Schizophrenia Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

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

Transcriptome-wide mega-analyses reveal joint dysregulation of immunologic genes and transcription regulators in brain and blood in schizophrenia

Jonathan L. Hess [a,b,c,1], Daniel S. Tylee [a,b,c,1], Rahul Barve [a,b,c], Simone de Jong [d,p], Roel A. Ophoff [d,e], Nishantha Kumarasinghe [f,g,h,i], Paul Tooney [h,j,k,l], Ulrich Schall [f,h,k,l], Erin Gardiner [h,j,l], Natalie Jane Beveridge [h,j,l], Rodney J. Scott [j,k], Surangi Yasawardene [g], Antionette Perera [g], Jayan Mendis [g], Vaughan Carr [h,m], Brian Kelly [f,k,l], Murray Cairns [h,j,k,l], the Neurobehavioural Genetics Unit: , Ming T. Tsuang [n,o], Stephen J. Glatt [a,b,c,*]

[a] Psychiatric Genetic Epidemiology & Neurobiology Laboratory (PsychGENe Lab), Syracuse, NY, USA
[b] Departments of Psychiatry and Behavioral Sciences & Neuroscience and Physiology, Syracuse, NY, USA
[c] SUNY Upstate Medical University, Syracuse, NY, USA
[d] Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Behavior, David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, California, USA
[e] Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
[f] School of Medicine & Public Health, The University of Newcastle, Callaghan, Newcastle, Australia
[g] Department of Anatomy, Faculty of Medical Sciences, University of Sri Jayawardenepura, Nugegoda, Sri Lanka
[h] Schizophrenia Research Institute, Sydney, New South Wales, Australia
[i] Faculty of Medicine, Sir John Kotelawala Defence University, Ratmalana, Sri Lanka
[j] School of Biomedical Sciences & Pharmacy, Faculty of Health, The University of Newcastle, New South Wales, Australia
[k] Hunter Medical Research Institute, Newcastle, Australia
[l] Centre for Translational Neuroscience & Mental Health, University of Newcastle, Callaghan, Newcastle, Australia
[m] School of Psychiatry, University of New South Wales, Kensington, New South Wales, Australia
[n] Center for Behavioral Genomics, Department of Psychiatry, Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA, USA
[o] Harvard Institute of Psychiatric Epidemiology and Genetics, Boston, USA
[p] MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK

ARTICLE INFO

Article history:
Received 14 March 2016
Received in revised form 7 July 2016
Accepted 11 July 2016
Available online xxxx

Keywords:
Schizophrenia
Gene expression
Transcriptome
Brain
Blood
Innate immunity
Support vector machine
Random forests

ABSTRACT

The application of microarray technology in schizophrenia research was heralded as paradigm-shifting, as it allowed for high-throughput assessment of cell and tissue function. This technology was widely adopted, initially in studies of *postmortem* brain tissue, and later in studies of peripheral blood. The collective body of schizophrenia microarray literature contains apparent inconsistencies between studies, with failures to replicate top hits, in part due to small sample sizes, cohort-specific effects, differences in array types, and other confounders. In an attempt to summarize existing studies of schizophrenia cases and non-related comparison subjects, we performed two mega-analyses of a combined set of microarray data from *postmortem* prefrontal cortices ($n = 315$) and from *ex-vivo* blood tissues ($n = 578$). We adjusted regression models per gene to remove non-significant covariates, providing best-estimates of transcripts dysregulated in schizophrenia. We also examined dysregulation of functionally related gene sets and gene co-expression modules, and assessed enrichment of cell types and genetic risk factors. The identities of the most significantly dysregulated genes were largely distinct for each tissue, but the findings indicated common emergent biological functions (*e.g.* immunity) and regulatory factors (*e.g.*, predicted targets of transcription factors and miRNA species across tissues). Our network-based analyses converged upon similar patterns of heightened innate immune gene expression in both brain and blood in schizophrenia. We also constructed generalizable machine-learning classifiers using the blood-based microarray data. Our study provides an informative atlas for future pathophysiologic and biomarker studies of schizophrenia.

1. Introduction

The molecular bases of schizophrenia (**SZ**) remain unresolved despite decades of intensifying research. This situation impedes progress toward biologically based risk assessment and diagnostic testing, early

* Corresponding author at: SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY 13210, United States.
E-mail address: stephen.glatt@psychgenelab.com (S.J. Glatt).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.schres.2016.07.006
0920-9964/

detection, and the development of rationally selected therapeutics to alter disease progression and clinical trajectory. As such, characterizing molecular correlates of SZ is of great potential interest and value. In the past 15 years, the transcriptome has received growing attention in SZ research, particularly in the effort to identify biomarkers—objective biological indicators of normal functioning or illness. Whole-transcriptome quantification (*e.g.*, by microarray) offers several attractive features: (1) it provides a relatively efficient and unbiased means of screening many analytes of a single molecule type (*i.e.*, messenger RNAs, **mRNAs**); (2) RNAs can be mapped reliably onto genes and proteins to assess a wide range of biological processes; (3) the measurement of large numbers of biological features allows for the assessment of network function; and (4) differences in mRNA expression reflect the combination of genetic and environmental factors, making it a more dynamic and responsive readout of biological function than static genetic variants. Indeed, transcriptome-wide studies of *postmortem* brain tissues have revealed altered molecular pathways and helped generate new hypotheses about the biological underpinnings of SZ. Similarly, studies of blood samples from individuals with SZ shed light on disturbances in circulating immune cells and could provide a basis for easily assessable SZ biomarkers.

