BrainSeq: Neurogenomics to Drive Novel Target Discovery for Neuropsychiatric Disorders

BrainSeq: A Human Brain Genomics Consortium*

*Correspondence: drweinberger@libd.org
http://dx.doi.org/10.1016/j.neuron.2015.10.047

We outline an ambitious project to characterize the genetic and epigenetic regulation of multiple facets of transcription in distinct brain regions across the human lifespan in samples of major neuropsychiatric disorders and controls. Initially focused on schizophrenia and mood disorders, the goal of this consortium is to elucidate the underlying molecular mechanisms of genetic associations with the goal of identifying novel therapeutic targets. The consortium currently consists of seven pharmaceutical companies and a not-for-profit medical research institution working as a precompetitive team to generate and analyze publicly available archival brain genomic data related to neuropsychiatric illness.

Motivation for the Consortium
It has become a virtual cliché to opine that there is an unmet medical need for the development of novel medicines for psychiatric disorders, as novel treatment approaches have stalled for decades (Hyman, 2012). Schizophrenia in particular is facing a crisis. There has been a complete lack of success in delivering drugs against novel therapeutic targets that show consistent efficacy in clinical trials, there has been only marginal success in treating the negative and cognitive domains that are not treated by current therapies, and there are lingering adverse effect profiles with these same compounds (Coyle, 2006). Developing the next generation of central nervous system (CNS) therapies for psychiatric disorders requires a deeper understanding of the underlying molecular mechanisms contributing to illness based on human data.

It is generally believed that the application of human genetic findings to drug discovery may provide a rational entry point to better understand disease mechanisms at the phenotypic, molecular/ cellular, and neurocircuitry levels. Recent successes in psychiatric genetics include the publication of 108 significantly associated common loci for schizophrenia by the Psychiatric Genomics Consortium (PGC) identified by genome-wide association analysis (GWAS) (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) and the identification of multiple rare de novo mutations associated with autism spectrum disorders (De Rubeis et al., 2014). Such advances instill optimism that new therapeutic hypotheses can be derived, and ultimately tested in the clinic, based on models that will be informed by an understanding of the molecular mechanism underlying the genetic association (Schubert et al., 2014). However, GWAS is a variant-based association method: it identifies often broad genomic loci containing multiple, sometimes hundreds, of correlated disease-associated variants and usually does not identify the disease-associated gene(s). More importantly, the association alone says little about the biological function or role of the locus in disease. GWAS results therefore lead to four important questions: (1) which variant(s) within the identified locus are responsible for the disease association, (2) which gene(s) do they act through, (3) is there a specific pathogenic transcript, and (4) do the functional consequences relate to disease-associated altered expression?

The first step in translating these exciting genetic findings into transformative medicines is the mapping of SNP-to-gene-to-function, which is critical to link the information from genetic association to disease biology through the functional impact of implicated genetic variations. The availability of large transcriptional datasets from relevant case and control tissues is a critical component for such an analysis. As well as enhancing our general understanding of disease biology, this will aid in drug development by providing a clue to a pathogenic gene product and insight to molecular directionality of the associated product, thus initiating model building based on a molecular mechanism of risk to close in on potential new targets.

