

## **Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection**

Antonio F. Pardiñas<sup>a</sup>, Peter Holmans<sup>a</sup>, Andrew J. Pocklington<sup>a</sup>, Valentina Escott-Price<sup>a</sup>, Stephan Ripke<sup>b,c</sup>, Noa Carrera<sup>a</sup>, Sophie E. Legge<sup>a</sup>, Sophie Bishop<sup>a</sup>, Darren Cameron<sup>a</sup>, Marian L. Hamshere<sup>a</sup>, Jun Han<sup>a</sup>, Leon Hubbard<sup>a</sup>, Amy Lynham<sup>a</sup>, Kiran Mantripragada<sup>a</sup>, Elliott Rees<sup>a</sup>, James H. MacCabe<sup>d</sup>, Steven A. McCarroll<sup>e</sup>, Bernhard T. Baune<sup>f</sup>, Gerome Breen<sup>g,h</sup>, Enda M. Byrne<sup>i</sup>, Udo Dannlowski<sup>j,k</sup>, Thalia C. Eley<sup>g</sup>, Caroline Hayward<sup>l</sup>, Nicholas G. Martin<sup>m,n</sup>, Andrew M. McIntosh<sup>o,p</sup>, Robert Plomin<sup>g</sup>, David J. Porteous<sup>l</sup>, Naomi R. Wray<sup>i</sup>, GERAD1 Consortium<sup>q</sup>, David A. Collier<sup>g,r</sup>, Dan Rujescu<sup>s,t</sup>, George Kirov<sup>a</sup>, Michael J. Owen<sup>a\*</sup>, Michael C. O'Donovan<sup>a\*</sup>, James T. R. Walters<sup>a\*</sup>

<sup>a</sup> MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK

<sup>b</sup> Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA

<sup>c</sup> Department of Psychiatry and Psychotherapy, Charité, Campus Mitte, 10117 Berlin, Germany

<sup>d</sup> Department of Psychosis Studies, Institute of Psychiatry Psychology and Neuroscience, King's College London

<sup>e</sup> Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

<sup>f</sup> Discipline of Psychiatry, University of Adelaide, Australia

<sup>g</sup> Medical Research Council, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

<sup>h</sup> NIHR Biomedical Research Centre for Mental Health, Maudsley Hospital and Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

<sup>i</sup> Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia

<sup>j</sup> Department of Psychiatry and Psychotherapy University of Muenster, Muenster, Germany

<sup>k</sup> Department of Psychiatry, University of Marburg, Marburg, Germany

<sup>l</sup> Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

<sup>m</sup> School of Psychology, University of Queensland

<sup>n</sup> QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

<sup>o</sup> Division of Psychiatry, University of Edinburgh, Edinburgh, UK

<sup>p</sup> Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

<sup>q</sup> Control data used in the preparation of this article were obtained from the Genetic and Environmental Risk for Alzheimer's disease (GERAD1) Consortium. As such, the investigators within the GERAD1 consortia contributed to the design and implementation of GERAD1 and/or provided control data but did not participate in analysis or writing of this report.

<sup>r</sup> Discovery Neuroscience Research, Eli Lilly and Company Ltd, Lilly Research Laboratories, Erl Wood Manor, Surrey, UK

<sup>s</sup> Department of Psychiatry, University of Halle, Halle, Germany

<sup>t</sup> Department of Psychiatry, University of Munich, Munich, Germany

\* Correspondence and material requests should be addressed to M.J.O. ([owenmj@cardiff.ac.uk](mailto:owenmj@cardiff.ac.uk)), M.C.O'D. ([odonovanmc@cardiff.ac.uk](mailto:odonovanmc@cardiff.ac.uk)) or J.T.R.W. ([waltersjt@cardiff.ac.uk](mailto:waltersjt@cardiff.ac.uk))

Schizophrenia is a debilitating psychiatric condition often associated with poor quality of life and decreased life expectancy. Lack of progress in improving treatment outcomes has been attributed to limited knowledge of the underlying biology, although large-scale genomic studies have begun to provide such insight. We report the largest single cohort genome-wide association study of schizophrenia (11,260 cases and 24,542 controls) and through meta-analysis with existing data we identify 50 novel GWAS loci. Using gene-wide association statistics we implicate an additional set of 22 novel associations that map onto a single gene. We show for the first time that the common variant association signal is highly enriched among genes that are intolerant to loss of function mutations and that variants in these genes persist in the population despite the low fecundity associated with the disorder through the process of background selection. Associations point to novel areas of biology (e.g. metabotropic GABA-B signalling and acetyl cholinesterase), reinforce those implicated in earlier GWAS studies (e.g. calcium channel function), converge with earlier rare variants studies (e.g. *NRXN1*, GABAergic signalling), identify novel overlaps with autism (e.g. *RBFOX1*, *FOXP1*, *FOXG1*), and support early controversial candidate gene hypotheses (e.g. *ERBB4* implicating neuregulin signalling). We also demonstrate the involvement of six independent central nervous system functional gene sets in schizophrenia pathophysiology. These findings provide novel insights into the biology and genetic architecture of schizophrenia, highlight the importance of mutation intolerant genes and suggest a mechanism by which common risk variants are maintained in the population.

Schizophrenia is characterised by psychosis and negative symptoms such as social and emotional withdrawal. While onset of psychosis typically does not occur until late adolescence or early adult life, there is strong evidence from clinical and epidemiological studies that schizophrenia reflects a disturbance of neurodevelopment<sup>1</sup>. It confers substantial mortality and morbidity, with a mean reduction in life expectancy of 15-30 years<sup>2,3</sup>. Although recovery is possible, most patients have poor social and functional outcomes<sup>4</sup>. No substantial improvements in outcomes have emerged since the advent of antipsychotic medication in the mid-20th century, a fact that has been attributed to a lack of knowledge of pathophysiology<sup>1</sup>.

Schizophrenia is both highly heritable and polygenic, with risk ascribed to variants spanning the full spectrum of population frequencies<sup>5-7</sup>. The relative contributions of alleles of various frequencies is not fully resolved, but recent studies estimate that common alleles, captured by genome-wide association study (GWAS) arrays, capture between a third and a half of the genetic variance in liability<sup>8</sup>. There has been a long-standing debate, from an evolutionary standpoint, as to how common risk alleles might be maintained in the population, particularly given the early mortality and decreased fecundity associated with schizophrenia<sup>9</sup>. Various hypotheses have been proposed including compensatory advantage (balancing selection), whereby schizophrenia alleles confer reproductive advantages in particular contexts<sup>10,11</sup>; hitchhiking, whereby risk alleles are maintained by their linkage to positively selected alleles<sup>12</sup>; or contrasting theories that attribute these effects to rare variants and gene-environment interaction<sup>13</sup>. Addressing these competing hypotheses is now tractable given advances from recent studies of common genetic variation in schizophrenia.

The largest published schizophrenia GWAS, that from the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC), identified 108 genome-wide significant loci and unequivocally demonstrated the value of increasing sample sizes for discovery in schizophrenia GWAS<sup>5</sup>. Here, we report by far the largest ancestrally and phenotypically

homogeneous GWAS study of schizophrenia to date (N=11 260 cases, 24,542 controls). We combine these new data with the previous published GWAS to identify novel facets of genetic architecture and biology, and demonstrate that the evolutionary process of background selection can contribute to the maintenance of risk alleles in the population.

## RESULTS

### *GWAS and Meta-analysis*

We obtained genome-wide genotype information for the largest single-country study of schizophrenia (CLOZUK) based on 15,000 cases from the UK, 96% of whom were taking clozapine, an antipsychotic for treatment-resistant schizophrenia. Control datasets with UK ancestry were obtained from public repositories or through collaboration. To maximise homogeneity beyond that conferred by our ascertainment strategy, we set stringent thresholds for inclusion based on genotype-derived ancestry, resulting in a final CLOZUK GWAS of 11,260 cases and 24,542 controls (5,220 cases and 18,823 controls not in previous schizophrenia GWAS - **Methods; Figure S1; Figure S2**).

The CLOZUK GWAS analysis ( $\lambda_{GC}=1.281$ ,  $\lambda_{1000}=1.018$ ) yielded 18 genome-wide significant (GWS= $p<5\times 10^{-8}$ ) independent loci, four of which are novel (**Extended Data Table 1; Figure S3; Figure S4**). The CLOZUK SNP-based heritability ( $h^2_{SNP}$ ) on the liability scale was 0.29 (assuming 1% prevalence; see **Extended Data Table 2** for a range of plausible prevalences) using linkage-disequilibrium score regression (LDSR)<sup>14</sup>. We demonstrate common genetic architecture between the CLOZUK and the independent PGC sample (excluding all CLOZUK samples) by showing high genetic correlation (95%), sign test enrichment (99% of PGC GWS SNPs show the same direction of effect in CLOZUK:

$p=2.04 \times 10^{-21}$ ) and strong polygenic risk score overlap ( $p < 1 \times 10^{-300}$ ) – see **Methods** and **Extended Data Table 3** and **4**.

Having demonstrated high genetic correlation between datasets, we performed meta-analysis of the CLOZUK and the independent PGC dataset, excluding related and overlapping samples (total 40,675 cases and 64,643 controls,  $\lambda_{GC}=1.586$ ,  $\lambda_{1000}=1.012$ ; **Figure S5**). We identified 177 independently associated SNPs (**Extended Data Table 5**) that map to 143 independent loci (**Figure 1**, **Extended Data Table 6**, **Methods**). Overall, the association signal at the majority of previously identified loci strengthened in the meta-analysis (**Figure S6**), although 14 of the PGC loci were no longer genome-wide significant (**Extended Data Table 7**). Of the 52 loci not identified by the PGC, two have been reported as genome-wide significant in other studies: *ZEB2*<sup>15</sup> and a locus on chromosome 8 (38.0-38.3 MB)<sup>16</sup>. Thus we identify 50 novel genetic loci for schizophrenia (**Extended Data Tables 6 and 8**).

### ***Gene-based analysis***

To exploit the additional information provided by multiple sub-genome-wide significant SNP associations within single genes, we undertook gene-wide analyses using MAGMA<sup>17</sup> (**Methods**). In the full meta-analytic dataset, excluding the extended Major Histocompatibility region (xMHC) due to its complex LD structure, we identified 535 genome-wide significant genes ( $p < 2.5 \times 10^{-6}$ ), 240 of which did not co-localize with a genome-wide significant SNP locus from either the present meta-analysis or the PGC study<sup>5</sup> (**Extended Data Table 9**). We then used the full genome-wide gene-based results to (i) investigate genetic architecture (ii) highlight single gene loci and (iii) conduct hypothesis and data-driven systems genomic analyses in schizophrenia.

