Mini-review: Update on the genetics of schizophrenia

Short running title: **Update on the genetics of schizophrenia**

[1]Leda Coelewij, [1,2]David Curtis*

[1]UCL Genetics Institute, University College London.

[2]Centre for Psychiatry, Barts and the London School of Medicine and Dentistry.

*Correspondence: Darwin Building, Gower Street London WC1E 6BT, UK.
d.curtis@ucl.ac.uk

Summary

A number of important findings have recently emerged relevant to identifying genetic risk factors for schizophrenia. Findings using common variants point towards gene sets of interest and also demonstrate an overlap with other psychiatric and non-psychiatric disorders. Imputation of variants of the gene for complement component 4, C4, from GWAS data has shown that the predicted expression of the C4A product is associated with schizophrenia risk. Very rare variants disrupting SETD1A, RBM12 or NRXN1 have a large effect on risk. Other rare, damaging variants are enriched in genes which are loss of function intolerant and/or whose products localise to the synapse. These and particular copy number variants (CNVs) can result in increased risk of schizophrenia but also of other neurodevelopmental disorders. The findings for C4 and for NRXN1 may be especially helpful for elucidating the biological mechanisms which can lead to disease.

Key words: Schizophrenia, genetics, CNV, C4, SETD1A, RBM12, NRXN.

Introduction

Research into the genetics of schizophrenia is advancing rapidly and several important findings have emerged since the comprehensive review by Harrison in 2015 (Harrison 2015). At that time, a large GWAS carried out by the Psychiatric Genomics Consortium (PGC) had just been published which implicated over 100 loci at genome-wide significance, all with small effect sizes (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). Associations were enriched for genes expressed in the brain and in tissues with important roles in immunity. This study demonstrated that the cumulative effect of common variants could not explain the observed high heritability of schizophrenia and the problem of the "missing heritability" remained (Harrison 2015). The present article aims to provide a brief summary of more recent findings in order for the non-specialist to gain an overview of where the field currently stands. We present findings for individual genes, for common and rare variants and for copy number variants. We also present some results regarding the genetic relationship between schizophrenia and other disorders, notably developmental disability.

Findings for individual genes

The first specific gene to be implicated in the aetiology of schizophrenia was DISC1, when it was shown that a balanced translocation disrupting this gene cosegregated with schizophrenia and other psychiatric diagnoses in an extended pedigree from Scotland (Millar et al. 2000). However it has since proved difficult to elucidate the mechanism of this effect or to conclusively link other variants in this gene to increased risk (Wang et al. 2018). Although the 2014 GWAS produced a number of hits in or near genes of interest, it did not implicate any specific variants which clearly had a direct functional effect (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). Since then, evidence has emerged which provides a handful of examples where effects can be ascribed to particular types of variant affecting particular genes, comprising complement component 4 (C4), neurexin 1 (NRXN1), RNA-binding motif 12 (RBM12) and SETD1A.

The most significant GWAS hit occurred in the major histocompatibility (MHC) locus, which had been consistently implicated as associated with schizophrenia in a number of previous studies (Harrison 2015). However, identifying the gene or genes driving this signal had proved problematic, partly because of the complex and long range linkage disequilibrium (LD) relationships which are observed in this region. Using droplet digital PCR (ddPCR), Sekar and colleagues characterised complex variation in C4, the gene coding for complement component 4 (Sekar et al. 2016). Three kinds of variation commonly exist: the total number of copies can vary between 0 and 5; each copy can be long or short (L or S), depending on whether or not it contains a HERV insertion; each copy may be either of two paralogous genes (isotypes) which are denoted C4A and C4B and whose products bind different molecular targets. Four structural forms of C4A/C4B are commonly observed (BS, AL–BS, AL–BL, and AL–AL) and each of these was shown to be associated with differing levels of expression and each could be reliably imputed using the SNPs from the GWAS. This made it possible to use the PGC sample of 28,799 schizophrenia cases and 35,986 controls to show that predicted C4A expression from imputed C4 variants was associated with schizophrenia risk ($p = 3.6 \times 10^{-24}$) (Sekar et al. 2016). In a smaller number of subjects they confirmed that directly measured C4A expression was indeed higher in schizophrenia cases. The authors also demonstrated that C4A is present on neurons and synapses and they postulated that increased expression could lead to increased synaptic pruning, which would produce the smaller number of synapses which have been observed in patients with schizophrenia. They demonstrated that mice deficient in C4 showed changes consistent with reduced synaptic pruning. Although the effect on schizophrenia risk is only moderate, with OR of 1.3 between the structural forms with highest and lowest risk, this important paper is the first to link genetic variants to clear changes in function which impact in a biologically meaningful way.

