Wadleigh SD,McMahon FJ (2006): G72/G30 in schizophrenia and bipolar disorder: Review and meta-analysis. *Biol Psychiatry* **60**: 106-114. Links

Deupree JD,Smith SD,Kratochvil CJ,Bohac D,Ellis CR,Polaha J,Bylund DB (2006): Possible involvement of α-2A adrenergic receptors in attention deficit hyperactivity disorder: Radioligand binding and polymorphism studies. *Am J Med Genet B Neuropsychiatr Genet* **141**: 877-884. Links

Fan JB,Oliphant A,Shen R,Kermani BG,Garcia F,Gunderson KL,Hansen M,Steemers F,Butler SL,Deloukas P,Galver L,Hunt S,McBride C,Bibikova M,Rubano T,Chen J,Wickham E,Doucet D,Chang W,Campbell D,Zhang B,Kruglyak S,Bentley D,Haas J,Rigault P,Zhou L,Stuelpnagel J,Chee MS (2003): Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* **68**: 69-78. Links

Fan JB,Ma J,Li XW,Zhang CS,Sun WW,He G,Gu NF,Feng GY,St Clair D,He L (2006): Population-based and family-based association studies of an (AC)n dinucleotide repeat in α-7 nicotinic receptor subunit gene and schizophrenia. *Schizophr Res* **84**: 222-227. Links

First MB,Spitzer RL,Gibbon M,Williams JB (1995): *Structured Clinical Interview for DSM-IV Axis I Disorders, patient edition (SCID-I/P, Version 2.0)*. New York: Biometrics Research Department, New York State Psychiatric Institute.

Frederikse M,Lu A,Aylward E,Barta P,Sharma T,Pearlson G (2000): Sex differences in inferior parietal lobule volume in schizophrenia. *Am J Psychiatry* **157**: 422-427. Links

Freedman R,Coon H,Myles-Worsley M,Orr-Urtreger A,Olincy A,Davis A,Polymeropoulos M,Holik J,Hopkins J,Hoff M,Rosenthal J,Waldo MC,Reimherr F,Wender P,Yaw J,Young DA,Breese CR,Adams C,Patterson D,Adler LE,Kruglyak L,Leonard S,Byerley W (1997): Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. *Proc Natl Acad Sci USA* **94**: 587-592. Links

Freedman R,Leonard S,Gault JM,Hopkins J,Cloninger CR,Kaufmann CA,Tsuang MT,Farone SV,Malaspina D,Svrakic DM,Sanders A,Gejman P (2001): Linkage disequilibrium for schizophrenia at the chromosome 15q13-14 locus of the α7-nicotinic acetylcholine receptor subunit gene (CHRNA7). *Am J Med Genet* **105**: 20-22. Links

Freedman R,Leonard S,Waldo M,Gault J,Olincy A,Adler LE (2006): Characterization of allelic variants at chromosome

15q14 in schizophrenia. *Genes Brain Behav* **5**(Suppl 1): 14-22. Links

Freire L,Mangin JF (2001): Motion correction algorithms may create spurious brain activations in the absence of subject motion. *Neuroimage* **14**: 709-722. Links

Friston K,Ashburner J,Frith CD,Poline JP,Heather JD,Frackowiak RS (1995): Spatial registration and normalization of images. *Hum Brain Mapp* **2**: 165-189. Links

Gaser C,Volz HP,Kiebel S,Riehemann S,Sauer H (1999): Detecting structural changes in whole brain based on nonlinear deformations - Application to schizophrenia research. *Neuroimage* **10**: 107-113. Links

Gault J,Hopkins J,Berger R,Drebing C,Logel J,Walton C,Short M,Vianzon R,Olincy A,Ross RG,Adler LE,Freedman R,Leonard S (2003): Comparison of polymorphisms in the α7 nicotinic receptor gene and its partial duplication in schizophrenic and control subjects. *Am J Med Genet B Neuropsychiatr Genet* **123**: 39-49. Links

Goldberg II,Harel M,Malach R (2006a): When the brain loses its self: Prefrontal inactivation during sensorimotor processing. *Neuron* **50**: 329-339. Links

Goldberg TE,Straub RE,Callicott JH,Hariri A,Mattay VS,Bigelow L,Coppola R,Egan MF,Weinberger DR (2006b): The g72/g30 gene complex and cognitive abnormalities in schizophrenia. *Neuropsychopharmacology* **31**: 2022-2032. Links

Goti D,Hrzenjak A,Levak-Frank S,Frank S,van der Westhuyzen DR,Malle E,Sattler W (2001): Scavenger receptor class B, type I is expressed in porcine brain capillary endothelial cells and contributes to selective uptake of HDL-associated vitamin E. *J Neurochem* **76**: 498-508. Links