### ***Mutation intolerant genes***

Recent studies have shown that mutation intolerant genes capture much of the rare variant architecture of neurodevelopmental disorders such as autism, intellectual disability and developmental delay<sup>18-20</sup>. Work in parallel to the present study has shown that this extends to rare variants in schizophrenia (Singh et al, submitted). To investigate whether this holds for common variation, we analysed the set of loss-of-function (LoF) intolerant genes as defined by the Exome Aggregation Consortium (ExAC)<sup>21</sup>, using their preferred gene-level constraint metric,  $pLI \geq 0.9$  (probability of being LoF intolerant). Using gene set analysis in MAGMA, we found LoF intolerant genes (N=3230) were highly significantly enriched for schizophrenia common variant associations in comparison with all other genes ( $p=4.1 \times 10^{-16}$ ).

To quantify the contribution of SNPs within LoF intolerant genes to schizophrenia SNP-based heritability ( $h^2_{SNP}$ ) we conducted heritability analyses using partitioned LDSR<sup>22</sup> (**Extended Data Table 10**). Overall, genic SNPs account for 64% of  $h^2_{SNP}$ , a 1.23-fold enrichment (compared to that expected from SNP content) whereas non-genic SNPs were depleted for SNP heritability content ( $h^2_{SNP}=33\%$ ; enrichment=0.69). Consistent with the analysis using MAGMA,  $h^2_{SNP}$  was highly significantly enriched in LoF intolerant genes (2.01-fold;  $p=2.78 \times 10^{-24}$ ). Common variation in LoF intolerant genes explained 30% of all  $h^2_{SNP}$  (equating to 47% of all genic  $h^2_{SNP}$ ) despite containing only 15% of all SNP content (29% of the total genic SNP content). In contrast, genes classed as non-LoF intolerant ( $pLI < 0.9$ ) were significantly depleted for  $h^2_{SNP}$  relative to their SNP content (0.90-fold;  $p=5.86 \times 10^{-3}$ ), although in absolute terms, SNPs in these genes accounted for 34% of  $h^2_{SNP}$ . A finer scale analysis of the relationship between LoF intolerance scores and enrichment for association showed that enrichment is restricted to genes with a pLI score above 0.9 (**Figure S7**).

### ***Common risk alleles maintained by background selection***

Our novel finding that LoF intolerant genes, which by definition are under strong selective pressure, are enriched for common schizophrenia risk variants raises the question of how such alleles are maintained in the population. While the contribution of ultra rare variation in mutation intolerant genes to disorders associated with low fecundity can be accounted for by *de novo* mutation<sup>23,24</sup>, this cannot explain the persistence of common alleles. To address this question, we used partitioned LDSR to test the relationship between schizophrenia associated alleles and SNP-based signatures of natural selection. These included measures of positive selection, background selection, and Neanderthal introgression. We examined the heritability of SNPs after thresholding them at extreme values for these metrics (top 2%, 1% and 0.5%), while adjusting for baseline annotation sets including LoF intolerant genes (**Methods**).

We observed strong evidence for schizophrenia  $h^2_{\text{SNP}}$  enrichment in SNPs under background selection (BGS), which was consistent across all the thresholds we examined (**Table 1**). We also found a significant and consistent depletion of  $h^2_{\text{SNP}}$  in SNPs subject to positive selection as indexed by the CLR statistic. These two results are mutually consistent, as the calculation of the CLR statistic explicitly controls for B-statistic values<sup>25</sup>. This suggests that SNPs under positive selection, but under weak or no BGS, are depleted for association with schizophrenia. No significant relationship between  $h^2_{\text{SNP}}$  and other positive selection or Neanderthal introgression measures was found after correction for multiple testing (**Table 1**).

To explicitly test the association of SNPs under background selection in LoF intolerant genes, we binned the B-statistic into four categories of increasing score, assigned SNPs in these bins to a LoF intolerant set, an “all other” genes set or a non-genic set (**Figure S8**). The lower boundary of the top bin (B-statistic >0.75) corresponds approximately to the top 2% BGS threshold in **Table 1** and is equivalent to a reduction in effective population size estimated at each SNP of 75% or more<sup>26</sup>. We found significant heritability enrichment across all B-



statistic intervals in LoF intolerant genes which increased progressively with higher B-statistic scores. We also found enrichment for SNPs under BGS pressure in genes that are not LoF intolerant, although this was restricted to the highest B-statistic bin. There was no notable enrichment in any B-statistic intervals for non-genic regions.

### ***Notable single gene associations***

In considering the novel GWAS loci and gene-based results from the meta-analysis we highlight the subset of the findings most likely pointing to single genes, given the difficulties in accurately locating causal genes for complex disorders in a GWAS framework<sup>27</sup>. These single gene associations were defined as GWAS SNP loci that span only a single gene +/- 20kb (N=22, 44% of novel loci, **Extended Data Table 6**) or MAGMA associated genes that were at least 100kb from any other genome-wide significant gene (N=76, **Extended Data Table 9**). We acknowledge this does not imply that these are inevitably the pathogenic genes but their likelihood of being so is higher than that for genes in multigenic loci.

Details of the findings we consider most noteworthy are provided in **Table 2**. These results provide novel potential insights into schizophrenia pathogenesis as well as supporting existing hypotheses. To our knowledge we provide the first genome-wide significant findings supporting GABA involvement in schizophrenia pathogenesis through both SNP-based GWAS (*GABBR2*) and gene-based analysis (*SLC6A11*). Previous association studies have implicated calcium and cholinergic signalling involvement in schizophrenia<sup>5,8</sup>; we extend these findings by demonstrating association with an L-type calcium channel gene (*CACNA1D*) and with the acetyl cholinesterase gene (*ACHE*), a novel potential therapeutic target in schizophrenia which has also recently been implicated in autism through de novo mutation analysis<sup>28</sup>. We provide the first common variant support for *NRXN1*, previously

implicated by single-gene copy number variation (CNV) in schizophrenia, autism and intellectual disability<sup>28,29</sup> and we also highlight the involvement of other genes that increase risk across neurodevelopmental disorders including *RBFOX1*. Finally we find evidence for association for two genes whose functions are intimately linked to two of the most prominent genes from schizophrenia candidate gene era, *PDE4B*, a *DISC1* interactor<sup>30</sup>, and *ERBB4* a binding partner of neuregulin1 (*NRG1*)<sup>31</sup>.

### ***Systems genomics***

We undertook gene set analysis in MAGMA to gain insights into the biological systems underpinning schizophrenia. To maximise power we restricted our primary analysis to 134 gene sets we have previously postulated to be of likely relevance to the disorder on the basis of their involvement in the central nervous system, and which we subsequently showed effectively capture the excess CNV burden in schizophrenia<sup>32</sup> (**Extended Data Table 11**). The set span subcellular neuronal function; neuronal cell physiology; cellular, brain region and fiber tract morphology; learning; behavior; and brain development,<sup>32</sup>. In a GWAS context, we now show that collectively, this set captures a disproportionately high proportion of  $h^2_{\text{SNP}}$  (30% of total heritability; enrichment =1.63; 46% of genic heritability; **Extended Data Table 10**).

Of the 134 sets, 54 were nominally significant of which 12 survived multiple-testing correction (family-wise error rate (FWER) p-value < 0.05, **Extended Data Table 11**). Stepwise conditional analysis, adjusting sequentially for the more strongly associated gene sets, resulted in six gene sets that were independently associated with schizophrenia (**Table 3 and Data Supplement**). These gene sets extend from low-level molecular and sub-cellular processes to broad behavioural phenotypes. The most strongly associated gene set is

constituted by the targets of the Fragile X Mental Retardation Protein (FMRP)<sup>33</sup>. FMRP is a neuronal RNA-binding protein that interacts with polyribosomal mRNAs (the 842 target transcripts of this gene set<sup>33</sup>) and is thought to act by inhibiting translation of target mRNAs, including many transcripts of pre- and post-synaptic proteins. The FMRP target set has been shown to be enriched for rare mutational burden in *de novo* exome sequencing studies of autism<sup>34</sup> and intellectual disability<sup>35</sup>. In schizophrenia studies, it has also been shown to be nominally significantly enriched for association signal in sequencing studies<sup>8,35</sup> and in GWAS<sup>5,8</sup> but only inconsistently in studies of copy number variation<sup>32,36</sup>. However, in none of these studies was enrichment at levels of significance that would be equivalent to genome wide significance (i.e. survive correction for a comprehensive testing of gene-sets in public databases). Thus we provide the most robust evidence to date for the involvement of this gene set in schizophrenia.

We highlight another five gene sets that are independently associated with schizophrenia. Three of these derive from the Mouse Genome Informatics database<sup>37</sup> and relate to behavioural and neurophysiological correlates of learning; Abnormal Behaviour (MP:0004924), Abnormal Nervous System Electrophysiology (MP:0002272) and Abnormal Long Term Potentiation (MP:0002207). We note that two of these gene sets (MP:0004924 and MP:0002207) were among the five most enriched of 134 gene sets tested in a recent schizophrenia CNV analysis<sup>32</sup>. The remaining two independently associated genes sets were voltage-gated calcium channel complexes<sup>38</sup> and the 5-HT<sub>2C</sub> receptor complex<sup>39</sup>. The calcium channel finding confirms extensive evidence from common and rare variant studies implicating calcium channel genes in schizophrenia<sup>5,8</sup>, including a novel GWAS locus in *CACNA1D* identified in our meta-analysis. Whilst there is less convergent evidence in support of the involvement of the 5-HT<sub>2C</sub> receptor complex in schizophrenia, the fact that we identify independent association for this gene set implicates these genes in schizophrenia

pathophysiology and potentially rejuvenates a previous avenue of 5-HT<sub>2C</sub> ligand therapeutic endeavour in schizophrenia research<sup>40</sup>. However we interpret this result with caution given the small size of this gene set and the fact that a number of its genes encode synaptic proteins that are structurally related to other receptor complexes<sup>39</sup>, not only 5-HT<sub>2C</sub>.

