Schizophrenia Bulletin vol. 42 no. 3 pp. 538–541, 2016
doi:10.1093/schbul/sbw014
Advance Access publication March 18, 2016

What Next in Schizophrenia Genetics for the Psychiatric Genomics Consortium?

Aiden Corvin*[,1] and Patrick F. Sullivan[2,3]

[1]Department of Psychiatry and Neuropsychiatric Genetics Research Group, Trinity College Dublin, Dublin, Ireland; [2]Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC; [3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

*To whom correspondence should be addressed; Department of Psychiatry and Neuropsychiatric Genetics Research Group, Trinity College Dublin, Trinity Centre for Health Sciences, St James's Hospital, Dublin 8, Ireland; tel: +353-1-8962467, fax: +353-1-89634045, e-mail: acorvin@tcd.ie

Introduction

Over the last 8 years the Psychiatric Genomics Consortium (PGC; http://pgc.unc.edu) has fundamentally changed the landscape for psychiatric genetics research. This has been achieved through unprecedented teamwork, involving more than 900 investigators from 40 countries, allied to rigorous methodology. Significantly, the PGC has an open-source approach with the main findings freely available for unrestricted use (http://pgc.unc.edu/downloads). Dozens of groups around the world are using PGC data to develop better analytical methods and to perform secondary analyses on a dataset representing more than 400 000 human participants.

In 2014, the PGC reported an exceptionally careful and systematic analysis of all available published and unpublished schizophrenia (SCZ) data in a multi-stage genome-wide association study (GWAS) of up to 36 989 cases and 113 075 controls.[1] This landmark study sheds light on the genetic architecture of SCZ using approaches that had previously been successful for other complex biomedical disorders. Consequently, advances in genomics that are revolutionizing other medical disciplines could and should be applied to SCZ and other psychiatric disorders. The study identified that a third of genetic risk for SCZ could be attributed to common genetic variation of individually small genetic effects. By aggregating these small effects it was possible to generate genetic risk scores (GRS) that not only have predictive power for risk of illness, but that also identified overlap with other psychiatric disorders. The study provided robust association at 108 risk loci and evidence that these associations are enriched at enhancers active in the brain and in immune tissues. Individual genes involved in glutamatergic transmission, calcium channels and immune function were implicated, providing important new avenues for biological research.

This 2014 *Nature* article can be seen as the end of the beginning. The field now faces at least 3 major challenges. First, although common genetic variants make a substantial contribution to risk, most of illness susceptibility is unexplained. Some of this difference is attributable to rare risk variants ranging from large genomic changes (resulting from copy number variation) to single base changes.[2–4] Large samples will be required to generate study power as the contribution of individual mutations to risk in the population will be very small even if there is a relatively large impact on risk for an individual.[5]

Second, even though there is now a list of more than 100 susceptibility loci, interpreting how the implicated genes map to identifiable molecular pathways pertinent to disease etiology is challenging.[6] However, we can learn from other complex disorders where the challenge has not proved insurmountable (eg, Type 2 Diabetes).[7] Recent work by Steve McCarroll's group suggests grounds for optimism despite the additional challenges that are posed by clinically defined brain disorders like SCZ. The extended MHC region has the strongest association with SCZ ($P \sim 10^{-31}$), but this is the most complicated region in the human genome with extremely high gene density and linkage disequilibrium. McCarroll and colleagues convincingly identify copy number variation in the *C4* gene as one source of the MHC association with SCZ. Moreover, *C4* mediated synapse elimination in a mouse model. This finding is important, because it directly connects a GWAS finding to a specific gene with a directional hypothesis, and directly leads to a novel hypothesis about the etiology of SCZ with immediate translational implications.[8]

Finally, in complex disease genetics most common loci are likely to have subtle, regulatory effects with only a small fraction predicted to impact protein function. Consequently, there is a drive to identify rare, highly

penetrant mutations as these may be easier to model in cellular and animal systems and will complement ongoing efforts to understand the function of common risk variants.