Despite the vast potential and initial enthusiasm surrounding the use of microarrays in SZ research, the cross-study replication of genes and pathways found to be disrupted in SZ is mixed (Mirnics et al., 2006). A variety of practical and technical limitations may contribute to this, including: the evolution of new array technologies over time, the likelihood of etiologic heterogeneity of SZ (Arnedo et al., 2015; Tsuang and Faraone, 1995), the use of small sample sizes, and the inability to adequately protect against type-I errors, all of which exacerbate the "winner's-curse" phenomenon that undermines replication. In light of these issues, several studies have sought to consolidate the knowledge of transcriptomic abnormalities in SZ *via* meta-analysis (Bergon et al., 2015; Mistry and Pavlidis, 2010; Mistry et al., 2013a; Pérez-Santiago et al., 2012). These studies bolstered confidence by employing consistent pre-processing methods and demonstrating some similar dysregulated genes and network features across different studies; the implicated biological functions included oxidative phosphorylation, protein and nucleotide metabolism, synaptic transmission, myelination and glial function, and immune function, each of which have been implicated in previous work (Åberg et al., 2006; Dean, 2011; Devor and Waziri, 1993; Fineberg and Ellman, 2013; Middleton et al., 2002; Potvin et al., 2008). However, the approaches employed in previous meta-analyses studies had some limitations: (1) for the detection of differentially expressed genes, meta-analysis of summary statistics is relatively underpowered compared with combined-samples reanalysis of individual level data; (2) summary statistic meta-analysis does not allow flexible and transcript-specific modeling of clinical covariates across the entire sample; and (3) meta-analysis is not amenable to co-expression network analyses.

We use the term mega-analysis to refer to a strategy of the pooling of individual-level clinical and biological data from multiple studies for statistical modeling with appropriate correction for between-study variations (Mistry et al., 2013b; Seifuddin et al., 2013). This strategy allows for explicit modeling of factors that are consistently reported across studies (*i.e.*, gender, age), as well as factors that are inconsistently reported across studies (*e.g.*, subject medication status). In this study, we conducted two separate mega-analyses to summarize existing microarray-based transcriptomic studies of SZ in *postmortem* brain and in blood tissue using mixed-effect linear modeling. We extend upon previous approaches by employing network and annotation-based analyses to assess emergent biological functions. Furthermore, we perform systematic cross-tissue comparison of dysregulated functional gene sets and co-expression networks in SZ. We also characterized gene co-expression networks that were preserved across brain and blood samples in unaffected comparison subjects. Finally, we examined the cross-study generalizability of blood-based transcriptomic classifiers that differentiated SZ cases from unaffected comparison subjects.

2. Methods

Methods are described briefly here due to space constraints. A comprehensive description of all methods can be found in the Supplementary Materials.

2.1. Literature search and study selection

We searched public databases (*i.e.*, NCBI dbGaP, PubMed, SCOPUS, and EMBL-EBI ArrayExpress) for microarray-based studies of gene expression in subjects with SZ, schizoaffective disorder, or psychosis. We conducted this literature search for eligible data up to January 1, 2015; otherwise-eligible studies published after this date were not included in our analyses, but are shown in Supplementary Table 1 and compared to our findings qualitatively in the Discussion section. Twenty-five studies of blood-based gene expression and 19 studies of brain-based gene expression were identified. The following criteria for study inclusion and sample inclusion were used: (1) we only included studies that compared cases with unaffected non-related controls, (2) we only included cases classified as SZ or schizoaffective disorder, depressive subtype, based on the original investigators' determinations, (3) we only included studies for which raw probe-level data and gene annotations were available, (4) we only included studies that utilized non-custom microarray platforms developed by Affymetrix or Illumina to minimize technical sources of heterogeneity, and (5) we only included *postmortem* brain studies with samples consisting of tissue homogenates. Ultimately, we included nine blood studies and nine brain studies (*postmortem* prefrontal cortex, **PFC**, only) were ultimately retained for analysis (Table 1). The rationale for excluding each of the 26 other studies is provided in Supplementary Table 1.

2.2. Data import, normalization, quality control and probe matching

Data from each study were processed and normalized independently. Affymetrix arrays underwent robust multi-array average (**RMA**) normalization (Irizarry et al., 2003), with additional GC-correction whenever possible (*e.g.,* not compatible with Affymetrix Human Exon 1.0ST array). Both Affymetrix and Illumina array data were quantile-normalized and log-2 transformed. We mapped probes to HGNC gene symbols and collapsed expression values of multiple probes to individual genes through median summarization. Finally, for each gene within each individual study, expression values were *z*-transformed in order to normalize the range and variance of expression *across* datasets generated on different array platforms; the effects of normalization and transformation are depicted in **Fig. S1**. In order to identify potential differences in the proportions of leukocyte subtypes between SZ cases and unaffected comparison subjects, we performed deconvolution analysis using previously described methods (Abbas et al., 2009) followed by an independent samples *t*-test with family-wise Benjamini-Hochberg (**BH**) correction for multiple testing.

2.3. Mixed-effect linear modeling and gene set analysis

Expression and covariate data from individual studies were combined, creating separate brain ($n = 315$) and blood datasets ($n = 578$). Independent mega-analyses were performed on these datasets using mixed-effect linear modeling. The brain analysis included covariates for age (continuous), ancestry (Caucasian, Asian, African-American), gender (male, female), *postmortem* interval (continuous) and tissue pH (continuous). The blood analysis included covariates for age, sample-type (whole blood, leukocytes, peripheral blood mononuclear cells), ancestry (Caucasian, Asian), gender, and anti-psychotic status (yes, no; as defined by original study authors). A total of 20,767 genes

Please cite this article as: Hess, J.L., et al., Transcriptome-wide mega-analyses reveal joint dysregulation of immunologic genes and transcription regulators in brain and blood i..., Schizophr. Res. (2016), http://dx.doi.org/10.1016/j.schres.2016.07.006