### ***Systems genomics and mutation intolerant genes***

Together, LoF intolerant genes and the conditionally independent (“significant”) CNS-related gene sets together account for 39% of schizophrenia SNP-based heritability, equating to 61% of genic heritability (**Figure 2A; Extended Data Table 10**). This is likely to be an underestimation of the true effect of these gene sets since distal non-genic regulatory elements (not included in this analysis) will add to the heritability explained by these genes. In examining the relationship between the two gene sets (**Figure 2A**), genes belonging to both categories were the most highly enriched (2.6-fold;  $p=7.90 \times 10^{-15}$ ), although LoF intolerant genes that were not annotated to our significant CNS gene sets still displayed robust enrichment for SNP-based heritability (1.74-fold;  $p=9.77 \times 10^{-10}$ ), while genes that were in the significant CNS gene sets but had  $pLI < 0.9$  showed more modest enrichment (1.39-fold;  $p=6.05 \times 10^{-4}$ ). Notably genes outside these categories showed a depletion in heritability relative to their SNP content (enrichment=0.79,  $p=1.82 \times 10^{-7}$ ).

This general pattern remained when we focussed on the six significant CNS gene sets individually, in that the enrichment in these gene sets derives primarily from their intersection with LoF intolerant genes (**Figure 2B**). Indeed, only the targets of FMRP showed significant enrichment for SNPs in genes that are not LoF intolerant (2.06-fold;  $p=4.23 \times 10^{-5}$ ).

### ***Data-driven gene set analysis***

To set the systems genomics results in context, we undertook a purely data-driven analysis of a larger comprehensive annotation of gene sets from multiple public databases, totalling 6,677 gene sets (**Methods, Extended Data Table 12**). The LoF intolerant gene set was the most strongly enriched followed by the two strongest associated functional gene sets we had specified in our CNS gene set analysis (FMRP targets and MGI Abnormal Behaviour genes). The other three gene sets that survive FWER correction and conditional analyses are calcium ion import (GO:0070509), membrane depolarisation during action potential (GO:0086010) and synaptic transmission (GO:0007268). These sets show a marked functional and genic overlap with the independently associated sets from our primary CNS systems genomic analysis. Indeed if we repeat the data-driven comprehensive gene set analysis whilst adjusting for the six independently associated CNS gene sets, then the only surviving enrichment term is the LoF intolerant genes. These results are consistent with those from CNV analysis<sup>32</sup> in that they do not support annotations other than those related to CNS function, and demonstrate that hypothesis based analysis to maximise power does not substantially impact on the overall pattern of results.

## **DISCUSSION**

In the largest genetic study of schizophrenia to date, we identify 50 novel genomic loci and a further 22 gene-wide associations that map onto to a single gene. We also explore the genomic architecture of, and the evolutionary pressures on, common variant associations to the disorder, and conduct extensive gene-based and systems genomic analyses. As discussed above and in table 1, novel findings point to GABAergic and cholinergic transmission as relevant to the aetiopathogenesis of schizophrenia. We widen the array of calcium channel

receptors implicated in the disorder, and identify associations that support historical candidate gene networks that have been the subject of intensive functional investigation, but which have remained controversial in the absence of support from large scale genome-wide studies (*PDE4B*, *ERBB4*). We demonstrate congruence between common variant implicated genes and multiple rare variant associations to neurodevelopmental disorders, including *NRXN1*, *RBFOX1*, and *FOXG1* indicating that studies of the impact of rare mutations in these genes are likely to be relevant to schizophrenia pathogenesis more widely. Systems genomic analysis highlights six gene sets that are independently associated with schizophrenia, and point to molecular, physiological and behavioural pathways involved in schizophrenia pathogenesis. We also provide the first support at a genome-wide significant equivalent level implicating targets of FMRP in the disorder.

Our study provides the first evidence linking common variation in loss-of-function intolerant genes to risk of developing schizophrenia and demonstrates that these genes account for a substantial proportion (30%) of schizophrenia SNP-based heritability. Our findings underline the importance and value of well-curated catalogues of human genetic diversity and indicate that such resources have relevance for common variation and common disease. The Exome Aggregation Consortium<sup>21</sup> reported that pLI score is positively correlated with gene expression across tissues, and this is also true for brain expression<sup>41</sup>. Furthermore high pLI genes have more protein-protein interaction partners, and pathways with the highest median pLI scores are ‘hub’ processes such as spliceosome and proteasome<sup>21</sup>. This evidence points to LoF intolerant genes being involved in core biological processes and presumably explains why they are under strong selective pressure. Given mutation intolerance and high selection pressure, our finding that common schizophrenia risk variants are enriched in loss-of-function intolerant genes appears paradoxical. However, we also present novel evidence that the major

mechanism responsible for the persistence of common schizophrenia variation, both in LoF intolerant genes and across the genome, is background selection (BGS).

Selection against deleterious variants (purifying selection) is the main driver of genomic conservation over evolutionary timescales<sup>42</sup>. Regions of the genome under strong purifying selection constitute around 8% of the human genome<sup>43</sup>, and primarily include sequences which are important functionally and associated with fitness and survival<sup>44</sup>. BGS is a consequence of purifying selection occurring in regions of low recombination<sup>45,46</sup>. In such regions, selection against strongly deleterious variants causes whole haplotypes to be removed from the gene pool, which reduces genetic diversity at the locus in a manner equivalent to a reduction in effective population size<sup>47</sup>. This has the effect of impairing the efficiency of the selective process, allowing alleles with neutral or small deleterious effects that are not on haplotypes carrying strongly deleterious variants to rise in frequency by drift<sup>45</sup>. Our finding that schizophrenia-associated common SNPs are associated with high levels of BGS suggests a mechanism by which such variants, even if they are mildly deleterious, can be maintained in the population. The pruning of haplotypes carrying strongly deleterious mutations, which is expected to be particularly intense in LoF intolerant genes<sup>21</sup>, would allow these variants to persist in the population for as long as they reside on a haplotype free of these mutations. High LD additionally inhibits their amalgamation or replacement by recombination<sup>45,48</sup>. Such mechanisms of selection at linked sites have been shown to be compatible with the current knowledge about the genomic architecture of complex human traits<sup>49</sup> and to influence phenotypes in model organisms, including gene expression<sup>50</sup>.

We did not find enrichment for any measure of positive selection or Neanderthal introgression. A recent study explained a negative correlation between schizophrenia associations and metrics indicative of a Neanderthal selective sweep as evidence for positive

selection or polygenic adaptation in schizophrenia<sup>12</sup>. We do not replicate this correlation in our model, which addresses the contribution of BGS, and hence our results are not consistent with large contributions of positive selection to the genetic architecture of schizophrenia (**Table 1**). Indeed positive selection is not widespread in humans, as reported by other studies that explicitly considered or accounted for BGS<sup>25,51</sup>. Polygenic adaptation, the co-occurrence of many subtle allele frequency shifts at loci influencing complex traits<sup>52</sup>, though an intriguing possibility, has not been implicated in psychiatric phenotypes, including schizophrenia, in recent analyses<sup>53,54</sup>. In contrast, BGS has been proposed as a mechanism driving Human-Neanderthal incompatibilities, as regions with higher estimated B-statistics have lower estimated Neanderthal introgression<sup>55</sup>. We therefore conclude that the bulk of the BGS signal we obtain is unlikely to be influenced by positive selection<sup>26</sup> and that previous results linking positive selection and genetic variation in schizophrenia are likely to have been confounded by failure to control for BGS<sup>46</sup>.

In summary, we demonstrate the capacity and continued importance of improving the power of genome-wide association studies for answering fundamental questions about the nature of schizophrenia. We identify large numbers of novel associations that both suggest novel, and support established, hypotheses about the biology of schizophrenia. We show convergence of risk variants on mutation intolerant genes, and identify background selection as the main mechanism by which common schizophrenia-related variation is maintained in the population.



## METHODS

### *Case sample collection*

We collected blood samples from those with treatment-resistant schizophrenia (TRS) in the UK through the mandatory clozapine blood-monitoring system for those taking clozapine, an antipsychotic licensed for TRS. Following national research ethics approval and in line with UK Human Tissue Act regulations we worked in partnership with the commercial companies that manufacture and monitor clozapine in the UK. We ascertained anonymous aliquots of the blood samples collected as part of the regular blood monitoring that takes place whilst taking clozapine due to a rare haematological adverse effect, agranulocytosis. The CLOZUK1 sample was assembled in collaboration with Novartis (Basel, Switzerland). The company, through their proprietary Clozaril® Patient Monitoring Service (CPMS), provided whole-blood samples and anonymised phenotypic information for 6,882 individuals with TRS (5528 cases post-QC), which were included in the in a recent schizophrenia GWAS by the PGC<sup>5</sup>. The CLOZUK2 sample, previously unreported, was assembled in collaboration with the other major company involved in the supply and monitoring of clozapine in the UK, Leyden Delta (Nijmegen, Netherlands). The company, through their proprietary Zaponex® Treatment Access System (ZTAS), provided whole-blood samples and anonymised phenotypic information for 7,417 of those taking clozapine (4973 cases post-QC). Both Clozaril® and Zaponex® are bioequivalent brands of clozapine licensed in the UK<sup>56</sup>.

We restricted the CLOZUK1 and CLOZUK2 samples to those with a clinician reported diagnosis of treatment-resistant schizophrenia. The UK National Institute for Health and Care Excellence (NICE) advise prescription of clozapine is reserved for those with schizophrenia in whom two trials of antipsychotics have failed (including one second-generation antipsychotic)<sup>57</sup> which mirrors the criteria for licensed use of clozapine. The sole alternative

licensed indication for clozapine in the UK is for the management of resistant psychosis in Parkinson's disease (PD)<sup>58</sup> and, although this is a rare indication, we excluded PD patients (n=8) from the case dataset. We also excluded those with off-license indications, which included those with alternative clinician diagnoses of bipolar affective disorder and personality disorders (n=56). Together with the clinical guidelines outlined, these exclusions ensure that CLOZUK1 and CLOZUK2 samples are from those patients that conform to a clinical description of TRS. We have reported the use of CLOZUK1 as a schizophrenia dataset in previous publications<sup>5,7,59,60</sup> and have presented evidence to support the use of TRS-defined individuals as valid schizophrenia samples<sup>32</sup>, which we have updated and present in **Supplementary Note 1**, including validation of a clinician diagnosis of TRS against research diagnostic criteria for schizophrenia.

In addition we also included in our analysis a more conventional cohort of UK-based patients with schizophrenia (CardiffCOGS). Recruitment was via secondary care, mainly outpatient, NHS mental health services in Wales and England. These patients were not exclusively taking clozapine at the time of their recruitment. All cases underwent a SCAN interview<sup>61</sup> and case note review followed by consensus research diagnostic procedures and were included if they had a DSM-IV schizophrenia or schizoaffective disorder-depressive type diagnosis, as previously reported<sup>7</sup>. The CardiffCOGS samples were recruited and genotyped in two waves: CardiffCOGS1 (512 cases, included in a previous GWAS<sup>5</sup>) and CardiffCOGS2 (247 cases).

