Schizophrenia at a Genetics Crossroads: Where to Now?

Aiden Corvin*1
1Department of Psychiatry & Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Ireland
*To whom correspondence should be addressed; tel: + 353-189-62468, fax +353-189-63405, e-mail: acorvin@tcd.ie

Key words: schizophrenia/network/biological pathways

Introduction
These are interesting times for schizophrenia genetics. The last 5 years has seen unprecedented progress, moving us from debate about putative genetic models to an understanding that susceptibility involves a complex interplay of both common and rare genetic risk variants. Critical advances, achieved through the confluence of high-throughput genomics platforms and unparalleled collaboration (eg, the Psychiatric Genomics Consortium), have been the subject of excellent reviews elsewhere. The focus of this article is on a fresh challenge. How to move past gene discovery to understand pathophysiology and disease etiology?

So what do we know? In brief, this year is likely to see the confirmation by the PGC, at stringent levels of statistical significance, of more than 70 independent common risk alleles for schizophrenia from genome-wide association study (GWAS) analysis.5 To this we can add at least a dozen rare, structural genomic risk variants (chromosomal microduplications or deletions). Much still needs to be done. Because the common risk alleles individually have small effects on risk, collectively the more than 70 identified common risk loci may explain <5% of the total genetic variance in schizophrenia susceptibility (Ripke, personal communication). The structural variants have much larger effects (Odds Ratio [OR]=2–30) but are individually rare (present in <0.1% of controls) and are likely to make an even smaller contribution to the total risk. Present findings represent the first tangible pieces of a much larger genetic puzzle. That the genetic etiology is complex and only partially resolved has provoked criticism of the value of genomic research in schizophrenia, and medicine in general. This is to miss the point. The purpose of these studies is to understand biology, and this list of identified risk genes offers many new avenues for hypothesis-based research. In attempting to add more pieces to the puzzle, we know where to look, but until relatively recently have lacked the means to do so.

A Roadmap to Future Gene Discovery?
Large numbers of small genetic effects, explaining at least 25% of the genetic variance in schizophrenia risk, remain to be identified.6 Some effects will be so small as to allude detection. However, the recent PGC findings represent a more than 5-fold increase in confirmed risk loci, achieved by doubling the sample size studied.5 This confirms a trajectory of discovery similar to that of other common disorders including Type 2 Diabetes (T2D), Ulcerative Colitis (UC), Crohn’s Disease (CD) and Psoriasis.9 So gene discovery through GWAS analysis will continue to be a powerful tool.

Rare sequence mutations represent the vast bulk of genomic variation. The rapid evolution of genomic sequencing opens up this unexplored genomic terrain and represents a philosophical transition from investigating what we hold in common as populations to what makes us individually unique. Shifting along this continuum, from relatively common variants investigated through GWAS analysis, the next step will be low-cost, high-throughput array-based assays of lower frequency variants, an approach that has been successful for other common disorders.10 Based on sequencing technology, there are likely to be improved calling methods to tackle the hitherto unexplored influence of smaller structural variants (<50kb).11 Most obviously, this technology allows us to access the huge reservoir of rare, or private sequence mutations in the human genome.12

The challenges for rare variant discovery are beyond the scope of this article (for review, see Sullivan et al.3). In short these include genetic (the “noise” generated by the background mutation rate, genetic heterogeneity, reduced penetrance); phenotypic (variable phenotype expression); and logistical (a coherent management strategy for the vast amount of data generated) factors. At this point too few schizophrenia exome studies have been published to
allow meaningful comparison of results. However, from early reports, there appears to be an increased rate of de novo sequence mutations in schizophrenia cases compared with controls, and this may reflect, in part, the impact of greater paternal age. The transition to rare compared with controls, and this may reflect, in part, the rate of de novo sequence mutations in schizophrenia cases from early reports, there appears to be an increased rate subsets of individuals, with distinct genomic disorders for most but not all cases. There are already identified to represent disruption across a complex genetic network, and for most cases genetic susceptibility is likely on what we know, schizophrenia is a highly polygenic disorder, and for most cases genetic susceptibility is likely to represent disruption across a complex genetic network rather than individual gene effects. This will be true for most but not all cases. There are already identified subsets of individuals, with distinct genomic disorders (eg, 22q11.2 deletion syndrome, 1q21.1 microdeletion syndrome), presenting as the schizophrenia syndrome. To what extent these, essentially rare diseases, contribute to the total syndrome is still uncertain. Functional studies of these rare high penetrance risk loci or specific risk genes (eg, Neurexin-1, VIPR2, ZNF804A) will be important. But deeper biological insight will require understanding of how a list of genes orchestrates the dynamic set of cellular processes that underpin schizophrenia.