The Third Set of PGC Aims

The task for the PGC and the field is to translate genetics discovery into an understanding of SCZ biology. The aims of the 2 prior versions of the PGC were focused on developing and proving a robust set of methodological approaches.[9–11] In the next 5 years, the overarching aim of what will be the third set of PGC (PGC3) is to identify biologically, clinically and therapeutically meaningful information from the growing resource of available genetic information. Achieving this will require an ambitious investigation of both common and rare genetic variation addressing 6 key questions.

What More Can Be Learned From GWAS in SCZ?

Identified common SCZ risk loci that pass a stringent threshold of medical significance explain ~5% of genetic risk variance. We know that at least another 25% of susceptibility is explained by common variants that have not yet met this threshold and that this may include contribution from hundreds or even thousands of small effects. As effects become smaller the discovery return from increasing sample size will diminish. The plan for PGC3 is to analyze 100 000 cases. In 2016, we expect to attain a total of ~65 000 cases. Based on previous average discovery of 4 associations per 1000 cases would expect to double the number of confirmed loci. Unpublished data supports this observation (O'Donovan MC, personal communication).

As larger imputation panels (eg, from the Haplotype Reference Consortium) become available it will be possible to impute and test much less common genetic variants (eg, to a minor allele frequency of 0.1% rather than the current 1%) widening the field of investigation. Larger sample sizes and a greater contribution from non-European samples will be important in providing greater resolution and narrower intervals of association and help to determine which variants are likely to be functionally relevant.

This increased yield of common risk variants will, in turn, inform analyses by the PGC Network/Pathway group to identify functionally or biologically connected gene sets in SCZ but also across psychiatric disorders. The group has developed a statistically robust approach that combines multiple analysis programs (including FORGE, INRICH, ALIGATOR, and MAGMA) that account for technical challenges such as linkage disequilibrium and gene size. They have published promising results from their provisional cross-disorder analyses of the existing data.[12] Significant efforts will be made to expand GWAS investigation through a central pipeline across bipolar disorder (BP), major depressive disorder (MDD), anorexia nervosa (AN), autism, attention deficit hyperactivity disorder (ADHD), substance use disorders, obsessive-compulsive disorder (OCD)/Tourette syndrome, and post-traumatic stress disorder. This will substantially increase study power for cross-disorder pathway analysis and will also facilitate the Brainstorm Initiative discussed below.

What More Can We Learn From GRS?

GRS are a weighted sum of the number of risk alleles in a GWAS and can be used to define a genetic "score" for each individual in an independent "target" sample. Since its original application by the International Schizophrenia Consortium, identifying sizable genetic correlations between SCZ and bipolar disorder, variations have emerged to allow better effect estimation.[13,14] It has been shown that GRS for SCZ predict BP cases with or without psychosis and also treatment response in SCZ patients.[15–17] By following an ongoing study of all 9-year-old twins born in Sweden since 1992, PGC3 will evaluate if GRS for disorders like SCZ can predict developmental trajectories that predate full development of the illness. By extension, interactions between genotype (measured by GRS) and environmental risk factors will also be explored. A second aim will be to predict if each disorder will predict signs, symptoms or comorbid diagnosis in other groups (eg, MDD GRS predict mood symptoms or disorder in SCZ).

"Brainstorm": What Can We Learn About the Relationships Between Brain Disorders?

Co-morbidity between psychiatric disorders is common in clinical practice and epidemiological and clinical correlations have been observed between SCZ and many other disorders. GRS and other methods have allowed cross disorder analyses confirming that observed relationships between SCZ-BP, SCZ-MDD, and BP-MDD reflect shared genetics rather than exogenous environmental factors.[18] Other methods such as multi-phenotype Mendelian randomization (MMR)[19] and LD score regression (LDSR)[20] can also be applied to allow evaluation of genetic relationships between phenotypes. Smaller versions of this aim are published, and a far more comprehensive analysis is nearing completion.[21,22] Provisional analysis suggests substantial genetic overlap between PGC disorders, but also that other correlations (SCZ-smoking) are driven by non-genetic factors.[23] Understanding these overlaps is essential to improving classification across disorders as supported by the NIMH RDoC initiative. Through the Brainstorm Initiative a more ambitious analysis is proposed to combine GWAS meta-analyses of all PGC disorders, neurological

diseases and brain-related traits to identify other genetic relationships.