Genotyping for these case samples was performed by the Broad Institute (Massachusetts, USA) for the CLOZUK1 sample and CardiffCOGS1 cases, using Illumina HumanOmniExpress-12 and OmniExpressExome-8 chips as described elsewhere<sup>7</sup>. The CardiffCOGS2 cases and the CLOZUK2 sample were genotyped by deCODE Genetics (Reykjavík, Iceland), using Illumina HumanOmniExpress-12 chips.

As all of these samples are intrinsically related and their recruitment and genotyping conforms to clinical and technical standards, thus we have combined them and used the term “CLOZUK” throughout this manuscript to describe the schizophrenia case dataset.

### ***Control sample collection***

Control samples were collected from publicly available sources (EGA) or through collaboration with the holders of the datasets. Individual datasets were curated using the same procedures as the case-only datasets. In order to maximize the numbers of individuals that could be effectively included in the GWAS without introducing confounders, these datasets were chosen on the basis of having recruited individuals with self-reported UK ancestry (either exclusively or primarily) and having been genotyped on Illumina chips. A summarized view of all the datasets included in the GWAS is provided in **Supplementary Note 2**, which includes further details of the control datasets.

### ***Genotype quality-control (QC)***

Given the many data sources used and the variety of genotyping chips available, a stringent quality control (allowing only 2% of missing SNP and individual data) was performed separately in each individual dataset, using PLINK v1.9<sup>62</sup> and following standard procedures<sup>63</sup>. To facilitate merging and to avoid common sources of batch effects<sup>64</sup>, all SNPs in each dataset were also aligned to the plus strand of the human genome (build 37p13), removing strand-ambiguous markers in the process. As most control datasets lacked any markers in the X and Y chromosomes, or in the mitochondrial DNA, every SNP from these regions was discarded in the combined genotype data. The final merge of all case and control datasets left 203,436 overlapping autosomal SNPs.

All individuals were imputed simultaneously in the Cardiff University high-performance computing cluster RAVEN<sup>65</sup>, using the SHAPEIT/IMPUTE2 algorithms<sup>66,67</sup>. As reference panels, a combination of the 1000 Genomes phase 3 (1KGPp3) and UK10K datasets was used, as this has previously been shown to increase the accuracy of imputation for individuals of British ancestry, particularly for rare variants<sup>68</sup>.

After imputation, a principal component analysis (PCA) of common variants (MAF higher than 5%) was carried out to obtain a general summary of the population structure of the sample, using the EIGENSOFT v6 toolset<sup>69</sup>. A plot of the first two PCs showed the existence of a large fraction of cases (~20%) with no overlapping controls (**Figure S1, A**). A comparison with the 1KGPp3 dataset, performed using PCA and ADMIXTURE<sup>70</sup> estimates, showed that most of these cases were similar in genetic ancestry to non-European individuals, namely from the East Asian or West African superpopulations (**Figure S1, B**). In order to ameliorate population stratification in the association analysis<sup>71</sup>, all individuals not falling into an area delimited by the mean and 3 standard deviations of the two first principal components of the control samples were excluded from further analyses (**Figure S1, C**). By repeating PCA only on the selected individuals, no outliers could be detected in the first two principal components, and ADMIXTURE plots were homogenised as well (**Figure S2**).

The CLOZUK sample was further pruned by removing all individuals with inbreeding coefficients ( $F$ ) higher than 0.2, and leaving only a random member of each pair with a relatedness coefficient ( $\hat{\pi}$ ) higher than 0.2. Furthermore, to ensure the independence of our analyses with previous GWAS conducted by the Schizophrenia Working Group of the PGC, relatedness coefficients of CLOZUK individuals were also calculated with all the individual datasets included in the latest PGC GWAS<sup>5</sup> following approval by the Consortium. Detected genetic relatives (or duplicates) were excluded in CLOZUK in the same way as intra-population relatives. After this imputation and curation process, 35,802 samples (11,260

cases and 24,542 controls) with 9.65 million imputed markers (INFO>0.3 and MAF>0.001) remained in the CLOZUK dataset.

### ***GWAS and reporting of independently-associated regions***

The CLOZUK schizophrenia GWAS was performed using logistic regression with imputation probabilities (“dosages”) adjusted for 11 PCA covariates. These covariates were chosen as those nominally significant ( $p < 0.05$ ) in a logistic regression for association with the phenotype<sup>72</sup>. To avoid overburdening the GWAS power by adding too many covariates to the regression model<sup>73</sup>, only the first twenty PCs were considered and tested for inclusion, as higher numbers of PCs only become useful for the analysis of populations that bear strong signatures of complex admixture<sup>74</sup>. The final set of covariates included the first five PCs (as recommended for most GWAS approaches<sup>75</sup>) and PCs 6, 9, 11, 12, 13 and 19. Quantile-quantile (QQ) and Manhattan plots are shown in **Figure S3** and **S4**.

In order to identify independent signals among the regression results, signals were amalgamated into putative associated loci using the same two-step strategy and parameters as PGC (**Extended Data Table 1**). In this procedure, regular LD-clumping is performed ( $r^2 = 0.1$ ,  $p < 1 \times 10^{-4}$ ; window size <3 Mb) in order to obtain independent index-SNPs. Afterwards loci were defined for each index SNP as the genomic region which contains all other imputed SNPs within an  $r^2 \geq 0.6$ . To avoid inflating the number of signals in gene-dense regions or in those with complex LD, all loci within 250kb of each other were annealed.

### ***Meta-analysis with PGC***

A total of 6,040 cases and 5,719 controls from CLOZUK were included in the recent PGC study<sup>5</sup>. We reanalysed the PGC data after excluding all these cases and controls, obtaining a

sample termed ‘INDEPENDENT PGC’ (29,415 cases and 40,101 controls). Adding the summary statistics from this independent sample to the CLOZUK GWAS results allowed for a combined analysis of 40,675 cases and 64,643 controls (without duplicates or related samples). This meta-analysis was performed using the fixed-effects procedure in METAL<sup>76</sup> with weights derived from standard errors. For consistency with the PGC analysis, additional filters (INFO>0.6 and MAF>0.01) were applied to the CLOZUK and INDEPENDENT PGC summary statistics, leaving 8 million markers in the final meta-analysis results. QQ and Manhattan plots are shown in **Figure S5** and **Figure 2**. The same procedure as above was used in order to report independent loci from this analysis (**Extended Data Table 5**, **Extended Data Table 6**). As raw PGC genotypes were not available for the LD-clumping procedure, 1KGPp3 was used as a reference.

### *Estimation and assessment of a polygenic signal*

Association signals caused by the vast polygenicity underlying complex traits can be hard to distinguish from confounders related to sample relatedness and population stratification. In order to effectively disentangle this issue, we used the software LD-Score v1.0 to analyse the summary statistics of the CLOZUK GWAS, and estimate the contribution of confounding biases to our results by LDSR<sup>14</sup>. An LD-reference was generated using 1KGPp3 using the standard parameters implemented in the software. In order to improve accuracy, indels were discarded and markers used in this procedure were restricted to those with INFO>0.9 and MAF>0.01, a total of 5.16 million SNPs. The resulting LD-score intercept was  $1.085 \pm 0.010$ , which compared to a mean  $\chi^2$  of 1.417 indicates a polygenic signal contribution of at least 80%. This is in line with other well-powered GWAS studies of complex human traits<sup>14</sup>, including schizophrenia<sup>5</sup>. SNP-based heritability ( $h^2_{\text{SNP}}$ ) in our datasets (CLOZUK, INDEPENDENT PGC and the CLOZUK+PGC meta-analysis) was also calculated in this

analysis, and transformed to a liability scale using a population prevalence of 1% (registry-based lifetime prevalence<sup>77</sup>). For reference and compatibility with epidemiological studies of schizophrenia, prevalence estimates of 0.7% (lifetime morbid risk<sup>78</sup>) and 0.4% (point prevalence<sup>78</sup>) were used for additional liability-scale  $h^2_{\text{SNP}}$  calculations (**Extended Data Table 2**).

We also used LD-Score to compare the genetic architecture of CLOZUK and INDEPENDENT PGC, by calculating the correlation of their summary statistics<sup>79</sup>. A genetic correlation coefficient of  $0.9541 \pm 0.0297$  was obtained, with a p-value of  $6.63 \times 10^{-227}$ . There were 76 independent SNPs at a genome-wide significant (GWS) level in the INDEPENDENT PGC dataset after excluding the extended major histocompatibility complex region (xMHC). Using binomial sign tests based on clumped subsets of SNPs we found all but 1 (98.6%) of the 76 GWS SNPs from the INDEPENDENT PGC were associated with the same direction of effect in the CLOZUK sample, a result highly unlikely to reflect chance<sup>80</sup> ( $p=2.04 \times 10^{-21}$ , **Extended Data Table 3**). Even of those with an association p-value less than  $10^{-4}$  in the INDEPENDENT PGC sample, 82% showed enrichment in the CLOZUK cases ( $p=3.44 \times 10^{-113}$ ), confirming very large numbers of true associations will be discovered amongst these SNPs with increased sample sizes. Additionally, the new sample introduced in this study (CLOZUK2) was compared by the same methods with the independent PGC dataset and showed results consistent with the full CLOZUK analysis, providing molecular validation of this sample as a schizophrenia sample (**Extended Data Table 3**).

We went on to conduct polygenic risk score analysis. Polygenic scores for CLOZUK were generated from INDEPENDENT PGC as a training set, using the same parameters for risk profile score (RPS) analysis in PGC<sup>5</sup>, arriving at a high-confidence set of SNPs for RPS estimation by removing the xMHC region, indels, and applying  $\text{INFO} > 0.9$  and  $\text{MAF} > 0.1$  cut-offs. Scores were generated from the imputation dosage data, using a range of p-value

thresholds for SNP inclusion<sup>81</sup> ( $5 \times 10^{-8}$ ,  $1 \times 10^{-5}$ , 0.001, 0.05 and 0.5). In this way, we can assess the presence of a progressively increasing signal-to-noise ratio in relation to the number of markers included<sup>82</sup>. As in the PGC study, we find the best p-value threshold for discrimination to be 0.05 and report highly significant polygenic overlap between the INDEPENDENT PGC and CLOZUK samples ( $p < 1 \times 10^{-300}$ ,  $r^2 = 0.12$ , **Extended Data Table 4**), confirming the validity of combining the datasets. For comparison with other studies we also report polygenic variance on the liability scale<sup>83</sup>, which amounted to 5.7% for CLOZUK at the 0.05 p-value threshold (**Extended Data Table 4**). As in the PGC study the limited  $r^2$  and AUC in this analysis restricts the current clinical utility of these scores in schizophrenia.

