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 Model Organism (MIAME) 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

In order to move forward in understanding the processes of schizophrenia, we must carefully examine the challenges we are facing in the field 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. 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. 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. 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.

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—these are often inferred from study to study.
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 brains. 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, empirical evidence suggests that there may be a strong link between gene expression changes and genetic susceptibility within human brain disorders (1,3,7). Regulator of G-protein signaling 4 (10,38-41), disrupted in schizophrenia 1 (42,43), glutamic acid decarboxylase 67 (25,44), dysbindin (45,46), metabolotropic 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.nlm.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 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-
scriptome datasets remains extremely challenging (1-3,69). In general, comparisons can occur by two different mechanisms. First, we can compare the outcome of the experiments, asking if a gene or pathway was changed across multiple datasets. This method relies on the already reported outcome, which is provided by the investigators who generated and compared the dataset. Thus, we accept the data normalization procedures and comparisons performed by the various investigators, regardless of the methodological differences between their analyses. If concordant data are identified across many datasets and cohorts, such data are very powerful and conclusive. However, negative outcomes of such comparisons are hard to interpret—methodological differences can obliterate the findings, giving rise to a high rate of type II errors (obtaining false-negative results). In the second approach, the analysis is performed using the raw data generated by different laboratories—disregarding the analyses and conclusions of the investigators who performed the microarray experiments (for an example of such study, see [70]). Ideally, the raw microarray data are preprocessed and analyzed together, considerably gaining statistical power. However, such post hoc comparisons postulate that there are no major technical confounds between the various experiments contained within the combined dataset. Unfortunately, this is almost never the case, as each of the platforms has inherent biases (71). Even if the same microarray platform and processing procedures are used (and most often they are not), the operators, batches of reagents and microarray processing equipment, and instrument settings are all different. Furthermore, the samples were processed in batches, and all these experimental variables may (yet again) give rise to a large number of type I and type II errors (72). Most of such post hoc meta-analysis should be coupled with stringent false discovery procedures (70), and whenever possible, new discoveries should be validated using an independent wet laboratory method (e.g., in situ hybridization or quantitative polymerase chain reaction).

Transcriptome Profiling of Postmortem Tissue Will Continue in New Directions

Alternative splicing is a critical mechanism by which protein diversity is increased without significantly increasing genome size (73). It is estimated that >80% of genes are alternatively spliced in the human genome (74). In the brain, alternative splicing is more a rule than the exception (75), and it is currently probable that at least 15% (and perhaps as many as 50%) of human genetic diseases arise from mutations either in consensus splice site sequences or in their enhancers and silencers. Over the last several years, it became clear that gene expression changes in schizophrenia do not equally affect all the alternate transcripts of the same gene. For example, neuregulin 1 (NRG1), one of the most replicated and best characterized schizophrenia susceptibility genes, gives rise to many structurally and functionally distinct isoforms, through alternative promoter usage (76). However, in the hippocampus of subjects with schizophrenia only, the NRG1 type I transcript showed significant upregulation, while a novel NRG1 type IV isoform expression was dependent on a single SNP within the risk haplotype in both diseased and control subjects (77). Based on these findings, the authors proposed that altered transcript regulation is a potential molecular mechanism behind the genetic association of NRG1 with schizophrenia. Unfortunately, to date, very few (if any) true transcriptome splice variant assessment studies have been performed on postmortem brain tissue. Yet, with the emergence of the exon-expression assessing DNA microarrays (78), these experiments can be performed at relatively low cost, opening an important new line of investigation.

However, the newly developed, sequencing-based expression analysis methods (commonly referred to as RNA-seq) will fundamentally change our whole-genome expression profiling efforts (79,80). The idea of sequence-based measurement of gene expression levels has been intriguing scientists for almost two decades, and led to the development of serial analysis of gene expression (81), cap analysis of gene expression (82) and massively parallel signature sequencing (83). Today, as the tools, methods, and throughput of the initial sequencing technologies mature, we entered the era of whole transcriptome sequencing (84), where the abundance of RNA-derived complementary DNA can be efficiently assessed across the whole transcriptome (80). The newest, next generation sequencing platforms encompass quite diverse technologies based on microelectrophoretic methods, sequencing by hybridization, real-time observation of single molecules, and cyclic-array sequencing (85), which can all be adopted for analysis of messenger RNA transcripts. Of these, the concept of cyclic-array sequencing appears to be the most promising technology and is based on sequencing of a dense array of DNA features by iterative cycles of enzymatic manipulation and imaging-based data collection (86,87). There are various implementations of this technology (e.g., 454 Genome Sequencers, Roche Applied Science, Branford, Connecticut; Illumina Genome Analyzer, San Diego, California; the SOLID platform by Applied Biosystems, Foster City, California) (84), but their workflows are conceptually similar: libraries are prepared by random fragmentation of DNA and subsequent in vitro ligation of common adapter sequences. Then, the generated fragments are sequenced by de novo synthesis using serial extension of primed templates. Finally, data are acquired by imaging of newly incorporated, fluorescently labeled nucleotides across the full array at each cycle. Once the sequencing is completed, the various sequence fragments are aligned, assembled, and annotated to the genome. The obtained data are more comprehensive than any report provided by DNA microarrays—the RNA-seq data provide information about the abundance of unknown transcripts, RNA processing events (e.g., alternative splicing, RNA editing), SNPs, and mutations, all in a single dataset (79). Furthermore, at least theoretically, as the data are sequence-based, they are largely platform independent, allowing direct comparisons across datasets generated at various time points and using diverse technologies (88). Interestingly, the power of RNA-seq approaches raises an interesting question: will DNA microarrays become obsolete in the near future (89)? Will the new technology give us fundamentally new insight into disease mechanisms of mental disorders? Only time will tell. However, we can be certain that the main challenge will remain the interpretation of the data: can we generate biologically salient interpretations of complex datasets that are worthy of follow-up experimentation.