Here, we discuss a new precompetitive initiative launched by the Lieber Institute for Brain Development (LIBD) with pharmaceutical industry partners (Astellas Pharma, AstraZeneca, Eli Lilly and Company, F. Hoffmann-La Roche, Johnson and Johnson, Lundbeck and Pfizer) to take advantage of the emerging and previously unprecedented genetic knowledge of psychiatric disorders and technical advances in the analysis of gene expression in brain tissue: BrainSeq, A Human Brain Genomics Consortium. Utilizing the growing brain tissue repository at the LIBD, with over 1,900 human postmortem neuropsychiatric disease and control samples available, the primary goal of BrainSeq is to generate and analyze spatial and temporal neurogenomics data (e.g., genotype, RNA sequence, and DNA methylation) and to establish a public database of these results. Short-term goals of the consortium are to determine how neurogenomics data can be used to elucidate the underlying molecular mechanisms of the genetic associations being discovered for psychiatric disorders, as well as the development of new bioinformatics tools and industry standards for scientific discovery in neurogenomics. The long-term goal is to develop new preclinical hypotheses that can drive the nomination of novel treatment targets. While initially focused on schizophrenia and mood disorders and recently identified common risk loci (Schizophrenia...
The genetic architecture of complex disorders involves primarily sequence variation that influences gene regulation (e.g., transcript abundance, splicing, novel transcript architecture, translation efficiency) rather than protein sequence (Fu et al., 2013). Thus, the molecular identification of genetic regulation of the transcriptome is a critical step in understanding genetic mechanisms and discovering novel isoforms that may underlie disease mechanisms and that may be targetable for the development of new therapies. In other words, regardless of the specific molecular mechanism of the genetic variation (e.g., promoter regulation, splicing, microRNA, long non-coding RNA, epigenetic processes), we argue that if the variant is not a protein coding variation, the primary readout should ultimately manifest as an effect in the transcriptome (see Figure 1). Even for regulatory variants affecting translation, it is possible to elucidate their impact by measuring the transcriptome from purified translating polyribosomes (Sterne-Weiler et al., 2013), a project that can be initiated in future stages of BrainSeq.

As illustrated in Figure 2, the principle of gene-to-drug approaches is based on identifying a molecular mechanism in the transcriptome that accounts for the clinical association and then building cell and animal models based on the molecular species identified. This represents an approach to model building with a high level of construct validity, modeling the specific molecular pathology of illness and genetic risk, in contrast to traditional models based on overall gene up- or downregulation. The ideal convergence of molecular association would be the identification of a specific transcript associated both with the illness state and with genetic risk, and with the risk-associated genotype predicting the same directionality of expression difference between cases and controls. In such a case, knowing the directionality of the molecular mechanism underlying the clinical association will suggest how to target it pharmacologically (i.e., with antagonists or agonists). Recent data suggest that risk-associated loci for psychiatric disorders based on common variation influence expression of specific, often previously unannotated splice variants (Tao et al., 2014), illustrating the importance of in-depth RNA characterization to identify potentially pathogenic transcripts, which may be of relatively low abundance. The risk-associated transcript and its expression differential is the specific construct for building preclinical models.

Figure 1. Mechanisms of Biological Risk from Noncoding Genetic Variation
Institute, and The Johns Hopkins University. The BrainSeq consortium has committed initial resources for completion of four phases of RNA-seq processing and analysis, based on specific brain regions and clinical diagnoses. The BrainSeq Phase I tissue cohort consists of a diagnostic mix of 746 postmortem human brain samples from the dorsolateral prefrontal cortex. Phase II samples consist of 200 patients with schizophrenia and 300 controls from mid hippocampus. Phase III involves head of caudate tissue from a similar sample of patients with schizophrenia and controls. Phase IV consists of hippocampal and medial prefrontal cortex tissues from controls and patients with bipolar disorder and major depression, who are psychotropic drug-free at time of death. All brain donations are obtained by verbal, witnessed informed consent with the next-of-kin (see Table S1).

Consenting and Clinical Review Process
The LIBD conducts retrospective clinical diagnostic reviews on every donation, which consist of: (1) a 34-item telephone screening with the next-of-kin on the day of donation conducted by a board-certified psychiatrist or board-certified neuropsychologist to gather demographic, medical, medication, social, psychiatric symptoms, treatment, and other clinical history; (2) a macroscopic and microscopic neuropathological examination by a board-certified neuropathologist to screen for neuritic pathology and to determine lifetime psychiatric and neurodegenerative disorders; (3) autopsies, to determine lifetime psychiatric diagnoses according to DSM-IV/V. Additionally, for every psychiatric case, extensive medical records review, contact with treating physicians, and/or family informant interviews (LIBD psychosocial autopsy interview, the Structured Clinical Interview for DSM-IV Disorders [SCID], or Mini International Neuropsychiatric Interview [MINI]) are conducted, where applicable and available. A master’s-level clinician summarizes all of these data into a neuropsychiatric narrative summary, which is then reviewed and scored by two board-certified psychiatrists, to determine lifetime psychiatric diagnoses according to DSM-IV/V.