### ***Gene set analysis***

In order to assess the enrichment of sets of functionally related genes, we used MAGMA v1.03<sup>84</sup> on the CLOZUK+PGC meta-analysis summary statistics, after excluding the xMHC region. First, gene-wide p-values were calculated by combining the p-values of all SNPs inside genes after accounting for linkage disequilibrium (LD) and outliers. This was performed allowing for a window of 35 kb upstream and 10 kb downstream of each gene in order to capture the signal of nearby SNPs that could fall in regulatory regions<sup>85,86</sup>. Next we calculated competitive gene set p-values on the gene-wide p-values after accounting for gene size, gene set density and LD between genes. For multiple testing correction in each gene set collection, a FWER<sup>87</sup> was computed using 100,000 re-samplings.

We performed sequential analyses using three approaches:

**1. Loss-of-function intolerant genes:** We tested the enrichment of the loss-of-function (LoF) intolerant genes described by ExAC<sup>21</sup>. This set comprises all genes defined in the ExAC database<sup>88</sup> as having a probability of LoF-intolerance (pLI) statistic higher than 90%. While



these genes do not form part of cohesive biological processes or phenotypes, they have been previously found to be highly expressed across tissues and developmental stages<sup>21</sup>. Also, they are enriched for hub proteins<sup>89</sup>, which makes them interesting candidates for involvement in the “evolutionary canalisation” processes that have been proposed to lead to pleiotropic, complex disorders<sup>90</sup>.

**2. CNS-related genes:** These gene sets were compiled in our recent study<sup>32</sup>, and include 134 gene sets related to different aspects of central nervous system function and development. These include, among others, gene sets which have been implicated in schizophrenia by at least two independent large-scale sequencing studies<sup>8,35</sup>: targets of the fragile-X mental retardation protein (FMRP<sup>33</sup>), constituents of the N-methyl-D-aspartate receptor (NMDAR<sup>91</sup>) and activity-regulated cytoskeleton-associated protein complexes (ARC<sup>92,93</sup>), as well as CNS and behavioural gene sets from the Mouse Genome Informatics database version 6 (<http://www.informatics.jax.org>)<sup>37</sup>.

**3. Data-driven:** The final systems genomic analysis was designed as an “agnostic” approach, and thus a large number of gene sets from different public sources was included, which summed to a total of 6,677 sets. Gene set sources were selected to encompass a comprehensive collection of biochemical pathways and gene regulatory interaction networks, not necessarily conceptually related to psychiatric disorders. Similar approaches have been successful elsewhere<sup>86,94</sup>. In building this analytic approach, the LoF intolerant gene set and all sets in the *CNS-related* collection were used. Additionally, 2,693 gene sets with direct experimental evidence and a size of 10-200 genes<sup>86</sup> were extracted from the Gene Ontology (GO<sup>95</sup>) database release 01/02/2016; 1,787 gene sets were extracted from the 4<sup>th</sup> ontology level of the Mouse Genome Informatics database version 6; 1,585 gene sets were extracted from REACTOME<sup>96</sup> version 55; 290 gene sets were extracted from KEGG<sup>97</sup> release 04/2015; and 187 gene sets were extracted from OMIM<sup>98</sup> release 01/02/2016.

Detailed results of the analyses of the CNS-related and data-driven collection are given in **Extended Data Table 11** and **Extended Data Table 12**. Reported numbers of genes in each gene set are those with available data in the meta-analysis. This may differ from the original gene set description as some genic regions (such as those in the X chromosome) had null or poor SNP coverage. The gene content of the CNS-related gene sets that survive conditional analysis (“significant”) is given in MAGMA format in the **Data Supplement**.

### *Partitioned heritability analysis of gene sets*

It is known that the power of a gene set analysis is closely related to the total heritability of the phenotype and the specific heritability attributable to the tested gene set<sup>99</sup>. In order to assess the heritability explained by the genes carried forward after the main gene set analysis, LD-Score was again used to compute a partitioned heritability estimate of CLOZUK+PGC using the gene sets as SNP annotations. As in the MAGMA analysis, the xMHC region was excluded from the summary statistics. These were also trimmed to contain no indels, and only markers with INFO > 0.9 and MAF > 0.01, for a total of 4.64 million SNPs. As a recognised caveat of this procedure is that model misspecification can inflate the partitioned heritability estimates<sup>22</sup>, all gene sets were annotated twice: Once using their exact genomic coordinates (extracted from the NCBI RefSeq database<sup>100</sup>) and another with regulatory regions taken into account using the same upstream/downstream windows as in the MAGMA analyses. Additionally, all SNPs not directly covered by our gene sets of interest were explicitly included into other annotations (“non-genic”, “genic but not LoF-intolerant”) based on their genomic location. Finally, the “baseline” set of 53 annotations from Finucane et al. 2015<sup>22</sup>, which recapitulates important molecular properties such as presence of enhancers or phylogenetic conservation, was also incorporated in the model. All of these annotations were then tested jointly for heritability enrichment. We note that using exact genic coordinates or

adding regulatory regions made little difference to the estimated enrichment of our gene sets, and thus throughout the manuscript we report the latter for consistency with the MAGMA gene set analysis (**Figure 2; Extended Data Table 10**).

### *Natural selection analyses*

We aimed to explore the hypothesis that some form of natural selection is linked to the maintenance of common genetic risk in schizophrenia<sup>101-103</sup>. In order to do this, for all SNPs included in the CLOZUK+PGC meta-analysis summary statistics, we obtained four different genome-wide metrics of positive selection (iHS<sup>104</sup>, XP-EEH<sup>105</sup>, CMS<sup>106</sup> and CLR<sup>25</sup>), one of background selection (B-statistic<sup>26</sup>, post-processed by Huber et al. 2016<sup>25</sup>) and one of Neanderthal introgression (average posterior probability LA<sup>107</sup>). The use of different statistics is motivated by the fact that each of them is tailored to detect a particular selective process that acted on a particular timeframe (see Vitti et al. 2013<sup>51</sup> for a review). For example, iHS and CMS are based on the inference of abnormally long haplotypes, and thus are better powered to detect recent selective sweeps that occurred during the last ~30,000 years<sup>108</sup>, such as those linked to lactose tolerance or pathogen response<sup>106</sup>. On the other hand, CLR incorporates information about the spatial pattern of genomic variability (the site frequency spectrum<sup>109</sup>), and corrects explicitly for evidences of background selection, thus being able to detect signals from 60,000 to 240,000 years ago<sup>25</sup>. The B-statistic uses phylogenetic information from other primates (chimpanzee, gorilla, orang-utan and rhesus macaque) in order to infer the reduction in allelic diversity that exists in humans as a consequence of purifying selection on linked sites over evolutionary timeframes<sup>48</sup>. As the effects of background selection on large genomic regions can mimic those of positive selection<sup>46</sup>, it is possible that the B-statistic might amalgamate both, though the rather large diversity reduction that it infers for the human genome as a whole suggests any bias due to positive

selection is likely to be minor<sup>44</sup>. Finally, XP-EEH is a haplotype-based statistic which compares two population samples, and thus its power is increased for alleles that have suffered differential selective pressures since those populations diverged<sup>105</sup>. Though methodologically different, LA has a similar rationale by comparing human and Neanderthal genomes<sup>107</sup>, in order to infer the probability of each human haplotype to have been the result of an admixture event with Neanderthals.

For this work, CLR, CMS, B-statistic and LA were retrieved directly from their published references, and lifted over to GRC37 genomic coordinates if required using the ENSEMBL LiftOver tool<sup>110,111</sup>. As the available genome-wide measures of iHS and XP-EEH were based on HapMap3 data<sup>112</sup>, both statistics were re-calculated with the HAPBIN<sup>113</sup> software directly on the EUR superpopulation of the 1KGp3 dataset, with the AFR superpopulation used as the second population for XP-EEH. Taking advantage of the fine-scale genomic resolution of these statistics (between 1-10 kb), all SNP positions present in CLOZUK+PGC were assigned a value for each measure, either directly (if the position existed in the lifted-over data) or by linear interpolation. To simplify the interpretation of our results, all measures were transformed before further analyses to a common scale, in which larger values indicate stronger effect of selection or increased probability of introgression.

Heritability enrichment of these statistics was tested by the LD-Score partitioned heritability procedure. We derived binary annotations from the natural selection metrics by dichotomising at extreme cut-offs defined by the top 2%, 1% and 0.5% of the values of each metric in the full set of SNPs. This approach is widely used in evolutionary genomics, due to the difficulty of setting specific thresholds to define regions under selection<sup>25,51</sup>. Consistent with the previously described LDSR partitioned heritability protocol, enrichment was estimated for all binary annotations after controlling for 3 main categories of our set-based analysis (“non-genic”, “genic” and “genic LoF-intolerant”) and the 53 “baseline” categories

of Finucane et al. 2015<sup>22</sup>. We note that the LD-Score software allows use of an extension of the partitioned heritability framework to test the metrics as fully quantitative annotations. Results of this analysis replicated those reported here (*data not shown*), though it did not constitute our main approach as methodological details on this feature of LDSR are yet to be published.

		Top 2% of scores (genome-wide)		Top 1% of scores (genome-wide)		Top 0.5% of scores (genome-wide)	
Metric	Ref.	Enrichment	2-sided p-value	Enrichment	2-sided p-value	Enrichment	2-sided p-value
Background Selection (B-statistic)	26	<u>2.0612</u>	<u>0.0008</u>	<u>2.6477</u>	<u>1.2721x10<sup>-4</sup></u>	<u>2.8073</u>	<u>1.0226x10<sup>-4</sup></u>
Positive selection (CLR)	25	0.6069	0.0506	<u>0.3154</u>	<u>1.1349x10<sup>-5</sup></u>	<u>0.2382</u>	<u>1.1120x10<sup>-4</sup></u>
Positive selection (CMS)	106	0.1703	0.0036	0.1902	0.0489	-0.0108	0.0736
Positive selection (XP-EEH)	105	0.9918	0.9824	0.6211	0.4882	-0.0018	0.1666
Positive selection (iHS)	104	0.8593	0.7095	0.7649	0.6992	1.3686	0.7130
Neanderthal posterior probability (LA)	107	0.7465	0.2175	0.7895	0.4930	0.9067	0.8428

**Table 1.** Heritability analysis of natural selection metrics.

Partitioned LDSR regression results for SNPs thresholded by extreme values (defined as top percentiles vs all other SNPs) of each natural selection metric. All tests have been adjusted for 56 "baseline" annotations, which include categories such as "LoF-intolerant" and "conserved" (see **Methods**). Negative enrichments should be considered zero (no contribution to  $h^2_{\text{SNP}}$  by these SNPs). Underlined values indicate results surviving correction after adjusting for all tests (Bonferroni  $\alpha = 0.05/18 = 0.0028$ ). Reference numbers for each metric indicate references in the main text.

