

Translating Genetic Risk Loci Into Molecular Risk Mechanisms for Schizophrenia

Nicholas J. Bray* and Matthew J. Hill

MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK

*To whom correspondence should be addressed; MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Hadyr Ellis Building, Maindy Road, Cardiff, CF24 4HQ, UK; tel: +44-(0)2920-688368, fax: +44-(0)2920-687068, e-mail: BrayN3@cardiff.ac.uk

Introduction

The past decade has witnessed major advances in our understanding of the genetics of schizophrenia. Large, consortia-led genomic studies involving thousands of patients and controls have identified genetic loci associated with risk for the disorder at unprecedented levels of confidence. As expected from a condition associated with reduced fecundity, variants that have a large impact on risk for schizophrenia are invariably rare, occurring even in patients at frequencies considerably less than 1%. These include copy number variants (CNVs) which delete or duplicate large segments of DNA, often encompassing multiple genes.¹ Alleles of strong effect on schizophrenia risk also potentially include loss-of-function de novo mutations within protein coding DNA sequence that have recently been identified through exome sequencing studies.² It is clear, however, that schizophrenia also involves the action of many genetic variants that are common in the general population (frequencies > 5%). Although these have individually small effects on risk for schizophrenia (odds ratios typically < 1.2), they collectively account for a sizeable fraction of the variance in liability to the disorder.³ In the largest genome-wide association study (GWAS) of schizophrenia to date, the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC) reported genome-wide significant association between the disorder and common variation at 108 independent genetic loci.⁴ In the following short article, we outline some of the questions that first need to be addressed in translating these latter findings into an improved understanding of common molecular risk mechanisms for schizophrenia, and introduce functional genomic technologies that can be applied for this purpose.

Which Genes at GWAS Risk Loci Confer Susceptibility?

It is first important to note that the identification of risk loci through GWAS does not necessarily mean that the

actual susceptibility genes at these loci (ie, those that are functionally altered by the risk variants) have been confidently identified. One reason for this is that genotypes at neighboring DNA variants often correlate within a population (a phenomenon known as “linkage disequilibrium”), with the result that association signals can span large genomic regions, often encompassing more than 1 gene. In addition, it is now known that genes typically give rise to multiple RNA transcripts, which may differ in their expression profile as well as function, and it is likely that, in many cases, only specific transcripts of a given gene will be affected by schizophrenia risk variation.⁵ Further, and as we will see in the next section, the functional nature of many of the risk variants themselves means that it might not always be the nearest gene that is affected.

How are Susceptibility Genes Functionally Altered by Genetic Risk Variants?

Genetic associations at 10 of the 108 schizophrenia risk loci reported by the PGC (2014) credibly index known, commonly-occurring exonic variants that change the amino acid sequence of the encoded proteins (“nonsynonymous polymorphism”).⁴ As these potentially constitute some of the functional variants mediating risk for schizophrenia, they are good candidates for which to directly assess effects on protein function. However, the vast majority of loci exhibiting genome-wide significant association with schizophrenia cannot be accounted for by variants that impact upon protein structure. Instead, these are likely to index functional variation within regulatory regions of the genome, which alter gene expression or splicing by interfering with the binding of molecules that drive these processes (eg, transcription factors [TFs], splicing factors, microRNA). This adds to the difficulty of identifying the actual susceptibility genes because, unlike variants that change amino acid sequence, regulatory variants can

be located large distances from the genes they regulate—even in neighboring genes.⁶ For accurate modeling of this type of genetic risk mechanism (and potentially also for therapeutic reasons), it is necessary to determine not only which gene transcripts are differentially regulated, but also the nature of the effect (eg, increased or decreased expression).

Where (and When) are These Effects Exerted?

Regulatory elements in which risk variants are likely to be located include promoters and enhancers that can be specific to particular transcripts and operate only in certain cells, at particular developmental stages or under certain biological conditions. It is therefore important to establish where and when schizophrenia risk variants exert their effects. Regulatory variation has been shown to have different effects across brain regions,^{7,8} and in the case of schizophrenia risk variation at the *ZNF804A* locus, to alter *ZNF804A* expression at a particular stage of human fetal brain development.^{5,9} Establishing the temporo-spatial nature of these effects will again be crucial for accurate modeling and for the potential development of therapeutic interventions that target these processes.

Approaches for the Functional Interrogation of Noncoding GWAS Risk Loci

In parallel with progress in schizophrenia genetics, there have been considerable advances in tools with which to functionally interrogate noncoding genomic loci. We now focus on some of these key functional genomic technologies and their application to schizophrenia risk loci.

