Schizophrenia Research xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

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

Integration of gene expression and GWAS results supports involvement of calcium signaling in Schizophrenia

L. Hertzberg [a,b], P. Katsel [c], P. Roussos [c], V. Haroutunian [c,d], E. Domany [a,*]

[a] Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel
[b] The Emotion-Cognition Research Center, Shalvata Mental Health Center, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[c] Department of Psychiatry, The Mount Sinai School of Medicine, New York, NY, USA
[d] Department of Psychiatry, James J Peters VA Medical Center, Bronx, NY, USA

ARTICLE INFO

Article history:
Received 11 June 2014
Received in revised form 1 February 2015
Accepted 2 February 2015
Available online xxxx

Keywords:
GWAS
Gene expression
Pathway analysis

ABSTRACT

The number of Genome Wide Association Studies (GWAS) of schizophrenia is rapidly growing. However, the small effect of individual genes limits the number of reliably implicated genes, which are too few and too diverse to perform reliable pathway analysis; hence the biological roles of the genes implicated in schizophrenia are unclear. To overcome these limitations we combine GWAS with genome-wide expression data from human post-mortem brain samples of schizophrenia patients and controls, taking these steps: 1) Identify 36 GWAS-based genes which are expressed in our dataset. 2) Find a cluster of 19 genes with highly correlated expression. We show that this correlation pattern is robust and statistically significant. 3) GO-enrichment analysis of these 19 genes reveals significant enrichment of ion channels and calcium-related processes. This finding (based on analyzing a small number of coherently expressed genes) is validated and enhanced in two ways: First, the emergence of calcium channels and calcium signaling is corroborated by identifying proteins that interact with those encoded by the cluster of 19. Second, extend the 19 cluster genes into 1028 genes, whose expression is highly correlated with the cluster's average profile. When GO-enrichment analysis is performed on this extended set, many schizophrenia related pathways appear, with calcium-related processes enriched with high statistical significance. Our results give further, expression-based validation to GWAS results, support a central role of calcium-signaling in the pathogenesis of schizophrenia, and point to additional pathways potentially related to the disease.

1. Introduction

Studies conducted by the International Schizophrenia Consortium attribute one-third of genetic susceptibility for schizophrenia (estimated to be 80% (Sullivan et al., 2003)) to the collective effect of hundreds of common polygenic variants, each contributing a small effect (Purcell et al., 2009; Gejman et al., 2011).

The results of GWAS (Johnson and O'Donnell, 2009) contain many false positives (Burmeister et al., 2008). Current replicable GWAS results account for only a small percentage of the estimated heritability (Ozomaro et al., 2013) and their systematic biological interpretation is lacking. GWAS results have generated several biological hypotheses, such as involvement of *ZNF804A* through regulation of gene expression (O'Donovan et al., 2008); of infection, through interaction with genes located in the major histocompatibility complex (MHC) region, and involvement of calcium channels, based on two GWAS-derived genes that encode for calcium channel subunits (Ripke et al., 2013). Focusing on a specific GWAS-based variant might be, however, misleading, as each specific variant may be a false positive, and in any case it confers only a small increase of the risk.

Higher-level interpretation of GWAS results in terms of implicated pathways is hindered by the number of reliable GWAS-derived genes (see (Ripke et al., 2013)), too small to yield robust and statistically meaningful pathway enrichment analysis. Ripke et al. (2013) dealt with this limitation by focusing from the outset on the set of SNPs located in genes encoding calcium channel subunits, and found enrichment of SNPS with small *p*-values in this set. Jia et al. (2010) performed a broader search and identified enrichment of pathways related to metabolism of glutamate, apoptosis, inflammation and immune system. In O'Dushlaine et al. (2011), the pathway of cell adhesion passed multiple testing correction. In Schizophrenia Working Group of the Psychiatric Genomics (2014), a multi-stage GWAS of up to 36,989 cases and 113,075 controls, 108 loci met genome-wide significance, 83 of which have not been previously reported. Protein-coding variants played a limited role, consistently with the hypothesis that most associated variants detected by GWAS exert their effects through altering

* Corresponding author at: Department of Physics of Complex Systems, Weizmann Institute of Science, 234 Herzl St. Rehovot 7610001, Israel. Tel.: +972 8 934 3964; fax: +972 8 934 4109.
E-mail address: Eytan.domany@weizmann.ac.il (E. Domany).

http://dx.doi.org/10.1016/j.schres.2015.02.001
0920-9964/

gene expression rather than protein structure. Genes encoding voltage-gated calcium channel subunits, *CACNA1C*, *CACNB2* and *CACNA1I*, were found to be significantly associated, extending previous findings implicating members of this family of proteins in schizophrenia.

Expression profiling studies in schizophrenia suggested changes in functional gene groups such as oligodendrocyte and myelin related genes, metabolism, synaptic transmission, GABAergic and glutamatergic pathways (Roussos et al., 2012; Katsel et al., 2005a; Mirnics et al., 2006). Often there is little agreement between different microarray studies regarding which transcripts are differentially expressed in the disease, reflecting the GWAS-based picture of a multifactorial heterogeneous disease, caused by a combination of many genetic and environmental factors. The reported results show only modest changes in genes' expression levels between cases and controls, calling for a more involved and integrated analysis. A model of a complex combination of genetic and environmental factors is concordant with the clinical heterogeneity of the disease (for example, gender difference in the age of onset and outcome) and with delicate structural brain changes that are found in both schizophrenia patients and their relatives (Cooper et al., 2014).