<b>Genes implicated by ‘single gene’ GWAS locus (locus overlaps with only one gene boundary (+/- 20kb))</b>			
<b>Gene</b>	<b>OR</b>	<b>P value</b>	<b>Comment</b>
<b><i>GABBR2</i></b>	0.92	1.02x10 <sup>-9</sup>	Encodes GABA-B receptor 2, a metabotropic G-protein coupled receptor. This is the first direct association of a GABA receptor gene in schizophrenia GWAS. Ionotropic GABA-A receptor complexes have recently been implicated at a pathway level in a schizophrenia CNV study <sup>32</sup> .
<b><i>PDE4B</i></b>	0.95	3.35x10 <sup>-8</sup>	An enzyme involved in cAMP signalling and an interactor of <i>DISC1</i> . Translocations of both genes have been implicated in multiple psychiatric disorders <sup>30,114</sup> .
<b><i>CACNA1D</i></b>	1.07	1.38x10 <sup>-8</sup>	A voltage gated, L-type calcium channel sub-unit that is implicated in dopamine neuron pathology in Parkinson’s disease <sup>115</sup> . There are four L-type alpha sub-units and this is the second to be associated with schizophrenia. The other, <i>CACNA1C</i> , is an established candidate gene (p=5.6x10 <sup>-20</sup> in this meta-analysis).
<b><i>RBFOX1</i></b>	0.92	1.44x10 <sup>-8</sup>	Encodes FOX1, a neuronal RNA binding protein and splicing factor implicated in autism and neurodevelopment <sup>116</sup> . A number of additional novel loci highlight genes involved in transcription including <i>FOXPI</i> and <i>FOXGI</i> , both of which have also been implicated in autism spectrum disorders and in regulating GABAergic/glutamatergic signalling <sup>117-119</sup> . In addition <i>ASCLI</i> (a single gene locus in both GWAS and gene-wide analysis is a novel transcription factor gene which is involved in neuronal cell differentiation <sup>120</sup> .
<b>Genes implicated by gene-based analysis (no other genome-wide significant gene within 100kb)</b>			
<b>Gene</b>	<b>P Value</b>	<b>Comment</b>	
<b><i>ACHE</i></b>	1.8x10 <sup>-6</sup>	Encodes acetyl cholinesterase (AChE) which hydrolyzes and inactivates the neurotransmitter acetylcholine. AChE inhibitors are used to treat cognitive impairment in dementia and have shown limited symptomatic benefits as adjunctive agents in schizophrenia trials <sup>121,122</sup> . Recently identified as enriched for de novo mutations in ASD <sup>28</sup> . We also implicate other cholinergic genes in the gene-based analysis: <i>CHRNA3/CHRNA5/CHRNB4</i> (p values <7.6x10 <sup>-7</sup> ), <i>CHRM4</i> (p=1.7x10 <sup>-12</sup> ); as well as GWAS analyses: <i>CHRNA2</i> (p= 7.6x10 <sup>-12</sup> ) and <i>CHRM3</i> (p=4.7x10 <sup>-10</sup> ).	
<b><i>NRXN1</i></b>	6.6x10 <sup>-7</sup>	The first genome-wide significant common variant finding for this gene. <i>NRXN1</i> has been robustly implicated in schizophrenia pathogenesis by single-gene copy number variation analysis <sup>29</sup> .	
<b><i>ERBB4</i></b>	1.3x10 <sup>-7</sup>	A member of the tyrosine kinase receptor family that binds and is activated by the schizophrenia candidate gene neuregulin1 ( <i>NRG1</i> ) <sup>123</sup> .	
<b><i>SLC6A11</i></b>	1.3x10 <sup>-6</sup>	Encodes a sodium-dependent GABA transporter, providing further support GABA involvement in schizophrenia pathogenesis. Another implicated gene, <i>CLCN3</i> , encodes a voltage-gated chloride channel involved in transmitter loading of GABAergic synapse vesicles. <i>SLC6A11</i> is one of two major GABA transporters in the brain, rare variants in which have been implicated in epilepsy, ID, and stereotypic behaviour <sup>124,125</sup> .	

**Table 2.** Genes with pre-existing genetic and functional hypotheses in schizophrenia or related neurodevelopmental disorders. Additional details an **Extended Data Table 6** and **Extended Data Table 9**.

<b>Gene Set</b>	<b>Number of genes</b>	<b>Enrichment p value (FWER)</b>	<b>Conditional p value</b>
<b>Targets of FMRP<sup>33</sup></b>	798	$1 \times 10^{-5}$	$1.9 \times 10^{-8}$
<b><i>Abnormal Behaviour</i> (MP:0004924)</b>	1939	$1.8 \times 10^{-4}$	$1.4 \times 10^{-5}$
<b>5-HT<sub>2C</sub> receptor complex<sup>39</sup></b>	16	0.029	0.001
<b><i>Abnormal Nervous System Electrophysiology</i> (MP:0002272)</b>	201	0.003	0.002
<b>Voltage-gated calcium channel complexes<sup>38</sup></b>	196	0.011	0.016
<b><i>Abnormal Long Term Potentiation</i> (MP:0002207)</b>	142	0.030	0.031

**Table 3. Functional gene set analysis highlights six independent gene sets associated with schizophrenia**

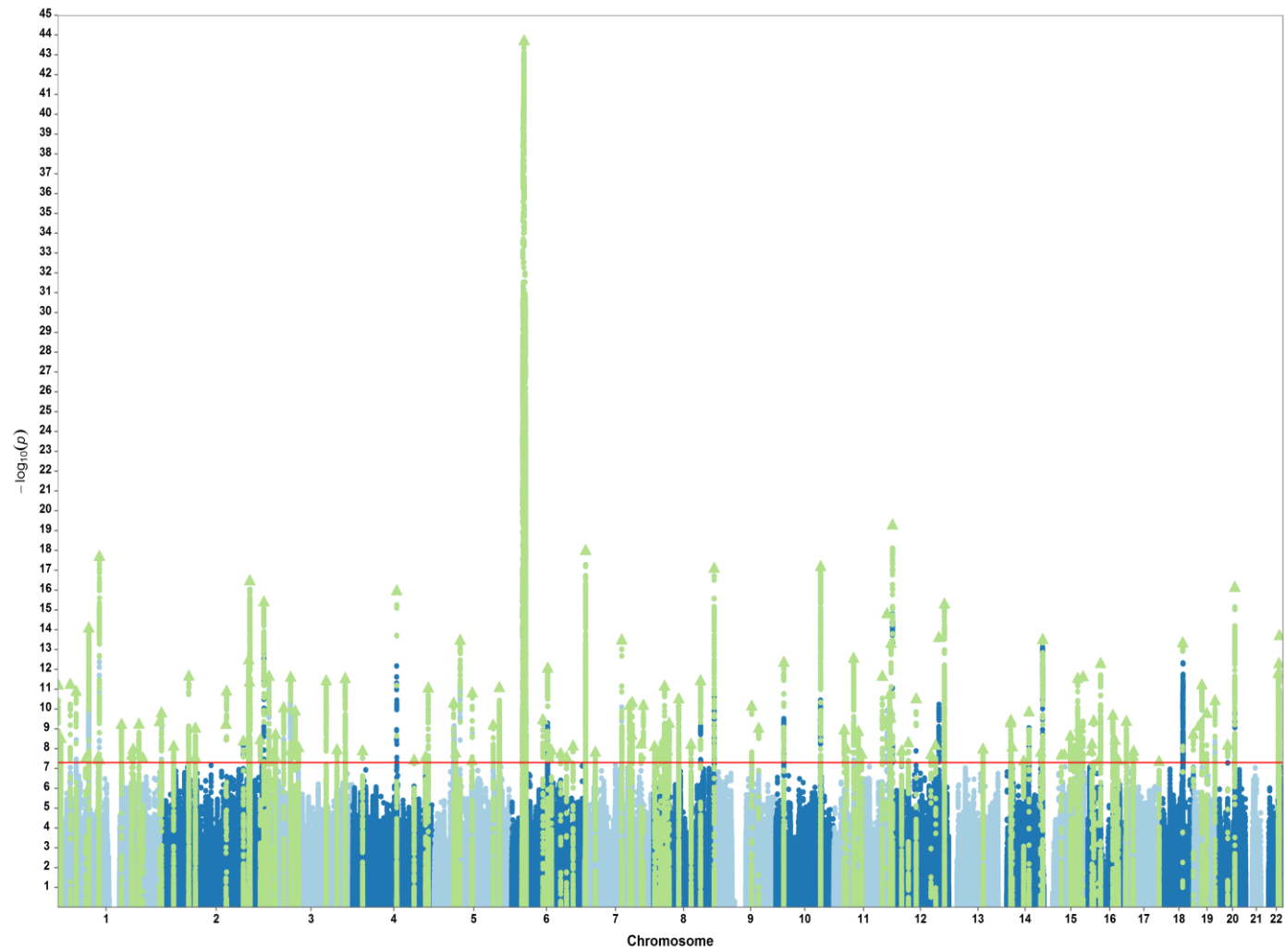
FMRP: Fragile X Mental Retardation Protein.

MP refers to Mammalian Phenotype Ontology term of the MGI: Mouse Genome Informatics (<http://www.informatics.jax.org>)<sup>37</sup>, from which gene sets were derived.

