Analyzing Schizophrenia by DNA Microarrays

Szatmár Horváth, Zoltán Janka, and Károly Mirnics

To understand the pathological processes of schizophrenia, we must embrace the analysis of the diseased human brain: we will never be able to recapitulate the pathology of uniquely human disorders in an animal model. Based on the outcome of the transcriptome profiling experiments performed to date, it appears that schizophrenia is associated with a global gene expression disturbance across many cortical regions. In addition, transcriptome changes are present in multiple cell types, including specific subclasses of principal neurons, interneurons, and oligodendrocytes. Furthermore, transcripts related to synaptic transmission, energy metabolism, and inhibitory neurotransmission are routinely found underexpressed in the postmortem brain tissue of subjects with schizophrenia. To put these transcriptome data in biological context, we must make our data publicly available and report our findings in a proper, expanded Minimum Information About a Microarray Experiment format. Cell-type specific expression profiling and sequencing-based transcript assessments should be expanded, with particular attention to understanding splice-variant changes in various mental disorders. Deciphering the pathophysiology of mental disorders depends on integrating data from across many research fields and techniques. Leads from postmortem transcriptome profiling will be essential to generate model animals, perform tissue culture experiments, and develop or evaluate novel drugs to treat this devastating disorder.

Key Words: DNA microarrays, gene expression, postmortem, RNA-seq, schizophrenia, transcriptome

Analysis of Postmortem Brain Tissue Is Challenging, yet Necessary

Limitations and cohort-specific datasets are an unwanted and unavoidable part of postmortem expression profiling studies (1-4), and for the successful interpretation of our experimental outcomes, we must carefully examine the challenges we are facing in this line of research. Postmortem brain availability is limited; building a brain bank is time consuming, expensive, and lasts decades; and most of the human brain banks around the country can collect only a few schizophrenic brains each year (5). As a result, postmortem experiments are performed on relatively few brains, limiting the power of our studies. Furthermore, schizophrenia is a spectrum disorder that encompasses many clinical presentations. Patients have diverse genetic backgrounds, different lifestyles, personality traits, and various habits (6,7). In addition, many of the patients have multiple comorbid health conditions, abuse alcohol and drugs, or have nicotine dependence. The postmortem brains are collected at various ages; the patients lived with a disease for various lengths of time, received different medications over the course of their lives, were exposed to various environmental influences, and died of different causes (1-4). All of these factors affect the brain transcriptome, making the interpretation of findings extremely challenging. Yet, despite these challenges, embracing the analysis of the diseased human brain is critical for understanding the pathological processes of schizophrenia: while the animal models can mimic certain aspects of the human pathology, we will never be able to recapitulate uniquely human disorders in an animal model.

From the Department of Psychiatry (SH, KM) and Vanderbilt Kennedy Center for Research on Human Development (KM), Vanderbilt University, Nashville, Tennessee; and Department of Psychiatry (ZJ), University of Szeged, Szeged, Hungary.

Address correspondence to Karoly Mirnics, M.D., Department of Psychiatry, Vanderbilt University, 8130A MRB III, 465 21st Avenue South, Nashville, TN 37232; E-mail: karoly.mirnics@vanderbilt.edu.

Received May 20, 2010; revised Jul 12, 2010; accepted Jul 14, 2010.

Transcriptome Changes Implicate Dysfunction Across Different Brain Regions, Cell Types, and Cellular Processes

With advancement of technology, we are making steady progress in understanding the pathophysiological processes associated with schizophrenia, and analysis of postmortem tissue has been a critical part in advancing our knowledge of this devastating disease. It has been a decade since the first DNA microarray experiments were performed on postmortem brain tissue (8-11). To date, dozens of DNA expression profiling experiments have been conducted on the brains of subjects with schizophrenia, providing us with an intriguing glance into the disturbed transcriptome networks (8,9,12-22).