Tissue Quality Measures
Every case has a postmortem interval (in hours) calculated as the difference between the time the brain was frozen from the time the patient was pronounced or estimated to have last been alive if the death was unattended. Every brain has a pH measure (in cerebellar tissue) for determining basic tissue quality. RNA Integrity Number (RIN) is measured on specific brain regions as they are dissected and processed for specific research studies (e.g., dorsolateral prefrontal cortex [DLPFC] RIN, hippocampal RIN, caudate RIN, medial prefrontal cortex RIN, etc.), prior to the RNA-seq data described below.

Sample Selection and Processing
All samples selected for molecular genetic analyses are dissected from deep-frozen tissue in a standardized routine by the same individual, using a dental drill to minimize tissue injury and RNA degradation. The initial four phases of the BrainSeq work are focused on dorsolateral prefrontal cortex (Brodmann area [BA] 9, 46), mid-hippocampus, caudate nucleus head, and medial prefrontal cortex (BA 32), because these are brain regions consistently and repeatedly identified as functionally altered in patients with schizophrenia and mood disorders, and the functional connection between these regions also shows alterations in patients (Cocchi et al., 2014; Rasetti et al., 2014; Rasetti and Weinberger, 2011; Tan et al., 2006). Moreover, recent work using in vivo functional imaging approaches has identified physiological alterations in these brain regions as being associated with increased genetic risk for these disorders in the healthy unaffected siblings of patients (Callicott et al., 2003; Rasetti et al., 2011; Sambataro et al., 2013).

Human Brain Neurogenomic Data in BrainSeq
The large-scale datasets generated by BrainSeq provide an unprecedented opportunity to interrogate the functional consequences of genetic risk variants on the transcriptome and the epigenome in the same subject samples (see Table S1).

Genotype/Genetic Data
Imputation from observed microarray-based genotyping yields more complete genetic information on every sample. Briefly, all subjects were genotyped using cerebellar DNA with an Illumina genotyping microarray platform—the majority of genotyping utilized the Illumina Human1M-Duo BeadChip with subsets of samples genotyped using the Illumina HumanHap650Yv3 BeadChip, Omni2.5 BeadChip, or Omni5 BeadChip. Pre-imputation quality control steps, performed...
separately for each chip type, largely mirror recommended guidelines for pre-GWAS data processing (Anderson et al., 2010) and involve removing low-quality and/or rare variants. Cleaned observed data on each genotyping platform are phased into haplotypes using SHAPEIT (Delaneau et al., 2013) and imputed to the 1000 Genomes Phase 3 variant set reference panel for the autosomal chromosomes and Phase 1 variant set for chromosome X, using Impute2 (Marchini et al., 2007). Post-imputed genetic data are then merged across genotyping platforms, resulting in over 80 million genetic variants, consisting of single nucleotide variants (SNVs), insertions and deletions (indels), duplications and repeat elements, per individual. There are approximately 7.5 million common and high-quality variants (minor allele frequency, MAF, >5%; Hardy-Weinberg Equilibrium, HWE, p value > 10^{-6}; and non-missing genotype data across >90% of the subjects), which can be correlated with transcriptome and epigenome data to identify the regulatory role of risk-associated genetic variation in the human brain.

**RNA Sequencing/Transcriptome Data**

Following RNA extraction from relevant brain regions, RNA sequencing libraries are constructed following the manufacturer’s protocols using two different Illumina kits: polyA+ selection for DLPFC samples (see additional details in Jaffe et al., 2015) and RiboZero Standard depletion for DLPFC, hippocampus, caudate nucleus, and medial prefrontal cortex samples. These RiboZero sequencing libraries are also prepared with ERCC spike-ins for data normalization (Risso et al., 2014) and also employ a strand-specific protocol. Sequencing libraries are barcoded to allow for multiple samples to be run across different flow cells and lanes to reduce potential biases induced by individual lanes, and samples resequenced using the Illumina HiSeq 2000 and HiSeq 3000 paired-end 100-bp reads targeting approximately 100 million reads per sample. The resulting sequencing reads are assessed using the FastQC tool, and the metrics are retained for each sample for downstream analysis. These sequencing reads are aligned to the genome using a splice-aware aligners TopHat2 (Kim et al., 2013), STAR (Dobin et al., 2013), or GSNAP (Wu and Nacu, 2010), and expression levels for different features can be summarized for each sample, for example across annotated genes and exons (Liao et al., 2014), splice junctions, expressed sequences (Jaffe et al., 2015), and transcripts (Frazee et al., 2015; Trapnell et al., 2013).