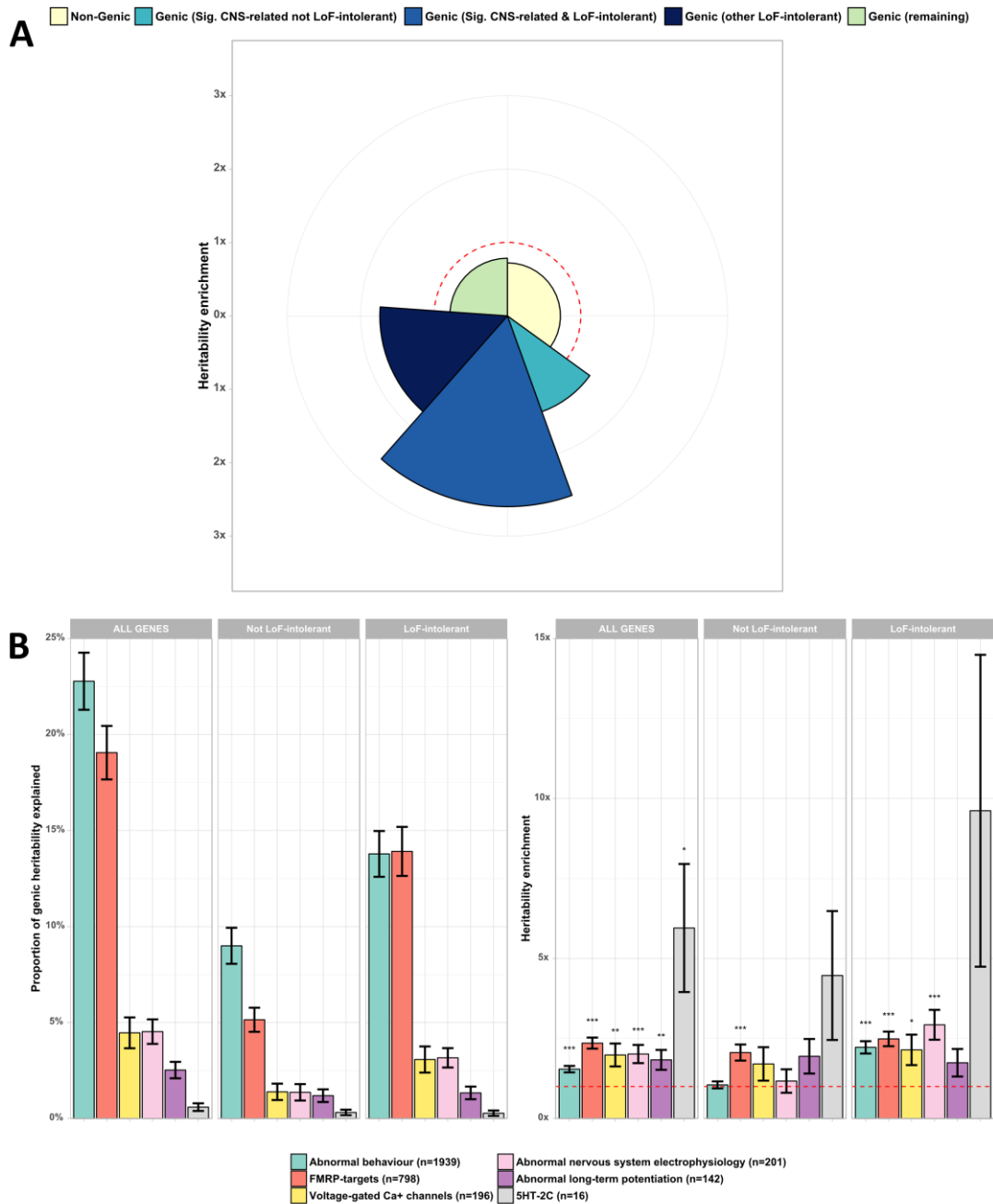
FWER: Westfall-Young family-wise error rate, as implemented in MAGMA<sup>17,87</sup>.

Conditional p value refers to stepwise conditional analysis that adjusts sequentially for ‘stronger’ associated gene sets.





**Figure 1.** Manhattan plot of schizophrenia GWAS associations from the meta-analysis of CLOZUK and an independent PGC dataset (N=105,318; 40,675 cases and 64,643 controls).



**Figure 2.** Partitioned heritability analysis of gene sets in schizophrenia. **A:** Heritability of genomic partitions and the six conditionally independent (“significant”) gene sets (**Table 3**). Radius of each segment indicates the degree of enrichment, while the arc (angle of each slice) indicates the percentage of total SNP-based heritability explained. No relative enrichment or depletion (enrichment=1) is shown by the dashed red line. **B:** Heritability of the significant CNS gene sets dissected by their overlap with LoF-intolerant genes. Asterisks indicate the significance of each heritability enrichment (\*  $\leq 0.05$ ; \*\*  $\leq 0.01$ ; \*\*\*  $\leq 0.001$ ).

## REFERENCES

1. Owen, M.J., Sawa, A. & Mortensen, P.B. Schizophrenia. *Lancet* (2016).
2. Thornicroft, G. Physical health disparities and mental illness: the scandal of premature mortality. *The British Journal of Psychiatry* **199**, 441-442 (2011).
3. Olfson, M., Gerhard, T., Huang, C., Crystal, S. & Stroup, T. Premature mortality among adults with schizophrenia in the united states. *JAMA Psychiatry* **72**, 1172-1181 (2015).
4. Morgan, C. *et al.* Reappraising the long-term course and outcome of psychotic disorders: the AESOP-10 study. *Psychological Medicine* **44**, 2713-2726 (2014).
5. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
6. Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nature Neuroscience* **19**, 571-7 (2016).
7. Rees, E. *et al.* Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry* **204**, 108-14 (2014).
8. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-190 (2014).
9. Power, R.A., Kyaga, S., Uher, R. & *et al.* Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry* **70**, 22-30 (2013).
10. Huxley, J., Mayr, E., Osmond, H. & Hoffer, A. Schizophrenia as a Genetic Morphism. *Nature* **204**, 220-221 (1964).
11. Shaner, A., Miller, G. & Mintz, J. Schizophrenia as one extreme of a sexually selected fitness indicator. *Schizophrenia Research* **70**, 101-109 (2004).
12. Srinivasan, S. *et al.* Genetic Markers of Human Evolution Are Enriched in Schizophrenia. *Biological Psychiatry* (2016).
13. Uher, R. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Mol Psychiatry* **14**, 1072-1082 (2009).
14. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-295 (2015).
15. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics* **45**, 1150-1159 (2013).
16. Shi, Y. *et al.* Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nature Genetics* **43**, 1224-1227 (2011).

17. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology* **11**, e1004219 (2015).
18. McRae, J.F. *et al.* Prevalence, phenotype and architecture of developmental disorders caused by de novo mutation. *bioRxiv* (2016).
19. Kosmicki, J. *et al.* Refining the role of de novo protein truncating variants in neurodevelopmental disorders using population reference samples. *bioRxiv* (2016).
20. Samocha, K.E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nature Genetics* **46**, 944-950 (2014).
21. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* (2015).
22. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* **47**, 1228-1235 (2015).
23. McRae, J.F. *et al.* Prevalence, phenotype and architecture of developmental disorders caused by de novo mutation. *bioRxiv* (2016).
24. Takata, A., Ionita-Laza, I., Gogos, Joseph A., Xu, B. & Karayiorgou, M. De Novo Synonymous Mutations in Regulatory Elements Contribute to the Genetic Etiology of Autism and Schizophrenia. *Neuron* **89**, 940-947 (2016).
25. Huber, C.D., DeGiorgio, M., Hellmann, I. & Nielsen, R. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Molecular Ecology* **25**, 142-156 (2016).
26. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread Genomic Signatures of Natural Selection in Hominid Evolution. *PLoS Genet* **5**, e1000471 (2009).
27. Chakravarti, A., Clark, A.G. & Mootha, V.K. Distilling Pathophysiology from Complex Disease Genetics. *Cell* **155**, 21-26 (2013).
28. Sanders, S.J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-33 (2015).
29. Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Human Molecular Genetics* **18**, 988-96 (2009).
30. Millar, J.K. *et al.* DISC1 and PDE4B Are Interacting Genetic Factors in Schizophrenia That Regulate cAMP Signaling. *Science* **310**, 1187-1191 (2005).
31. Stefansson, H. *et al.* Neuregulin 1 and susceptibility to schizophrenia. *American Journal of Human Genetics* **71**, 877-892 (2002).
32. Pocklington, A.J. *et al.* Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia. *Neuron* **86**, 1203-14 (2015).
33. Darnell, J.C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247-61 (2011).

34. Iossifov, I. *et al.* De Novo Gene Disruptions in Children on the Autistic Spectrum. *Neuron* **74**, 285-299 (2012).
35. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179-184 (2014).
36. Szatkiewicz, J.P. *et al.* Copy number variation in schizophrenia in Sweden. *Molecular Psychiatry* **19**, 762-73 (2014).
37. Blake, J.A., Bult, C.J., Eppig, J.T., Kadin, J.A. & Richardson, J.E. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Research* **42**, D810-7 (2014).
38. Müller, C.S. *et al.* Quantitative proteomics of the Cav2 channel nano-environments in the mammalian brain. *Proceedings of the National Academy of Sciences* **107**, 14950-14957 (2010).
39. Becamel, C. *et al.* Synaptic multiprotein complexes associated with 5-HT(2C) receptors: a proteomic approach. *The EMBO journal* **21**, 2332-42 (2002).
40. Liu, J. *et al.* Prediction of Efficacy of Vabicaserin, a 5-HT(2C) Agonist, for the Treatment of Schizophrenia Using a Quantitative Systems Pharmacology Model. *CPT: Pharmacometrics & Systems Pharmacology* **3**, e111 (2014).
41. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
42. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476-482 (2011).
43. Rands, C.M., Meader, S., Ponting, C.P. & Lunter, G. 8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage. *PLoS Genet* **10**, e1004525 (2014).
44. Fu, W. & Akey, J.M. Selection and Adaptation in the Human Genome. *Annual Review of Genomics and Human Genetics* **14**, 467-489 (2013).
45. Charlesworth, B. Background Selection 20 Years on: The Wilhelmine E. Key 2012 Invitational Lecture. *Journal of Heredity* (2013).
46. Charlesworth, B., Betancourt, A.J., Kaiser, V.B. & Gordo, I. Genetic Recombination and Molecular Evolution. *Cold Spring Harbor Symposia on Quantitative Biology* **74**, 177-186 (2009).
47. Comeron, J.M., Williford, A. & Kliman, R.M. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Heredity* **100**, 19-31 (2007).
48. Nordborg, M., Charlesworth, B. & Charlesworth, D. The effect of recombination on background selection. *Genetics Research* **67**, 159-174 (1996).