One of the major scientific advances of recent years has been the identification of regulatory regions in non-coding DNA sequence throughout the human genome. These regions are characterized by open chromatin, making them accessible to the TFs that regulate gene expression. It is now possible to map these regions on a genome-wide scale using Next Generation Sequencing (NGS) after treating cell nuclei with enzymes that preferentially target accessible chromatin (eg, DNase-seq, ATAC-seq). At greater sequencing depths, it is possible to identify “footprints” in these regions created by the TFs bound to the DNA. The genome-wide binding of individual TFs can also be mapped using chromatin immunoprecipitation followed by NGS (ChIP-seq). A TF is hereby covalently linked to its interacting DNA, which is then sheared and selectively recovered using TF-specific antibodies prior to identification through sequencing. ChIP-Seq can also be used to predict the regulatory status of genomic regions by targeting characteristic histone modifications in chromatin. For example, promoters and enhancers are typically marked by histone methylations *H3K4me3* and *H3K4me1*, with the additional histone acetylation mark

H3K27ac indicating activation and the histone methylation mark *H3K27me3* indicating repression. Genomic regions subject to DNA methylation, a form of epigenetic regulation typically associated with transcriptional repression, can also be mapped by combining methods that assess the methylation status of cytosine residues with microarray or NGS technology.

As previously noted, it is now clear that regulatory regions of the genome can be highly cell-specific. The ENCYClopedia Of DNA Elements (ENCODE) and NIH Roadmap Epigenomics projects have released DNase-seq, ChIP-seq and DNA methylation data from a wide variety of human tissues and cell types, available as information on the UCSC Genome Browser (www.genome.ucsc.edu). Although it is difficult to confidently identify the functional variants underpinning a GWAS signal, such data can be used to prioritize those narrowed down by fine mapping at associated loci.¹⁰ In addition, it is possible to use these data to test whether GWAS signals are statistically enriched in regulatory regions utilized by particular cell types, thereby suggesting where (and possibly when) these variants are active. For example, the PGC (2014) tested for enrichment of credible risk variants at the 108 genome-wide significant schizophrenia loci in active enhancers in 56 human cell lines and tissues, finding significant enrichment in enhancers active in human brain tissue.⁴ As most existing data are from nonneural cells and tissues, the PsychENCODE project (www.psychencode.org) has recently been launched to map regulatory elements in regions of the developing and adult human brain, as well as in human neural cell systems. The potential of these DNA sequences (as well as variants within them) to drive gene expression can now be assessed in a high-throughput manner using massively parallel reporter assays, which use NGS to measure the number of barcoded molecules transcribed in association with particular sequences.

Underlining the principle that it might not always be the closest gene that is affected by regulatory risk variation, there has been a growing appreciation that chromosomal regions frequently fold in order to bring distant regulatory regions (eg, enhancers) in closer proximity to the genes they regulate. These chromosomal interactions can be studied using the chromosome conformation capture (3C) technique and its derivatives (eg, 4C, 5C, Hi-C). 3C-based techniques involve formaldehyde cross-linking of interacting sites in cells of interest, cutting of DNA with a restriction enzyme and a ligation reaction to join cross-linked DNA fragments. Whereas 3C uses polymerase chain reaction to investigate chromosomal interactions at specific candidate loci, recent methods (eg, Hi-C) make use of high-throughput technologies such as NGS to study chromosomal interactions on a genome-wide scale.

Although, to date, few studies have used 3C-based methods to interrogate schizophrenia GWAS loci,

these are likely to prove extremely valuable in identifying the primary transcripts targeted by genetic risk variation. Recently, Roussos and colleagues (2014) used 3C to elucidate the mechanism through which schizophrenia risk variation in an intron of the *CACNA1C* gene regulates *CACNA1C* expression.¹¹ A predicted enhancer region within this intron containing schizophrenia-associated variants was found to interact with the *CACNA1C* promoter region in human dorsolateral prefrontal cortex and neurons derived from human induced pluripotent stem cells (hiPSCs). Using a reporter gene assay, the authors further showed that the schizophrenia-associated allele within this enhancer drives lower transcriptional activity, consistent with its association with decreased *CACNA1C* expression in human brain tissue.