Based on the outcome of the transcriptome profiling experiments performed to date, we can conclude that gene expression changes in schizophrenia affect multiple brain regions, including the prefrontal (8,9,11,21) and temporal cortices (17,19,23) and hippocampus (23,24). First, it appears that schizophrenia is associated with a global gene expression disturbance across many cortical regions, although some of the disturbances might be more pronounced in the brain regions that are primarily associated with the positive, negative, and cognitive symptoms of the disease (23). For example, underexpression of regulator of G-protein signaling 4 (RGS4) in subjects with schizophrenia is present in the prefrontal, visual, and motor cortices (10), while systemic gamma-aminobutyric acid system transcript reductions can be observed in prefrontal, anterior cingulate, primary motor, primary visual, and temporal cortices (18,25). In the context of the neurodevelopmental origin of schizophrenia (26), these common, widespread gene expression disturbances across the entire brain are not surprising and suggest that the region-related symptoms might arise as a combined result of gene expression changes and specific regional connectivity or work demand of the individual brain structures.

Second, transcriptome changes are present in multiple cell types, including specific subclasses of principal neurons (20,27,28), interneurons (18,25,29), and oligodendrocytes (8,23). This wealth of data makes the interpretation of the combined findings extremely challenging: the pathophysiological processes affecting the various cell populations are clearly interrelated, but a reliable time line or causality between them cannot be easily established. Furthermore, while there is a consensus that both the neuronal and glial transcriptomes are affected by the pathophysiology of schizophrenia, there is considerably less agreement about which individual transcripts are the most affected ones-they often vary from study to

0006-3223/$36.00
doi:10.1016/j.biopsych.2010.07.017

study, perhaps reflecting various differences across the studied postmortem cohorts.

Third, schizophrenia affects the expression of transcripts related to genes involved in multiple intracellular processes. Transcripts related to synaptic transmission (9,11,30), energy metabolism (14,15,24), immune response (12,31,32) and inhibitory neurotransmission (13,18,25) are routinely found altered in the postmortem brain tissue of subjects with schizophrenia. The relationship between these expression changes is still unclear, but one can hypothesize that the energy requirement of the brain cells is tightly associated with the number of connections that the neurons support. Thus, reduction in metabolic activity can result in elimination of synapses (33) or reduction in synapse number can lead to compensatory decrease in metabolic activity (34,35). However, it is also possible that the postmortem brain collection procedures can introduce cohort biases (1-3,36), leading to enrichment of diseased subjects with a similar disease subphenotype across the different brain collections. For example, a collection of brains obtained from chronically hospitalized individuals (who have a severe phenotype and do not respond well to medications) (8,23) is likely to show transcriptome changes that are quite different from a collection that is obtained through the assistance of the coroner's office (where the brains are obtained from patients who died an accidental death, were living and interacting in the outside community, and presumably responded well to medication) (9,10,18).

Fourth, empiric evidence suggests that there may be a strong link between gene expression changes and genetic susceptibility within human brain disorders (1,3,37). Regulator of G-protein signaling 4 (10,38-41), disrupted in schizophrenia 1 (42,43), glutamic acid decarboxylase 67 (25,44), dysbindin (45,46), metabotropic glutamate receptor subtype 3 (47), neuregulin 1 (48-50), gamma-aminobutyric acid type A receptor beta 2 (51,52), and 14-3-3 isoforms (30,53-56) have been identified as both schizophrenia susceptibility genes and genes with altered transcription in the diseased brain. Yet, the gene expression changes cannot be explained by predisposing genetic variants: the significant gene expression changes are present in the majority of subjects with schizophrenia, while the genetic predisposition of any single gene or copy number variant can explain only the minor proportion of the diagnosis. For example, RGS4 underexpression is present in approximately 70% to 90% of the postmortem brain of diseased subjects (10); yet, having a disease-predisposing single nucleotide polymorphism (SNP) in the RGS4 gene only slightly (but significantly) increases the odds of developing schizophrenia in multiple studies (38-41) with various degrees of consistency across cohorts (57,58). Similarly, in addition to the genetic signal in the minority of the diseased individuals, disrupted in schizophrenia 1 binding partner expression is also altered in the postmortem brains of a much larger proportion of subjects with schizophrenia (42). The possible explanation for this magnitude discrepancy is that many of the genes, critically important for the pathophysiology of schizophrenia, are convergence points (molecular hubs) of the transcriptome networks (59). Thus, underexpression of a hub gene might arise by two independent mechanisms: a genetic susceptibility within its own regulatory sequence or by an independent, upstream genetic or adaptational event (1).