Harrison PJ,Owen MJ (2003): Genes for schizophrenia? Recent findings and their pathophysiological implications. *Lancet* **361**: 417-419. Links

Ho BC,Wassink TH,O'Leary DS,Sheffield VC,Andreasen NC (2005): Catechol-*O*-methyl transferase Val158Met gene polymorphism in schizophrenia: Working memory, frontal lobe MRI morphology and frontal cerebral blood flow. *Mol Psychiatry* **10**: 229, 287-298. Links

Ho BC,Milev P,O'Leary DS,Librant A,Andreasen NC,Wassink TH (2006): Cognitive and magnetic resonance imaging brain morphometric correlates of brain-derived neurotrophic factor Val66Met gene polymorphism in patients with schizophrenia and healthy volunteers. *Arch Gen Psychiatry* **63**: 731-740. Links

Hodgkinson CA,Goldman D,Jaeger J,Persaud S,Kane JM,Lipsky RH,Malhotra AK (2004): Disrupted in schizophrenia 1 (DISC1): Association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am J Hum Genet* **75**: 862-872. Links

Holt DJ,Bachus SE,Hyde TM,Wittie M,Herman MM,Vangel M,Saper CB,Kleinman JE (2005): Reduced density of cholinergic interneurons in the ventral striatum in schizophrenia: An in situ hybridization study. *Biol Psychiatry* **58**: 408-416. Links

Horne BD,Camp NJ (2004): Principal component analysis for selection of optimal SNP-sets that capture intragenic genetic variation. *Genet Epidemiol* **26**: 11-21. Links

Houy E,Raux G,Thibaut F,Belmont A,Demily C,Allio G,Haouzir S,Fouldrin G,Petit M,Frebourg T,Campion D (2004): The promoter -194 C polymorphism of the nicotinic α7 receptor gene has a protective effect against the P50 sensory gating deficit. *Mol Psychiatry* **9**: 320-322. Links

Husemann J,Silverstein SC (2001): Expression of scavenger receptor class B, type I, by astrocytes and vascular smooth muscle cells in normal adult mouse and human brain and in Alzheimer's disease brain. *Am J Pathol* **158**: 825-832. Links

Ikemoto K (2002): [Localization of non-monoaminergic aromatic **L**-amino acid decarboxylase neurons (D-neurons) in the human striatum and their functional significance]. *Kaibogaku Zasshi* **77**: 71-75. Links

Ikemoto K (2004): Significance of human striatal D-neurons: Implications in neuropsychiatric functions. *Prog Neuropsychopharmacol Biol Psychiatry* **28**: 429-434. Links

Ikemoto K,Nishimura A,Oda T,Nagatsu I,Nishi K (2003): Number of striatal D-neurons is reduced in autopsy brains of schizophrenics. *Leg Med (Tokyo)* **5**(Suppl 1): S221-S224. Links

Karson CN,Mrak RE,Husain MM,Griffin WS (1996): Decreased mesopontine choline acetyltransferase levels in schizophrenia. Correlations with cognitive functions. *Mol Chem Neuropathol* **29**: 181-191. Links

Kiehl KA,Liddle PF (2003): Reproducibility of the hemodynamic response to auditory oddball stimuli: A six-week test-retest study. *Hum Brain Mapp* **18**: 42-52. Links

Kiehl KA,Stevens MC,Laurens KR,Pearlson G,Calhoun VD,Liddle PF (2005): An adaptive reflexive processing model of neurocognitive function: Supporting evidence from a large scale (*n* = 100) fMRI study of an auditory oddball task. *Neuroimage* **25**: 899-915. Links

Kircher T,Liddle P,Brammer M,Murray R,McGuire P (2003): [Neural correlates of "negative" formal thought disorder]. *Nervenarzt* **74**: 748-754. Links

Lee C,McGlashan TH,Woods SW (2005): Prevention of schizophrenia: Can it be achieved? *CNS Drugs* **19**: 193-206. Links

Lee SI,Batzoglou S (2003): Application of independent component analysis to microarrays. *Genome Biol* **4**:R76.1-21. Links

Leonard S,Gault J,Hopkins J,Logel J,Vianzon R,Short M,Drebing C,Berger R,Venn D,Sirota P,Zerbe G,Olincy A,Ross RG,Adler LE,Freedman R (2002): Association of promoter variants in the α7 nicotinic acetylcholine receptor subunit gene with an inhibitory deficit found in schizophrenia. *Arch Gen Psychiatry* **59**: 1085-1096. Links

Li Y,Adali T,Calhoun VD. (2007b): Estimating the number of independent components for functional magnetic resonance imaging data. *Hum Brain Mapp* **28**: 1251-1266. Links