In a genome-wide analysis of copy number variants (CNVs), eight loci were individually genome-wide significant, including a deletion at 2p16.3 (OR=5.9, $p = 4.9 \times 10^{-9}$) (Marshall et al. 2016). This is of especial interest because the deletion varies in size and is in some cases so small that it only effects exons of a single gene, NRXN1, which codes for neurexin 1. This result is aligned with previous reports that deletions of NRXN1 were associated with schizophrenia and other developmental disorders (Rujescu et al. 2009; Ching et al. 2010). Neurexins mediate synaptic function and mice in which NRXN1 is deleted have impaired social memory (Dachtler et al. 2015). Thus, there now seems to be good evidence that loss

of one copy of NRXN1 can substantially increase the risk of schizophrenia and this suggests that abnormal synaptic functioning may form part of the pathophysiology.

A study of an Icelandic family containing ten subjects with a psychiatric disorder, of whom six had a diagnosis of schizophrenia, revealed that all subjects with psychosis had a single nonsense mutation in RNA-Binding Motif 12 (RBM12) that was carried by all individuals with psychosis (Steinberg et al. 2017). The authors performed an extensive search in an attempt to find other subjects with variants disrupting RBM12 and were able to identify a Finnish case with schizophrenia who had a different truncating variant in this gene, c.2532delT. This subject was found to have four siblings with psychosis who all carried the same variant. Thus, it seems that truncating mutations of RBM12 can substantially increase the risk of schizophrenia. However these are extremely rare and at time of writing very little is known about the biological function of this gene.

Earlier studies had provided some evidence that loss of function (LoF) variants affecting SETD1A were associated with schizophrenia (Takata et al. 2014; Fromer et al. 2014). These were followed up in a whole exome sequenced sample containing 4,264 schizophrenia cases, 9,343 controls and 1,077 trios (cases with schizophrenia and their parents) (Singh et al. 2016). A combined case-control and de novo mutation analysis confirmed the association of SETD1A LoF variants with schizophrenia ($p = 3.3 \times 10^{-9}$). In a follow-up investigation of over 20,000 exomes, 10 LoF additional variants were identified in cases and none in controls. Out of 45,376 exomes in ExAC, only two were found to have SETD1A LoF variants. Examination of an exome-sequenced sample of 3,148 children with diverse, severe developmental disorders revealed that four had LoF variants in SETD1A and a fifth had a *de novo* deletion which included it. Overall, this study provided strong evidence that these very rare variants can have a substantial effect on increasing risk of schizophrenia or developmental disorder. SETD1A codes for a component of a histone methyl transferase and disruption of its function might be expected to have effects on the expression of many other genes. Thus, this finding does not immediately provide clear insights relating genetic variation to relevant biological functionality.

Common variants

A recent GWAS using a Chinese cohort consisting of 12,083 schizophrenia cases and 24,097 controls identified 7 loci associated to schizophrenia, 4 of which were novel, and when a subset of this cohort was analysed in combination with the PGC GWAS a total of 30 novel loci were found with genome wide significance (Li et al. 2017)**.** The authors used factors such as proximity, missense mutations and expression quantitative locus information to highlight 247 genes, of which 85 could be defined as "prioritized candidate genes". However, none can be conclusively declared to be involved in disease causation.

A major emphasis of studies of common variants has been not only to search for SNPs associated with schizophrenia risk but also to investigate relationships with the genetic susceptibility to different outcomes. Methods such as the polygenic risk score (PRS) and LD-regression aim to detect an overlap of multiple common variants associated with different phenotypes even if variants are not individually statistically significant. A recent review of the application of PRS analysis to schizophrenia identified 31 articles examining association with another phenotype (Mistry et al. 2017). There seems to be some overlap of common genetic risk factors for schizophrenia with other psychiatric disorders, notably depression, bipolar

disorder and borderline personality disorder (Cross-Disorder Group of the Psychiatric Genomics Consortium 2013; Lee et al. 2013; Witt et al. 2017). There are also results suggesting that the polygenic contribution to schizophrenia risk is associated with poorer cognitive functioning and a number of non-psychiatric conditions including diabetes, rheumatoid arthritis, Crohn's disease and amyotrophic lateral sclerosis (McIntosh et al. 2013; Hubbard et al. 2016; Stringer et al. 2014; McLaughlin et al. 2017).

Rare variants

A study of a whole exome sequenced Swedish sample of 4,877 cases and 7,455 controls focussed on damaging ultra-rare variants (dURVs), defined as nonsense, frame-shift and splice site variants found in only a single subject and absent from ExAC (Purcell et al. 2014; Lek et al. 2016). Although no individual genes were implicated, dURVs were enriched among cases in a number of gene sets, some overlapping each other, including genes near GWAS hits, genes known to cause intellectual disability, neuron specific genes, genes whose mRNAs are bound by fragile X mental retardation protein (FMRP) and genes encoding interactors with PSD-95, ARC, and NMDA receptors (Genovese et al. 2016). These results conclusively demonstrated that rare variants contribute to schizophrenia pathogenesis. They also point towards some of the biological processes likely to be involved.