Approaching Schizophrenia as a Network Disorder

Translating genetic variation into disease mechanisms requires the coordination of a complex network of cellular and intercellular functions. This network, termed the human interactome, incorporates interactions at the level of gene families, wider protein-protein interaction (PPI), metabolic pathways, regulatory networks, microRNA-gene networks, and interaction between genes and environmental factors, which will vary over time throughout the human lifespan. Despite this complexity, network-based approaches to human disease are increasingly feasible. Resources for investigating the interactome are expanding rapidly (reviewed by Vidal and colleagues). Intuitively, progress is dependent on the quality of information available for network analysis. At the present time the disease literature reflects the fact that some aspects of the network (eg, PPI) are better annotated than others (eg, regulatory networks). From a genetics perspective, notable successes have been achieved for other common traits and disorders where larger samples have been put together, quicker, than for psychiatric disorders.

Until this year, the number of common risk loci available for schizophrenia analysis has been small (<15), so by necessity, analyses have included variants with much weaker association signals introducing more “noise” to the analysis. This has involved two types of approach. The first is based on risk profile scores accrued across large numbers of genetic markers as part of the “polygene score method” described by the International Schizophrenia Consortium. Hidden within these “noisy” schizophrenia risk profile scores are hundreds of small, unconfirmed genetic risk effects contributing to schizophrenia genetic variance. In support of the network concept, these effects are not random but are more likely to affect gene expression in adult brain than would be expected by chance and are disproportionately attributable to 2725 genes expressed in the central nervous system. The second, more traditional, approach has involved investigation of specific pathways with interesting provisional findings for PPI analysis of the DISC1 interactome; for association with Cell Adhesion Molecule pathways, and for a gene network potentially regulated by the microRNA mir-137.

Harnessing information on a larger set of genetic risk variants will allow for more powerful analysis of genetic networks in schizophrenia. Maximizing the information available will ideally require methods of incorporating common and rare genetic risk variants into pathway or network analysis. But stepping beyond genetics discovery, this will also require integration across different levels of the network. To understand gene function we need to incorporate information about gene expression across human brain development. But networks are dynamic, and it will also be important to consider how transcriptional mechanisms respond to discrete environmental risk factors and also to the impact of social networks.

What Can We Learn From Biological Networks?

Biological networks are not random but evolve and are structured based on a set of underlying organizing principles. By understanding these principles we can apply them to undertake a structured approach to investigating schizophrenia as a network disorder. First, if a network is random, most nodes in the network will have approximately the same number of links, and highly connected nodes (also termed hubs) will be uncommon. In contrast, biological networks tend to have a small number of highly connected hubs that hold the network together. Identifying hubs within the network may be an important step in understanding biological processes as they may represent targets for therapeutic intervention.

Second, in any network the paths between nodes are usually short. This means that most nodes in a biological network are separated by only a few interactions as has been demonstrated for both protein interaction...
and metabolic networks. Extending this principle, proteins that are involved in the same disease show a greater propensity to interact with each other. So if we identify disease genes, other risk genes are likely to be found in the same local vicinity, or module, of the network. The problem becomes one of trying to identify the topographical modular structure of the network. This can be achieved through a growing number of network clustering algorithms. That more approaches are being developed is indicative that none are ideal, and where modules are identified it may be important to validate through functional annotation experimentally (eg, in PPI networks) or across different levels of network function.

The attraction of this approach is that it can substantially reduce the search space for genomic studies. Identified modules or local subnetworks can be directly tested to identify enrichment of common risk variation and rare functional mutations. The classic example, from the rare variant literature, is the 23 genes implicated in inherited forms of ataxia identified through a subnetwork of 54 proteins. A very different approach, described by Rossin et al. (2011), evaluates the quality of evidence for PPI between identified risk genes and uses this information in building a network. With this method the authors demonstrated that proteins encoded by genetic loci associated with a number of disorders or traits (RA, CD, height and lipid levels) are significantly connected compared with random networks. Two different applications of the classical approach have been reported in schizophrenia. One of these methods, described by Gilman and colleagues, integrated both common and rare risk variation in their analysis. Both this, and the GWAS-based analysis described by Jia and colleagues, reported association with numerous different aspects of neuronal function. With more risk loci being identified for schizophrenia, these types of analytical approaches are likely to become more common with an increasing focus on identifying evidence for convergence across methods and data sets.