Can We Evaluate the Role of Rare Copy Number Variants in SCZ?

A number of robustly associated copy number variant (CNV) loci have been identified already through merging of summary data at specific candidate loci across datasets held by PGC PI's (reviewed in Kirov[2]). The PGC has developed a centralized pipeline applying multiple calling algorithms to raw intensity data followed by QC, curation and analysis to allow a more systematic genome-wide analysis across the PGC disorders. A CNV analysis is nearing completion for the SCZ dataset, providing genome-wide significance evidence for a number of loci, but highlighting the power challenges in identifying rare events. Extending the sample to 100 000 cases will certainly increase power. Almost all confirmed risk CNVs for SCZ also increase susceptibility to other neurodevelopmental phenotypes (called pleiotropy).[24] As datasets across other disorders expand, cross-disorder comparisons will facilitate a deeper understanding of the potential phenotypes associated with each risk locus. Gene set analysis shows enrichment of genes from the *N*-methyl-D-aspartate receptor (NMDAR) and neuronal activity-regulated cytoskeleton-associated protein (ARC) postsynaptic signaling complexes at the small number of known SCZ CNV loci and it will be interesting to revisit this analysis if more loci can be identified.[25]

Can We Perform Well-Powered Studies of Rare Sequence Variation?

Rare mutations are a sizable reservoir of genetic variation, often missed by SNP arrays but particularly actionable.[26] Rare variant discovery is challenging and the SCZ data suggest overlap between common and rare variation at the same loci.[3,4,25] Genome sequencing has suggested that current sample sizes are insufficient. The PGC will perform target sequencing of 200 genes, based on SCZ results, in at least 20 000 subjects. This will focus on top hits from the PGC2 GWAS study with an emphasis on gene sets implicated by GWAS (eg, voltage-gated calcium channel signaling and NMDAR).

What Can We Learn From Densely Affected Pedigrees?

De novo mutations can have large effects on SCZ or ASD risk but most risk alleles are inherited. Although it is essentially impossible for these alleles to become common in the population they can segregate over a number of generations within a pedigree. Many densely affect pedigrees have been identified in SCZ and whole genome sequencing (WGS) is potentially powerful in helping to identify these mutations where they occur. Until now very few of these pedigrees have been assessed with WGS. As was the case for GWAS, there is substantial work to be done to develop standardized methodology and statistical analytical frameworks, and data sharing to allow rapid identification and validation of key findings. The PGC3 could provide an important forum for this development. Crucially, the PGC represents the efforts of hundreds of clinicians who collectively have performed many thousands of family history evaluations. By using this remarkable reach it should be possible to systematically identify the pedigrees most likely to yield rare variants of strong effect. Participation will be open to PGC and non-PGC clinicians from around the world (further details are available from author A.C.). The goal is to sequence from at least 100 of these pedigrees with a focus on unusually densely affected pedigrees, or large pedigrees where there is extensive comorbidity within the family.

Conclusions

These 6 research questions are not independent, and it will be the ability to integrate data across these questions that will determine the project's success. Beyond developing robust methodological approaches this has been a major strength of the PGC approach so far. An expanded list of common risk loci, with better understanding of whether these contribute to other brain disorders will be highly informative for Network/Pathway analysis as would the ability to incorporate data from rare variant discovery into these analyses. GRS methods will make it possible to test how these data may relate to the development and trajectory of illness, or to core symptomatology. This can frame the hypotheses tested in model systems. Rare variant analysis is likely to be helpful in defining particularly actionable mutations for analysis in model systems.