Liebermeister W (2002): Linear modes of gene expression determined by independent component analysis. *Bioinformatics* **18**: 51-60. Links

Lin Z,Altman RB (2004): Finding haplotype tagging SNPs by use of principal components analysis. *Am J Hum Genet* **75**: 850-861. Links

Mancama D,Mata I,Kerwin RW,Arranz MJ (2007): Choline acetyltransferase variants and their influence in schizophrenia and olanzapine response. *Am J Med Genet B Neuropsychiatr Genet* **144**: 849-853. Links

McKeown MJ,Sejnowski TJ (1998): Independent component analysis of fMRI data: Examining the assumptions. *Hum Brain Mapp* **6**: 368-372. Links

McKeown MJ,Makeig S,Brown GG,Jung TP,Kindermann SS,Bell AJ,Sejnowski TJ (1998): Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp* **6**: 160-188. Links

Meyer-Lindenberg A,Weinberger DR (2006): Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nat Rev Neurosci* **7**: 818-827. Links

Neckelmann G,Specht K,Lund A,Ersland L,Smievoll AI,Neckelmann D,Hugdahl K (2006): Mr morphometry analysis of grey matter volume reduction in schizophrenia: Association with hallucinations. *Int J Neurosci* **116**: 9-23. Links

Nocedal J,Wright SJ (1999): Line search methods. In: Glynn P ,Robinson SM , editors. *Springer series in operations research: Numerical optimization*. New York, NY: Springer-Verlag. p 32-63.

Numata S,Ueno S,Iga J,Yamauchi K,Hongwei S,Kinouchi S,Shibuya-Tayoshi S,Tayoshi S,Aki H,Sumitani S,Itakura M,Ohmori T (2007): Interaction between catechol-O-methyltransferase (COMT) Val108/158Met and brain-derived neurotrophic factor (BDNF) Val66Met polymorphisms in age at onset and clinical symptoms in schizophrenia. *J Neural Transm* **114**: 255-259. Links

Ohnishi T,Hashimoto R,Mori T,Nemoto K,Moriguchi Y,Iida H,Noguchi H,Nakabayashi T,Hori H,Ohmori M,Tsukue R,Anami K,Hirabayashi N,Harada S,Arima K,Saitoh O,Kunugi H (2006): The association between the Val158Met polymorphism of the catechol-*O*-methyl transferase gene and morphological abnormalities of the brain in chronic schizophrenia. *Brain* **129**(Part 2): 399-410. Links

Oliphant A,Barker DL,Stuelpnagel JR,Chee MS (2002): BeadArray technology: Enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* Suppl **56-58**: 60-61. Links

Owen MJ,Craddock N,O'Donovan MC (2005): Schizophrenia: Genes at last? *Trends Genet* **21**: 518-525. Links

Park L,Nigg JT,Waldman ID,Nummy KA,Huang-Pollock C,Rappley M,Friderici KH (2005): Association and linkage of α-2A adrenergic receptor gene polymorphisms with childhood ADHD. *Mol Psychiatry* **10**: 572-580. Links

Pearlson GD,Petty RG,Ross CA,Tien AY (1996): Schizophrenia: A disease of heteromodal association cortex? *Neuropsychopharmacology* **14**: 1-17. Links

Porteous DJ,Thomson P,Brandon NJ,Millar JK (2006): The genetics and biology of DISC1 - An emerging role in psychosis and cognition. *Biol Psychiatry* **60**: 123-131. Links

Reith J,Benkelfat C,Sherwin A,Yasuhara Y,Kuwabara H,Andermann F,Bachneff S,Cumming P,Diksic M,Dyve SE,Etienne P,Evans AC,Lal S,Shevell M,Savard G,Wong DF,Chouinard G,Gjedde A (1994): Elevated dopa decarboxylase activity in living brain of patients with psychosis. *Proc Natl Acad Sci USA* **91**: 11651-11654. Links

Roffman JL,Weiss AP,Goff DC,Rauch SL,Weinberger DR (2006): Neuroimaging-genetic paradigms: A new approach to investigate the pathophysiology and treatment of cognitive deficits in schizophrenia. *Harv Rev Psychiatry* **14**: 78-91. Links

Ruano G,Thompson PD,Windemuth A,Smith A,Kocherla M,Holford TR,Seip R,Wu AH. (2005a): Physiogenomic analysis links serum creatine kinase activities during statin therapy to vascular smooth muscle homeostasis. *Pharmacogenomics* **6**: 865-872. Links

Ruano G,Windemuth A,Holford TR (2005b): Physiogenomics: integrating systems engineering and nanotechnology for personalized medicine. In: Bronzino J , editor. *The Biomedical Engineering Handbook*.3rd ed. Boca Raton, FL: CRC Press. p 281-289.