Here we focused on functional changes by analyzing a large-scale genome-wide gene expression dataset of postmortem brain samples, for the identification of biological pathways and processes through which GWAS-derived genes affect schizophrenia. Integration of data from different platforms potentially increases the biological and statistical reliability of the results, reducing significantly the false positives rate resulting from each of the separate analyses. A short list of genes identified by GWAS as relevant to the disease were used as "seed", and the expression data of these and related genes were analyzed. Correlations between their expression profiles were calculated; high pairwise correlations suggest shared biological pathways. A cluster of genes with high pairwise correlations of expression was identified, and gene ontology (GO) enrichment analysis was performed on the members of the cluster. Since the genes of the cluster have correlated expression, the chance that they belong to the same pathway increases. Indeed we identified significantly enriched pathways—ion channels and, specifically, calcium channel activity.

Since the enrichment analysis was based on a small number of genes, we sought ways to validate this result further. First, we searched for proteins that interact with members of our cluster. The interaction partners were mostly related to calcium signaling; calcium channel subunits and members of the calmodulin pathway emerged. For further validation we extended the list of genes to be searched for enrichment, assembling all genes with high correlation of expression with the average expression profile of the cluster. Enrichment analysis of this extended list identified many pathways that are known to be associated with schizophrenia, and calcium-related pathways re-emerged with high statistical significance.

2. Materials and methods

2.1. Subjects

Human brain samples were obtained from the Brain Bank of the Department of Psychiatry of the Mount Sinai Medical Center (New York, NY)/JJ Peters Veterans Administration Medical Center (Bronx, NY). Dissections were performed blind to diagnosis. All cortical dissections and sample preparation were performed as described previously (Hakak et al., 2001; Katsel et al., 2005a; Katsel et al., 2005b). Brain banking activities were approved by the Institutional Review Board of the Mount Sinai School of Medicine, and written consent for brain donation was obtained from the next-of-kin of all subjects.

Table 1 shows the main characteristics of the population analyzed. All subjects were clinically and neuropathologically assessed as described previously (Purohit et al., 1998).

Table 1
Patients' characteristics. Control samples were derived from subjects with no evidence of dementia, neuro or psychopathology. The cases with schizophrenia were diagnosed by clinical investigators to meet criteria for schizophrenia by DSM-III/IV and to evidence no other comorbid psychopathology.

Characteristics	Schizophrenia	Control
Number of patients	28	22
Sex (M/F)	20/8	8/14
Age (years)	74.2 (11.4)	82.8 (11.7)
Neuritic plaque density (#/mm^2)	1.1 (0.4)	1.3 (0.6)
Cognitive dementia Rating	2.4 (1.7)	0.5 (0.9)
Braak staging	0.3 (0.5)	2.0 (1.2)
Brain pH	6.4 (0.2)	6.6 (0.3)
PM1 (min)	683 (493)	419 (329)

2.2. Gene expression pre-processing

The computational analysis starts with the raw data of Affymetrics HG-U133A arrays, listed in Table 2 by their distribution across 17 brain regions, in cases and controls. Standard MAS-5 algorithm was used for normalization. Then expression levels below 20 were set 20 and log2-transformation was applied. Probe-sets without assigned Affymetrix gene symbols annotation were removed. 12,033 probe-sets were left for the rest of the analysis after filtering (out of 22,283), representing 8542 gene symbols. Probe sets of the same gene were combined. For full details see supplementary Methods.

2.3. Analysis of the expression patterns of the 36 GWAS-based genes

We used the SPIN tool (Tsafrir et al., 2005) to order genes (samples), such that genes (samples) with similar expression pattern are grouped together. SPIN applies an iterative search algorithm, and sorts items into neighborhoods of similar patterns. Then, for each pair of the 36 genes, Pearson correlation of expression, along all 480 samples, was calculated.

2.4. Estimating the statistical significance of the observed correlation pattern

We randomly selected a group of 36 genes and calculated their pairwise correlation values. Repeating this 1000 times we generated

Table 2
Distribution of Affymetrix HG-133A arrays across brain tissues and populations.

Anatomical gr.		Brodmann area	Schizophrenia	Control
Frontal lobe	Anterior prefrontal cortex	BA10	18	14
	Frontal eye fields	BA8	14	12
	Dorsolateral prefrontal cortex	BA46	20	18
	Primary motor cortex	BA4	15	6
	Pars opercularis, part of Broca's area	BA44	15	11
Cingulate gyrus	Anterior cingulate gyrus	BA32	15	17
	Posterior cingulate gyrus	BA23	11	13
Temporal lobe	Hippocampus	HIPP	12	12
	Temporopolar area	BA38	14	14
	Parahippocampal cortex	BA36	14	17
	Inferior temporal gyrus	BA20	17	14
	Middle temporal gyrus	BA21	17	16
	Superior temporal gyrus	BA22	19	14
Parietal lobe	Somatosensory association cortex	BA7	15	13
Occipital lobe	Primary visual cortex (V1)	BA17	13	13
Basal ganglia	Caudate nucleus	CD	15	11
	Putamen	PT	11	10
		Total	255	225

Please cite this article as: Hertzberg, L., et al., Integration of gene expression and GWAS results supports involvement of calcium signaling in Schizophrenia, Schizophr. Res. (2015), http://dx.doi.org/10.1016/j.schres.2015.02.001