Other developments in the field will also be important for PGC3. More than 90% of the human genome represents non-coding DNA sequences.[27] It has long been known that noncoding elements have an important role in regulating gene expression but efforts to catalogue this variation are relatively recent.[28] PGC3 is timely as it coincides with international efforts to develop a better understanding of the genetic regulation of the developing and adult human brain. The PsychENCODE project aims to produce a publicly available multidimensional genomics resource using tissue- and cell type-specific samples from ~1000 well characterized health and disease-affected human post-mortem brains.[29] This will provide important context to efforts to understand the biological relevance of mutations or genetic variants detected through PGC3. Multiple groups around the world are already working on the neurobiology of emerging genomics findings. Success for PGC3 will be measured by the evaluation of the biological significance of common and rare genomic findings and their relevance to the etiology, diagnosis or treatment of SCZ. If successful this will help to guide and focus the

search for those core molecular etiological mechanisms most likely to improve patient care.

Funding

The PGC is supported by National Institute of Mental Health R01 MH094421.

Acknowledgment

P.F.S. consults for Pfizer, Inc.

References

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427.
2. Kirov G. CNVs in neuropsychiatric disorders. *Hum Mol Genet*. 2015;24:R45–49.
3. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*. 2014;506:185–190.
4. Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014;506:179–184.
5. Zuk O, Schaffner SF, Samocha K, et al. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A*. 2014;111:E455–464.
6. McCarthy SE, McCombie WR, Corvin A. Unlocking the treasure trove: from genes to schizophrenia biology. *Schizophr Bull*. 2014;40:492–496.
7. Gaulton KJ, Ferreira T, Lee Y, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet.* 2015;47:1415–1425.
8. Sekar A, Bialas AR, de Rivera H, et al. Schizophrenia risk from complex variation of complement component 4. *Nature*. 2016;530:177–183.
9. Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron*. 2010;68:182–186.
10. Cichon S, Craddock N, Daly M, et al. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*. 2009;166:540–556.
11. Psychiatric GWAS Consortium Steering Committee. A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol Psychiatry*. 2009;14:10–17.
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.
13. Ligthart L, Hottenga JJ, Lewis CM, et al. Genetic risk score analysis indicates migraine with and without comorbid depression are genetically different disorders. *Human Genet*. 2014;133:173–186.
14. Heck A, Fastenrath M, Coynel D, et al. Genetic analysis of association between calcium signaling and hippocampal activation, memory performance in the young and old, and risk for sporadic Alzheimer disease. *JAMA Psychiatry*. 2015;72:1029–1036.
15. Meier SM, Agerbo E. High loading of polygenic risk in cases with chronic schizophrenia [published online ahead of print September 1, 2015] . *Mol Psychiatry*. doi:10.1038/mp.2015.130.
16. Frank J, Lang M, Witt SH, et al. Identification of increased genetic risk scores for schizophrenia in treatment-resistant patients. *Mol Psychiatry*. 2015;20:913.
17. Ruderfer DM, Fanous AH, Ripke S, et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry*. 2014;19:1017–1024.
18. Lee SH, Ripke S, Neale BM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45:984–994.
19. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32:1–22.
20. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47:291–295.
21. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47:1236–1241.
22. Cross-Disorder Group of the Psychiatric Genomics Consortium; Lee SH, Ripke S, Neale BM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45:984–994.
23. Antilla V. The Brainstorm Project. Poster session presented at: American Society of Human Genetics; October 2014; San Diego, CA.
24. Tansey KE, Owen MJ, O'Donovan MC. Schizophrenia genetics: building the foundations of the future. *Schizophr Bull*. 2015;41:15–19.
25. Kirov G, Pocklington AJ, Holmans P, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry*. 2012;17:142–153.
26. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.
27. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
28. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
29. Akbarian S, Liu C, Knowles JA, et al. The PsychENCODE project. *Nat Neurosci*. 2015;18:1707–1712.