**DNA Methylation/Epigenetic Data**

Following DNA extraction from relevant brain regions, Illumina HumanMethylation450 BeadChip microarrays (Sandoval et al., 2011) are run on each sample following bisulfite conversion using the manufacturer’s protocols. Resulting image files (.idat) are processed and quality controlled using the minfi Bioconductor package (Aryee et al., 2014), resulting in proportion DNA methylation levels for >485,000 CpG sites across the epigenome on each sample.

**Leveraging Neurogenomics Data to Advance Treatment of Schizophrenia**

Schizophrenia and bipolar disorder are common, debilitating, neuropsychiatric disorders with enormous public health costs. Although efficacious, current treatments are far from ideal, as many patients fail to respond, as about one third of patients with schizophrenia are treatment-resistant (Smith et al., 2009). Of those who do respond, many only improve partially, leaving residual negative symptoms and cognitive deficits. In addition, adverse drug reaction and side effects remain major issues with both typical and atypical antipsychotics (Leucht et al., 2013), leading to treatment failure because of lack of tolerability and poor compliance. Antipsychotics, especially atypicals such as clozapine, are multiple-target drugs, i.e., they bind to multiple neurotransmitter receptors and other molecules. However, of all the molecular targets of antipsychotics, the occupancy of the dopamine D2 receptor remains a necessary condition for antipsychotic activity of all currently marketed drugs, despite attempts to develop alternative treatments (Ginovart and Kapur, 2012).

The success of the PGC (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) in identifying 108 loci associated with schizophrenia provides the opportunity to refine, and move away from, prior simple neurotransmitter hypotheses of schizophrenia. Interestingly, the dopamine D2 receptor maps to one of the 108 loci, possibly supporting the promise of the GWAS approach in identifying drug targets. Of the many other genes within loci identified by the PGC, many are highly attractive and tractable drug targets, some of which have previously been the target of drug development programs. These include GRIN2A, which encodes the NMDA (N-methyl-aspartate) receptor subunit NR2A, a key mediator of synaptic plasticity (Yashiro and Philpot, 2008), and GRIA1, which encodes glutamate receptor 1 (GluR1; also known as GluA1), an AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazol- lepropionic acid) receptor subunit that mediates fast synaptic transmission and is important for learning and memory (Sanderson et al., 2008). The PGC regions also contain SRR, which encodes serine racemase, an enzyme that catalyzes the conversion of L-serine to D-serine and is an essential co-agonist and activator of NMDA receptors (Campanini et al., 2013); CACNA1C, which encodes the alpha 1C subunit of the L-type voltage-gated calcium channel (Bhat et al., 2012); and CACNA1I, which encodes the pore-forming alpha subunit of the Cav3.3 T-type calcium channel involved in synaptic plasticity and long-term potentiation. GRM3 encodes metabotropic glutamate receptor 3 (mGluR3) predominantly expressed in astrocytes and is also implicated by the PGC GWAS results. Drugs with activity against mGluR2 and mGluR3 have been extensively explored as a potential therapeutic target in schizophrenia, but clinical studies to date have shown mixed results with regard to efficacy compared to current standard of care, and Phase III trials have not been successful so far (Li et al., 2015). It is conceivable that the failure to show efficacy in large Phase III studies is because mGluR3 has not been selectively targeted. Other reasons for this failure have been debated elsewhere (Dunlop and Brandon 2015). Characterization of the molecular association in human brain may clarify how the genetic association influences the gene product, specifically whether there is a novel GRM3 isoform in brain and whether it is up- or downregulated in illness.
The key to advancing genetic findings from psychiatric disorders into drug development is the definition of the precise molecular isoforms of these genes and their products that contribute to disease and their dysfunctional role in physiological pathways. By first defining the pathways and networks perturbed according to transcriptome analyses, it becomes possible to prioritize the most critical components of these pathways to target through in silico systems pharmacology or other bioinformatics approaches. We are optimistic that the possibility of identifying disease-associated gene isoforms, pathways, networks, and pathophysiological mechanisms through genomic research will catalyze drug discovery and development with novel mechanisms of action based on causation, not phenomenology. In so doing, this will also allow us to revisit previously unsuccessful targets armed with new knowledge about their biological role in disease and about patient subpopulations in which a therapeutic aimed at these targets may be efficacious. Indeed, by focusing on gene variants with functional impact at molecular levels, their effect on clinical intermediate phenotypes could be investigated in living patients, suggesting subgroups that may benefit more from target modulation.