Ultimately, the effects of schizophrenia risk variation on RNA transcription, splicing and stability should be detectable in terms of the RNA expression of the affected transcripts in the relevant cells and tissues. Although in-depth studies of individual risk variants/genes continue to be informative,^{5,9} it is now possible to freely access datasets that combine genome-wide genotyping and gene expression analysis in the human brain in order to assess potential regulatory effects of risk variants. The majority of existing expression quantitative trait loci (eQTL) data have been generated using microarrays for gene expression analysis. For example, the BRAINEAC (www.braineac.org) resource uses exon-level array data to identify eQTL in 10 regions of the adult human brain,⁸ allowing assessment of effects on individual RNA transcripts. Several recent initiatives (eg, the CommonMind, Leiber Institute Pharma RNA-Seq and GTEx Consortia), are generating human brain gene expression data using NGS technology (RNA-Seq), which can be used to assess genetic effects on the splicing and expression of both known and unknown RNA transcripts. As for the approaches discussed above, assessments in a variety of tissues (eg, brain regions, cell types) are likely to be highly informative, because these studies have the potential to elucidate not only how individual susceptibility transcripts are affected by genetic risk variants, but also where and when these effects are manifest. Given that schizophrenia is hypothesized to have an early neurodevelopmental component, we are currently generating RNA-Seq and genotype data from a large collection of human fetal brain samples with which to assess regulatory effects of genetic risk variation in the developing brain.

Moving to Biology and Treatment

It is important to emphasize that the impact of a given DNA variant on risk for a disorder will not necessarily reflect the therapeutic value of targeting

the affected molecule or its biological pathway. For example, the closest known gene to one of the 108 genome-wide significant associations reported in the PGC (2014) study⁴ is *DRD2*, encoding the dopamine D2 receptor. Although the associated variant at this locus increases risk for schizophrenia by less than 10% (odds ratio of risk allele: 1.08), the dopamine D2 receptor is a primary target of all antipsychotic drugs. It is likely that the small effects on schizophrenia risk conferred by individual common variants reflect, in many cases, the impact of the functional variant on the encoded molecule: variants that have subtle effects on a gene's function (eg, minor changes in its expression) might have small effects on susceptibility, whereas those that have more damaging effects on the same molecule (eg, a CNV or loss-of-function exonic mutation) could have a greater effect on risk. Ultimately, the translation of GWAS into improved schizophrenia treatments may depend upon the extent to which the molecules encoded by susceptibility transcripts fall into common biological pathways, as well as their capacity to be targeted. Bioinformatic pathway analyses of genes at schizophrenia GWAS risk loci are already yielding interesting findings.¹² Future analyses will benefit from greater resolution of the affected genes/transcripts and their biological functions. As many of the molecules encoded by schizophrenia susceptibility transcripts are likely to be poorly characterized at the functional level, a huge research effort will be required to better understand the biology of these molecules once they have been confidently identified, particularly in cells and at developmental stages where they are affected by genetic risk variants. In this endeavor, investigators can already benefit from major advances in the derivation of human neural cell types (eg, through hiPSCs) as well as methods for editing genetic sequence in cell and whole animal systems (eg, CRISPR).

Conclusions

Large-scale GWAS have been successful in identifying many common genetic loci conferring risk for schizophrenia. A crucial next step is to determine which gene transcripts at these loci are affected by the risk variants, how they are functionally altered, and where and when these effects are manifest. There are now a variety of powerful functional genomic technologies that can be applied to address these fundamental questions. Given that effects of regulatory variants can be specific to cell and developmental stage, assessments in a variety of cells and tissues are required. These efforts should serve to guide functional investigations of schizophrenia susceptibility genes in human and animal model systems, with the ultimate aim of developing improved treatments for the disorder.

Acknowledgment

The authors have declared that there are no conflicts of interest in relation to the subject of this article.

References

1. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*. 2012;148:1223–1241.
2. Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014;506:179–184.
3. Ripke S, O’Dushlaine C, Chambert K, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*. 2013;45:1150–1159.
4. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427.
5. Tao R, Cousijn H, Jaffe AE, et al. Expression of ZNF804A in human brain and alterations in schizophrenia, bipolar disorder, and major depressive disorder: a novel transcript fetally regulated by the psychosis risk variant rs1344706. *JAMA Psychiatry*. 2014;71:1112–1120.
6. Smemo S, Tena JJ, Kim KH, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014;507:371–375.
7. Buonocore F, Hill MJ, Campbell CD, et al. Effects of cis-regulatory variation differ across regions of the adult human brain. *Hum Mol Genet*. 2010;19:4490–4496.
8. Ramasamy A, Trabzuni D, Guelfi S, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci*. 2014;17:1418–1428.
9. Hill MJ, Bray NJ. Evidence that schizophrenia risk variation in the ZNF804A gene exerts its effects during fetal brain development. *Am J Psychiatry*. 2012;169:1301–1308.
10. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93:779–797.
11. Roussos P, Mitchell AC, Voloudakis G, et al. A role for noncoding variation in schizophrenia. *Cell Rep*. 2014;9:1417–1429.
12. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci*. 2015;18:199–209.