Microarray Data Should Be Shared in a Proper Format

Sharing of transcriptome datasets is feasible and can be submitted to several major data repositories. The two leading repositories are the Gene Expression Omnibus (GEO) (http://www.ncbi.nih.gov/geo/) (60), maintained by National Center for Biotechnology Information, and ArrayExpress Archive (http://www.ebi.ac.uk/microarray-as/ae/) (61), maintained by the European Bioinformatics Institute. The GEO is a public repository that archives and freely distributes high throughput gene expression data submitted by the scientific community. The GEO volumes currently contain over a billion gene expression measurements obtained from over 100 organisms, with tools that efficiently explore, query, and visualize the datasets using user-friendly, Web-based tools (62). Similarly, ArrayExpress represents a free repository and a discovery resource, with tools that allow comprehensive mining and comparison of various datasets (63). Their new tool, the Public Interface for Human Gene Expression Map (http://wwwdev.ebi.ac.uk/microarray/hge/HGE.jsp), is built on the recently published global map of gene expression, which is developed on data collected from 163 laboratories worldwide involving 5372 human samples from various tissues, cell types, and diseases (64). Unfortunately, both of these resources contain very few expression datasets derived from expression profiling of psychiatric disorders, resulting in sparing utilization of the resources by disease-oriented brain researches. Importantly, neither of these databases makes judgment about the quality of the deposited datasets, and one has to carefully review the experimental descriptions associated with the deposited files.

To ensure that the users can evaluate the quality and nature of the deposited transcriptome data, the Microarray and Gene Expression Data Society (MGED) (http://www.mged.org/) was established in 1999 as a grass roots movement by major DNA microarray users and developers. It is made up by biologists, engineers, computer scientists, and data analysts with a goal to facilitate biological and biomedical discovery through data integration. In an effort to standardize microarray data reporting, MGED proposed a set of guidelines, Minimum Information About a Microarray Experiment (MIAME) (http://www.mged.org/Workgroups/MIAME/miame.html), which is needed to enable the interpretation of the results of the experiment unambiguously and potentially to reproduce the experiment (65,66). There are six critical parameters to this reporting: 1) raw data for each hybridization; 2) final, normalized data for the study from which the conclusions are drawn; 3) sample annotation including experimental manipulations; 4) experimental design; 5) annotation of the DNA microarray; and 6) laboratory and data processing protocols. The MGED recommends the use of a spreadsheet-based, MicroArray Gene Expression Tabular (MAGE-TAB) format (67), as well as use of MGED Ontology for the description of the key experimental concepts (68). Most of the journals now require public disclosure of the data in this format at the time of publication. Sharing of microarray data was also required from all researchers using the National Institutes of Health established microarray core facilities (http://arrayconsortium.tgen.org). Unfortunately, after June 1, 2010, the National Institutes of Health Neuroscience Microarray Consortium will cease all operations.

While the MGED-proposed MIAME guidelines are a huge step toward meaningful comparison, integration, and interpretation of datasets, they represent a one-size-fits-all model, which might not be sufficient for postmortem brain researchers. In addition to the MIAME guidelines, in postmortem gene expression profiling experiments, we strongly encourage reporting the following parameters for each studied subject: 1) age, 2) race, 3) gender, 4) brain pH, 5) RNA integrity, 6) Diagnostic and Statistical Manual of Mental Disorders diagnosis, 7) postmortem interval, 8) cause of death, 9) agonal state, 10) comorbidity, 11) drug abuse and smoking history, 12) treatment history, 13) family history of neurological or psychiatric disorders, and 14) hospitalization history. A subject table, detailing as many of these parameters as possible, should accompany every transcriptome profiling dataset.

Even with comprehensive patient information and full disclosure of the experimental data in MIAME format, comparing tran-