Ruano G,Seip RL,Windemuth A,Zollner S,Tsongalis GJ,Ordovas J,Otvos J,Bilbie C,Miles M,Zoeller R,Visich P,Gordon P,Angelopoulos TJ,Pescatello L,Moyna N,Thompson PD (2006a): Apolipoprotein A1 genotype affects the change in high density lipoprotein cholesterol subfractions with exercise training. *Atherosclerosis* **185**: 65-69. Links

Ruano G,Windemuth A,Kocherla M,Holford T,Fernandez ML,Forsythe CE,Wood RJ,Kraemer WJ,Volek JS (2006b): Physiogenomic analysis of weight loss induced by dietary carbohydrate restriction. *Nutr Metab (Lond)* **3**: 20. Links

Ruano G,Zollner S,Goethe JW (2006c): Drug-induced metabolic syndrome (DIMS) in psychiatry: A diagnostic need uniquely suited to pharmacogenomics. In: Wong SHY ,Linder M ,Valdes RJ , editors. *Pharmacogenomics and Proteomics: Enabling the Practice of Personalized Medicine*. Washington, DC: AACC Press. pp 277-282.

Sawamura N,Sawa A (2006): Disrupted-in-schizophrenia-1 (DISC1): A key susceptibility factor for major mental illnesses. *Ann NY Acad Sci* **1086**: 126-133. Links

Schmitz M,Denardin D,Silva TL,Pianca T,Roman T,Hutz MH,Faraone SV,Rohde LA (2006): Association between α-2a-adrenergic receptor gene and ADHD inattentive type. *Biol Psychiatry* **60**: 1028-1033. Links

Speight G,Turic D,Austin J,Hoogendoorn B,Cardno AG,Jones L,Murphy KC,Sanders R,McCarthy G,Jones I,McCandless

F,McGuffin P,Craddock N,Owen MJ,Buckland P,O'Donovan MC (2000): Comparative sequencing and association studies of aromatic **L**-amino acid decarboxylase in schizophrenia and bipolar disorder. *Mol Psychiatry* **5**: 327-331. Links

Spitzer RL,Williams JB,Gibbon M (1996): *Structured Clinical Interview for DSM-IV: Non-Patient Edition (SCID-NP)*. New York: Biometrics Research Department, New York State Psychiatric Institute.

Srivastava RA (2003): Scavenger receptor class B type I expression in murine brain and regulation by estrogen and dietary cholesterol. *J Neurol Sci* **210**: 11-18. Links

Stassen HH,Bridler R,Hagele S,Hergersberg M,Mehmann B,Schinzel A,Weisbrod M,Scharfetter C (2000): Schizophrenia and smoking: Evidence for a common neurobiological basis? *Am J Med Genet* **96**: 173-177. Links

Vawter MP,Ferran E,Galke B,Cooper K,Bunney WE,Byerley W (2004): Microarray screening of lymphocyte gene expression differences in a multiplex schizophrenia pedigree. *Schizophr Res* **67**: 41-52. Links

Wang B,Wang Y,Zhou R,Li J,Qian Q,Yang L,Guan L,Faraone SV (2006): Possible association of the α-2A adrenergic receptor gene (ADRA2A) with symptoms of attention-deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* **141**: 130-134. Links

Whalley HC,Kestelman JN,Rimmington JE,Kelso A,Abukmeil SS,Best JJ,Johnstone EC,Lawrie SM (1999): Methodological issues in volumetric magnetic resonance imaging of the brain in the Edinburgh High Risk Project. *Psychiatry Res* **91**: 31-44. Links

Yang HH,Amari S (1997): Adaptive on-line learning algorithms for blind separation-maximization entropy and minimum mutual information. *Neural Comput* **9**: 1457-1482. Links

Zhang B,Jia Y,Yuan Y,Yu X,Xu Q,Shen Y (2004): No association between polymorphisms in the DDC gene and paranoid schizophrenia in a northern Chinese population. *Psychiatr Genet* **14**: 161-163. Links

Zhang WB,Zhang Z,Ni YX,Wu YL,Pei G (2003): A novel function of Goα: Mediation of extracellular signal-regulated kinase activation by opioid receptors in neural cells. *J Neurochem* **86**: 1213-1222. Links

Zhou SY,Suzuki M,Takahashi T,Hagino H,Kawasaki Y,Matsui M,Seto H,Kurachi M (2007): Parietal lobe volume deficits in schizophrenia spectrum disorders. *Schizophr Res* **89**: 35-48. Links