Third, interactions may be at the level of coexpression or shared function, rather than physical protein interaction. Applying a gene coexpression analysis method, Voineagu et al. (2011) recently reported discrete modules of coexpressed genes associated with autism and wider differences in transcriptome organization between autistic and normal brain. Interestingly, the transcriptome level changes showed convergence with known autism risk genes and genetic association signals, suggesting a plausible mechanism whereby a network of genes alters regulation of a transcriptional network mediating autism risk. Similar efforts to link genetic risk variants to gene expression networks are beginning to emerge in schizophrenia.

Fourth, highly connected nodes (hubs) that represent essential components in the network have particular properties. By virtue of having more interactions, hub proteins are likely to be involved in more biological processes and more diseases. One interpretation of the diverse developmental outcomes associated with structural risk variants (an illustrative example is 1q21.1 deletions) is that these loci represent examples of such “disease” hubs (see figure 1). So, careful characterization of the range of phenotypes associated with such loci will be important, and they are likely to represent key network components for targeted functional interrogation.

Fifth, common and rare risk variants may impact differently on the network. Much of our understanding of biological systems comes from simple model systems and is based on analysis of loss-of-function mutations. Most common risk variants are likely to have much more subtle effects on gene function. By way of illustration, the list of risk genes implicated in schizophrenia by common risk variants includes CACNA1C and TCF4, where functional mutations cause severe developmental syndromes (Timothy syndrome and Pitt-Hopkins syndrome, respectively). As such these might be seen as disease hubs. However, there are also likely to be points of divergence for common and rare risk variants in gene networks, where loss-of-function mutations of essential proteins are not compatible with life.

Finally, diseases may share hubs, but disease networks are also likely to overlap (see figure 1). This is a key emerging theme from investigation of other common disorders. For example, there are many shared loci involving autoimmune and inflammatory processes being identified across a range of common disorders (eg, CD, UC, T1D, Graves Disease, Coeliac Disease, and Psoriasis). There is significant evidence already for network overlap between schizophrenia and bipolar disorder and, to a lesser extent, autism. Cross disorder analysis is underway comparing schizophrenia, bipolar disorder, major depressive disorder, attention deficit hyperactivity disorder (ADHD), and autism. The findings from structural variants and for metabolic dysfunction in schizophrenia, argue for a wider analysis including other developmental (intellectual disability, seizure disorder) and metabolic disorders (obesity, lipid metabolism, T2D).

Conclusion

Beginning to put together the relationships between genes strikes to the heart of a fundamental challenge in schizophrenia research. Because we are learning from cancer research and other fields in medicine, clinical or even pathological diagnostics may bear little relation to the underlying molecular mechanisms of disease. Progress in these fields has required a transition from conceptual models of disease (or disorder) to mechanistic models of disease processes. The schizophrenia syndrome is likely to capture different molecular mechanisms, which represent discrete subtypes or even different diseases.
with the same common endpoints. By understanding the relationship between genes and identifying network subcomponents, these will become the models or building blocks for schizophrenia research. Such building blocks can be validated and functionally interrogated by model systems. This is likely to involve high-throughput screening in simple systems to identify common functions (eg, of mechanisms involving ion-channel function or cytoskeleton dynamics). Such mechanisms will, in turn, require more complex dissection, eg in mouse or human iPS models, with comparison across systems (for an elegant example, see Nithianantharajah). From such foundations we can advance new, focused hypotheses. Do subnetworks or molecular mechanisms map to neural circuits, perhaps identifiable by diffusion tensor imaging? How do network modules relate to clinical symptoms, outcome or treatment response? Or address wider questions. Is the burden of risk variants important to illness progression in high-risk populations? Does total genetic burden relate to outcome? How does genetic burden interact with environmental risk? None of these are abstract questions. We urgently need new strategies for prevention, early intervention, and treatment to tackle this most challenging of disorders.

**Funding**

Science Foundation Ireland (08/IN.1/B1916 to A.C.).

**Acknowledgments**

The author has declared that there are no conflicts of interest in relation to the subject of this study.

**References**

5. Ripke S. XXth World Congress of Psychiatric Genetics; 14-18 October 2012; Hamburg, Germany.


Schizophrenia at a Genetics Crossroads: Where to Now?


58. Corvin AP. Two patients walk into a clinic...a genomics perspective on the future of schizophrenia. *BMC Biol.* 2011;9:77.