Rare variants were also investigated in a UK cohort consisting of 5,585 schizophrenia cases and 8,103 controls. Only uncommon variants with minor allele frequency (MAF) < 1% were considered. No variant or gene was found to reach genome wide significance, and the only gene set to reach an adjusted significance threshold was that consisting of FMRP targets (which tend to be localised to the synapse) (Richards et al. 2016). A study combining the Swedish cohort and UK cohorts, including a total of 10,011 schizophrenia cases and 13,791 controls, demonstrated enrichment of rare variants in two gene sets, targets of FMRP and genes which are loss of function intolerant (Leonenko et al. 2017). One variant in MCPH1, rs61749465, approached but failed to reach exome-wide significance, a result possibly of interest because mutations in this gene are associated with primary autosomal recessive microcephaly.

Copy number variants

CNVs have been strongly implicated in schizophrenia risk, as previously reviewed (Harrison 2015). A large collaborative study, incorporating 43 separate datasets, included 21,094 schizophrenia cases compared to 20,227 controls (Marshall et al. 2016). In this sample, the CNV burden was increased in cases even after removing previously implicated loci (OR = 1.07, $p = 1.7 \times 10^{-6}$). There was a larger enrichment for copy number losses than copy number gains ($OR_{Loss} = 1.4$, $p_{Loss} = 4 \times 10^{-16}$; $OR_{Gain} = 1.1$, $p_{Gain} = 2 \times 10^{-7}$). This study implicated 8 CNV loci which were individually genome wide significant, including the NRXN1 deletion referred to above.

36 gene sets were tested for CNV burden, of which 15 were significantly enriched for rare copy number losses, where the gene set “GO synaptic” was ranked the most statistically significant ($p = 2.8 \times 10^{-11}$) and the gene set “ARC complex” was found to have the largest effect size (OR > 1.8). NMDA receptor complex genes, brain expressed genes and

prenatally expressed brain genes were significantly enriched for duplications (Marshall et al. 2016).

A study of CNVs identified associated with schizophrenia demonstrated that although they had a substantial effect on risk of schizophrenia they were in fact more likely to result in other phenotypes, consisting of developmental disorder (DD), autism spectrum disorder (ASD) and congenital malformations (CM) (Kirov et al. 2014). Thus, for the NRXN1 and the 22q11.21 deletions the estimated penetrance values for schizophrenia were 0.064 and 0.12 respectively, compared to a background population risk of about 0.01. However these values for the DD/ASD/CM phenotypes were 0.26 and 0.88. In general it seems that these CNVs produce a substantial risk of some kind of developmental abnormality but that schizophrenia represents only one possible outcome. Although their effect sizes are large, overall they remain a rare cause of schizophrenia and together are found in only 2-3% of cases.

Overlap with developmental disorders

Studies of dURVs, LoF variants affecting SETD1A and CNVs all demonstrate that disruption of the same genes can lead to either schizophrenia or a different neurodevelopmental disorder (Genovese et al. 2016; Singh et al. 2016; Kirov et al. 2014). To investigate this issue further, a meta-analysis was carried out using SNV calls from the whole-exome sequences of 4,133 schizophrenia cases and 9,274 matched controls, *de novo* mutations identified in 1,077 schizophrenia parent–proband trios and CNV calls from genotyping array data of 6,882 cases and 11,255 controls (Singh et al. 2017). This study showed that in schizophrenia cases there was an enrichment for rare, damaging variants in LoF-intolerant genes but that the effect size was smaller than for ASD and severe developmental disorders. Similar genes were involved, in that genes implicated in ASD and severe developmental disorders were significantly enriched for rare, damaging variants in schizophrenia cases ($p = 9.5 \times 10^{-6}$ and 2.3×10^{-6} respectively). The authors were able to acquire the cognitive phenotype for 2,971 schizophrenia cases, of whom 279 were also diagnosed with intellectual disability (ID). In this group of 279 SCZ-ID cases, the burden of rare, damaging variants in LoF-intolerant genes was observed to be significantly higher than in the other schizophrenia cases ($OR = 1.3$, $p = 2.6 \times 10^{-4}$) or in controls ($OR = 1.61$, $p < 5 \times 10^{-7}$). Likewise, these SCZ-ID cases also had a significant enrichment of rare LoF variants in developmental disorder–associated genes as compared to the other schizophrenia cases ($OR=2.4$, $P = 9 \times 10^{-4}$) or to controls ($OR = 3.4$, $P = 9.5 \times 10^{-6}$). Schizophrenia cases without ID did not have this enrichment in neurodevelopmental disorder–associated genes but did still have some enrichment of rare, damaging variants in LoF intolerant genes relative to controls ($OR=1.3$, $P = 2 \times 10^{-6}$). Overall, these analyses demonstrated that the burden of rare, damaging variants is highest in subjects with schizophrenia who also have lower cognitive functioning and that these subjects account at least in part for the observed sharing of rare genetic risk factors between schizophrenia and other neurodevelopmental disorders.