**BrainSeq Expected Outcomes to Identify Novel Treatments**

We expect that the BrainSeq data analysis and integration approach will provide us with sufficient genetic substrate to generate novel and testable biological and therapeutic hypotheses for schizophrenia, mood disorders, and related neuropsychiatric conditions. Based on an improved understanding of the transcriptomic and epigenomic effects of genetic disease-associated risk variants in relevant brain regions, we expect to be able to identify a significant subset of the truly causal genes and isoforms implicated by GWAS signals. Such causal gene mapping, combined with the ability to assign a directionality hypothesis to a disease locus as well as prior literature evidence, will significantly improve our chances to accurately assess the underlying disease biology and identify a path toward novel and better therapeutics.

**Data- and Biospecimen-Sharing Plans**

In addition to an active precompetitive research project, a fundamental principle of this consortium is to provide the scientific community at large with a valuable resource for future research. Data will be made available to approved investigators through a web-based portal hosted at LIBD through the BrainCloud application (in raw and/or processed forms) to allow for additional analyses and comparison with other datasets outside this consortium. The consortium will endeavor to publish data and consortium research results promptly to make them publically available. As a rule, all data will be released within one year of completion of the consortium data analysis. Data from the control and schizophrenia cohort of the Phase I of the consortium work plan is expected to be released by early 2016. The Phase II data results will be released in 2016. Phase III and IV data will be available in 2017. At any time, the consortium is open to accepting new members to expand the scope of the work plan and to allow early access to the research results and participation in ongoing analyses.

As the ability to perform additional studies on brain and tissue samples contained in the consortium data is anticipated, a procedure will be put in place to propose such studies. This may include molecular analysis of specific cell populations using laser capture and cell-sorting techniques, targeted resequencing with longer read technologies, whole-genome bisulfite sequencing, and chromatin state assays, as well as peptide sequencing.

**Future Plans**

We envision that BrainSeq will produce valuable information in its first planned phases about the molecular pathways associated with schizophrenia, mood disorders, and related conditions in the DLPFC, medial prefrontal cortex, hippocampus, and caudate, which shall facilitate the development of novel mechanistic and therapeutic hypotheses for the discovery and development of novel treatment targets. As the cost of next-generation sequencing technologies continues to decrease, the consortium anticipates the potential addition of whole-genome sequencing approaches to the data generation pipeline, as well as analysis of small RNA species to the analytic pipeline.

In closing, the BrainSeq precompetitive multi-partner consortium, applying a multi-omics approach with a highly collaborative spirit, combined with clear data-sharing policies and pre-defined analysis plans, allows for ready scalability and the potential to significantly accelerate psychiatric drug discovery. Yet we hope that our efforts will ultimately have broader applicability to other complex disorders, where the understanding of genotype-phenotype relationships is critical to elucidate the underlying molecular nature of disease etiology.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes two tables and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2015.10.047.

**CONSORTIA**


**ACKNOWLEDGMENTS**

We thank the staff of the Office of the Chief Medical Examiner of the State of Maryland, under the direction of Dr. David Fowler, for their collaboration. We also gratefully acknowledge the brain donors and their families, without whom this work would not be possible. We wish to acknowledge the vision and support of the Lieber and Maltz families. C.R.S., P.O., J.J., J.R.W., H.S.X., and A.R.W. are employees and stockholders of Pfizer, Inc. E.D., L.E., and T.K.-T. are employees and stockholders of F. Hoffmann-La Roche, AG. D.C.A., J.N.C., D.A.C., H.W., B.E., P.E., Y.L., L.N., C.R., J.E.S., R.M.S., and H.-R.Q. are employees and stockholders of Eli Lilly and Company. K.M. is an employee of TransThera Consulting Co. and an ex-employee of Eli Lilly and Company. M.D. is an employee of Genentech, Inc.

REFERENCES