Finally, we must also shy away from expression profiling bulk postmortem tissue (90). While the newly developed bioinformatic tools can assign gene expression changes to various types of brain cells (59,91,92), they will never be able to replace targeted, cell-type specific transcriptome profiling efforts. The tools to perform such analysis have been in our hands for at least a decade (93): laser dissection microscopy, coupled with immunohistochemistry and DNA microarray profiling, is a very powerful, yet underutilized, experimental approach. To date, we have created relatively few cell-type specific transcriptome datasets from postmortem tissue (20,27,29,94-96), and if we want to put the transcriptome data in biological context, this line of research must be expanded.

Integration of Knowledge Must Become Our Priority

Although a number of transcriptional changes in schizophrenia have been clearly identified, validated, and replicated, many of the
most critical questions remain unanswered. Which of the most important gene expression changes are related to the pathophysiology? Which of the identified gene expression changes related to the fundamental disease process reflect the lifestyle of patients, treatment of the disease, or relationship to compensational mechanisms? What parts of the transcriptome changes translate to behavioral changes in patients? Are there subclasses of gene expression patterns that correspond to patient endophenotypes? How can we use the disturbed transcriptome signature to develop novel, knowledge-based therapeutic agents? These questions cannot be answered by transcriptome profiling of postmortem brain or by any single other technical method or experimental model. The answers will have to come from obtaining converging evidence across multiple lines of research, data integration, and cross talk between various disciplines. Animal models mimicking postmortem findings (97), electrophysiology, imaging studies, in vitro assessment techniques, neurochemical methods, large-scale genetic and various “omic” studies, and neurocognitive evaluations are all various pieces of the same puzzle. Perhaps the best example of this converging evidence approach is the recent study from the Weinberger laboratory on the role of the KCNJ2 gene in schizophrenia (98,99). The initial observation of a disease-associated SNP near the KCNJ2 gene transcription start site was followed up by bioinformatic analysis, structural and functional brain imaging, postmortem gene expression studies, and electrophysiology. Alone, none of these individual findings are very compelling, but together they strongly support a notion that KCNJ2 gene dysfunction may be a strong contributor to the disease phenotype.

However, we need to integrate findings on a larger scale. Ultimately, the integration of all the existing archival data can uncover symptom-related molecular pathways, identify evidence-based novel targets, and develop novel drugs that target the fundamentally involved disease pathophysiology. The recently developed Stanley Neuropathology Consortium Integrative Database (http://sncid.stanleyresearch.org) (100) is a huge step toward establishing easy to use, flexible, and freely accessible data integration resources. The Stanley Neuropathology Consortium Integrative Database includes 1749 neuropathological markers measured in 12 different brain regions in 60 human subjects (15 each schizophrenia, bipolar disorder, depression, and unaffected control subjects). This resource allows us to use various statistical tools and it integrates data from DNA microarray, RNA expression, protein expression, cytarchitectonict, and other studies, greatly facilitating novel discovery.

Conclusions

In summary, postmortem brain expression profiling will continue to be critical for identifying disease-associated pathology and pathophysiology. We must better understand the cell-type and region-specific transcriptome changes in schizophrenia, with special attention to splice variant assessments. To achieve this, new technologies (such as RNS-seq methods) must be explored, validated, and adapted for use in postmortem experiments. The identified molecular disturbances should be considered valuable leads for mechanistic investigations, including generating model animals. Finally, we must recognize that without better integration of data from all fields of schizophrenia-related research, we will just continue to examine the proverbial elephant in the dark room.

KM’s work is supported by National Institute of Mental Health Grants R01MH067234 and R01 MH079299.

www.sobp.org/journal

We are grateful to Dr. Ron Emeson and Martin Schmidt for helpful suggestions and edits of the manuscript.

All authors report no biomedical financial interests or potential conflicts of interest.


15. Iwamoto K, Bundo M, Kato T (2005): Altered expression of mitochon-
dria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. Hum Mol Genet 14:241–253.


vides concurrent validation of SELENBP1 up-regulation in schizophre-


S. Horváth et al.