Conclusion

Studies of common variants demonstrate that they do modify risk of schizophrenia but have not been able to implicate individual genes. No variant has an individually large effect, as is to be expected given the selection pressures which result from the impaired reproductive fitness associated with schizophrenia (Power et al. 2013). By contrast, extremely rare CNVs and LoF variants, sometimes affecting individual genes, can result in a high risk of

schizophrenia, as is the case for NRXN1, SETD1A and RBM12. The findings for NRXN1 and C4 point to specific biological processes but otherwise it is difficult to say more than that the evidence seems to point towards genes involved in neurodevelopment and genes whose products are located in synapses. As further sequence data becomes available for analysis it is reasonable to expect that additional rare variants in additional genes may be recognised and that these may lead on to a better understanding of pathogenesis.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Ching, M.S.L. et al., 2010. Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 153B(4), pp.937–47.

Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, 381(9875), pp.1371–1379.

Dachtler, J. et al., 2015. Heterozygous deletion of α-neurexin I or α-neurexin II results in behaviors relevant to autism and schizophrenia. *Behavioral neuroscience*, 129(6), pp.765–76.

Fromer, M. et al., 2014. De novo mutations in schizophrenia implicate synaptic networks. *Nature*, 506(7487), pp.179–184.

Genovese, G. et al., 2016. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nature Neuroscience*, 19(11), pp.1433–1441.

Harrison, P.J., 2015. Recent genetic findings in schizophrenia and their therapeutic relevance. *Journal of Psychopharmacology*, 29(2), pp.85–96.

Hubbard, L. et al., 2016. Evidence of Common Genetic Overlap Between Schizophrenia and Cognition. *Schizophrenia Bulletin*, 42(3), pp.832–842.

Kirov, G. et al., 2014. The Penetrance of Copy Number Variations for Schizophrenia and Developmental Delay. *Biological Psychiatry*, 75(5), pp.378–385.

Lee, S.H. et al., 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*, 45(9), pp.984–994.

Lek, M. et al., 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), pp.285–291.

Leonenko, G. et al., 2017. Mutation intolerant genes and targets of FMRP are enriched for nonsynonymous alleles in schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*.

Li, Z. et al., 2017. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nature Genetics*, 49(11).

Marshall, C.R. et al., 2016. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics*, 49(1), pp.27–35.

McIntosh, A.M. et al., 2013. Polygenic Risk for Schizophrenia Is Associated with Cognitive Change Between Childhood and Old Age. *Biological Psychiatry*, 73(10), pp.938–943.

McLaughlin, R.L. et al., 2017. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nature Communications*, 8, p.14774.

Millar, J.K. et al., 2000. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Human molecular genetics*, 9(9), pp.1415–23.

Mistry, S. et al., 2017. The use of polygenic risk scores to identify phenotypes associated with genetic risk of schizophrenia: Systematic review. *Schizophrenia Research*.

Power, R.A. et al., 2013. Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry*, 70(1), pp.22–30.

Purcell, S.M. et al., 2014. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, 506(7487), pp.185–90.

Richards, A.L. et al., 2016. Exome arrays capture polygenic rare variant contributions to schizophrenia. *Human molecular genetics*, 25(5), pp.1001–7.

Rujescu, D. et al., 2009. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*, 18(5), pp.988–996.

Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511, pp.421–427.

Sekar, A. et al., 2016. Schizophrenia risk from complex variation of complement component 4. *Nature*.

Singh, T. et al., 2016. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nature neuroscience*, 19(4), pp.571–577.

Singh, T. et al., 2017. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nature Genetics*.

Steinberg, S. et al., 2017. Truncating mutations in RBM12 are associated with psychosis. *Nature Genetics*.

Stringer, S. et al., 2014. Genetic liability for schizophrenia predicts risk of immune disorders. *Schizophrenia Research*, 159(2–3), pp.347–352.

Takata, A. et al., 2014. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron*, 82(4), pp.773–780.

Wang, H.-Y. et al., 2018. Gene polymorphisms of DISC1 is associated with schizophrenia: Evidence from a meta-analysis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 81, pp.64–73.

Witt, S. et al., 2017. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. *Translational Psychiatry*, 7115.