49. North, T.L. & Beaumont, M.A. Complex trait architecture: the pleiotropic model revisited. *Scientific Reports* **5**, 9351 (2015).
50. Rockman, M.V., Skrovanek, S.S. & Kruglyak, L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* **330**, 372-6 (2010).
51. Vitti, J.J., Grossman, S.R. & Sabeti, P.C. Detecting Natural Selection in Genomic Data. *Annual Review of Genetics* **47**, 97-120 (2013).
52. Stephan, W. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular ecology* **25**, 79-88 (2016).
53. Field, Y. *et al.* Detection of human adaptation during the past 2,000 years. *bioRxiv* (2016).
54. Key, F.M., Fu, Q., Romagne, F., Lachmann, M. & Andres, A.M. Human adaptation and population differentiation in the light of ancient genomes. *Nat Commun* **7**(2016).
55. Harris, K. & Nielsen, R. The Genetic Cost of Neanderthal Introgression. *Genetics* **203**, 881-891 (2016).
56. Couchman, L., Morgan, P.E., Spencer, E.P., Johnston, A. & Flanagan, R.J. Plasma Clozapine and Norclozapine in Patients Prescribed Different Brands of Clozapine (Clozaril, Denzapine, and Zaponex). *Therapeutic Drug Monitoring* **32**, 624-627 (2010).
57. National Collaborating Centre for Mental Health. Psychosis and schizophrenia in adults: The NICE guideline on treatment and management. Vol. National Clinical Guideline Number 178 (NICE, London, 2014).
58. Davie, C.A. A review of Parkinson's disease. *British medical bulletin* **86**, 109-127 (2008).
59. Hamshere, M.L. *et al.* Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Molecular Psychiatry* **18**, 738 (2013).
60. Richards, A.L. *et al.* Exome arrays capture polygenic rare variant contributions to schizophrenia. *Human Molecular Genetics* **25**, 1001-7 (2016).
61. Wing, J.K. *et al.* Scan - Schedules for Clinical-Assessment in Neuropsychiatry. *Archives of General Psychiatry* **47**, 589-593 (1990).
62. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**(2015).
63. Anderson, C.A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protocols* **5**, 1564-1573 (2010).

64. Zuvich, R.L. *et al.* Pitfalls of merging GWAS data: lessons learned in the eMERGE network and quality control procedures to maintain high data quality. *Genetic Epidemiology* **35**, 887-898 (2011).
65. Advanced Research Computing @ Cardiff (ARCCA). Introduction to RAVEN. (accessed: 29/03/2016).
66. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics* **44**, 955-959 (2012).
67. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods* **10**, 5-6 (2013).
68. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nature communications* **6**(2015).
69. Patterson, N.J., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genetics* **2**, e190 (2006).
70. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655-1664 (2009).
71. Tian, C., Gregersen, P.K. & Seldin, M.F. Accounting for ancestry: population substructure and genome-wide association studies. *Human Molecular Genetics* **17**, R143-R150 (2008).
72. Peloso, G.M. & Lunetta, K.L. Choice of population structure informative principal components for adjustment in a case-control study. *BMC Genetics* **12**, 64-64 (2011).
73. Pirinen, M., Donnelly, P. & Spencer, C.C.A. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nat Genet* **44**, 848-851 (2012).
74. Bouaziz, M., Ambroise, C. & Guedj, M. Accounting for Population Stratification in Practice: A Comparison of the Main Strategies Dedicated to Genome-Wide Association Studies. *PLoS ONE* **6**, e28845 (2011).
75. Tucker, G., Price, A.L. & Berger, B.A. Improving the Power of GWAS and Avoiding Confounding from Population Stratification with PC-Select. *Genetics* **197**, 1045-1049 (2014).
76. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
77. Perälä, J., Suvisaari, J., Saarni, S.I. & *et al.* Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Archives of General Psychiatry* **64**, 19-28 (2007).
78. McGrath, J., Saha, S., Chant, D. & Welham, J. Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. *Epidemiologic Reviews* **30**, 67-76 (2008).

79. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241 (2015).
80. Ripke, S. *et al.* Genome-wide association study identifies five new schizophrenia loci. *Nature genetics* **43**, 969-976 (2011).
81. Tansey, K.E. *et al.* Common alleles contribute to schizophrenia in CNV carriers. *Molecular Psychiatry* (2015).
82. Dudbridge, F. Polygenic Epidemiology. *Genetic Epidemiology* **40**, 268-272 (2016).
83. Lee, S.H., Goddard, M.E., Wray, N.R. & Visscher, P.M. A better coefficient of determination for genetic profile analysis. *Genetic epidemiology* **36**, 214-224 (2012).
84. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
85. Maston, G.A., Evans, S.K. & Green, M.R. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics* **7**, 29-59 (2006).
86. The Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* **18**, 199-209 (2015).
87. Cox, D.D. & Lee, J.S. Pointwise testing with functional data using the Westfall–Young randomization method. *Biometrika* **95**, 621-634 (2008).
88. Exome Aggregation Consortium. ExAC browser. (accessed: 04/02/2016).
89. Batada, N.N., Hurst, L.D. & Tyers, M. Evolutionary and Physiological Importance of Hub Proteins. *PLoS Comput Biol* **2**, e88 (2006).
90. Parikshak, N.N., Gandal, M.J. & Geschwind, D.H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet* **16**, 441-458 (2015).
91. Pocklington, A.J., Cumiskey, M., Armstrong, J.D. & Grant, S.G.N. The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. *Molecular Systems Biology* **2**, 2006.0023 (2006).
92. Fernández, E. *et al.* Targeted tandem affinity purification of PSD - 95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Molecular systems biology* **5**, 269 (2009).
93. Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* **17**, 142-153 (2012).
94. Pers, T.H. *et al.* Comprehensive analysis of schizophrenia-associated loci highlights ion channel pathways and biologically plausible candidate causal genes. *Human Molecular Genetics* **25**, 1247-1254 (2016).



95. Consortium, T.G.O. Gene Ontology Consortium: going forward. *Nucleic Acids Research* **43**, D1049-D1056 (2015).
96. Fabregat, A. *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Research* **44**, D481-D487 (2016).
97. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* **44**, D457-D462 (2016).
98. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research* **43**, D789-D798 (2015).
99. de Leeuw, C.A., Neale, B.M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat Rev Genet* **17**, 353–364 (2016).
100. O'Leary, N.A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**, D733-D745 (2016).
101. van Dongen, J. & Boomsma, D.I. The evolutionary paradox and the missing heritability of schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **162**, 122-136 (2013).
102. Srinivasan, S. *et al.* Genetic Markers of Human Evolution Are Enriched in Schizophrenia. *Biological Psychiatry* [**in press**](2015).
103. Xu, K., Schadt, E.E., Pollard, K.S., Roussos, P. & Dudley, J.T. Genomic and Network Patterns of Schizophrenia Genetic Variation in Human Evolutionary Accelerated Regions. *Molecular Biology and Evolution* **32**, 1148-1160 (2015).
104. Voight, B.F., Kudravalli, S., Wen, X. & Pritchard, J.K. A Map of Recent Positive Selection in the Human Genome. *PLoS Biol* **4**, e72 (2006).
105. Sabeti, P.C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-918 (2007).
106. Grossman, Sharon R. *et al.* Identifying Recent Adaptations in Large-Scale Genomic Data. *Cell* **152**, 703-713 (2013).
107. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354-357 (2014).
108. Sabeti, P.C. *et al.* Positive Natural Selection in the Human Lineage. *Science* **312**, 1614-1620 (2006).
109. Ronen, R., Udpa, N., Halperin, E. & Bafna, V. Learning Natural Selection from the Site Frequency Spectrum. *Genetics* **195**, 181-193 (2013).
110. Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006-1007 (2014).

111. Cunningham, F. *et al.* Ensembl 2015. *Nucleic acids research* **43**, D662-D669 (2015).
112. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58 (2010).
113. Maclean, C.A., Chue Hong, N.P. & Prendergast, J.G.D. hapbin: An Efficient Program for Performing Haplotype-Based Scans for Positive Selection in Large Genomic Datasets. *Molecular Biology and Evolution* **32**, 3027-3029 (2015).
114. Millar, J.K. *et al.* Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Human Molecular Genetics* **9**, 1415-1423 (2000).
115. Chan, C.S. *et al.* 'Rejuvenation' protects neurons in mouse models of Parkinson's disease. *Nature* **447**, 1081-1086 (2007).
116. Fogel, B.L. *et al.* RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Human Molecular Genetics* **21**, 4171-86 (2012).
117. Shoichet, S.A. *et al.* Haploinsufficiency of novel FOXP1B variants in a patient with severe mental retardation, brain malformations and microcephaly. *Hum Genet* **117**, 536-44 (2005).
118. Bacon, C. *et al.* Brain-specific Foxp1 deletion impairs neuronal development and causes autistic-like behaviour. *Mol Psychiatry* **20**, 632-639 (2015).
119. Mariani, J. *et al.* FOXP1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* **162**, 375-90 (2015).
120. Pattyn, A., Guillemot, F. & Brunet, J.F. Delays in neuronal differentiation in Mash1/Ascl1 mutants. *Dev Biol* **295**, 67-75 (2006).
121. Singh, J., Kour, K. & Jayaram, M.B. Acetylcholinesterase inhibitors for schizophrenia. *Cochrane Database Syst Rev* **1**, CD007967 (2012).
122. Choi, K.-H., Wykes, T. & Kurtz, M.M. Adjunctive pharmacotherapy for cognitive deficits in schizophrenia: meta-analytical investigation of efficacy. *The British Journal of Psychiatry* **203**, 172-178 (2013).
123. Stefansson, H. *et al.* Neuregulin 1 and Susceptibility to Schizophrenia. *The American Journal of Human Genetics* **71**, 877-892 (2002).
124. Carvill, G.L. *et al.* Mutations in the GABA Transporter SLC6A1 Cause Epilepsy with Myoclonic-Atonic Seizures. *American Journal of Human Genetics* **96**, 808-15 (2015).
125. Dikow, N. *et al.* 3p25.3 microdeletion of GABA transporters SLC6A1 and SLC6A11 results in intellectual disability, epilepsy and stereotypic behavior. *American journal of medical genetics. Part A* **164A**, 3061-8 (2014).

## ACKNOWLEDGEMENTS

### *General*

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n° 279227 (CRESTAR Consortium; <http://www.crestar-project.eu/>). The work at Cardiff University was funded by Medical Research Council (MRC) Centre (MR/L010305/1), Program Grant (G0800509) and the European Community's Seventh Framework Programme HEALTH-F2-2010-241909 (Project EU-GEI). U.D. received funding from the German Research Foundation (DFG; FOR 2107 grant no. 1151/5-1).

### *Case data*

We thank the participants and clinicians who took part in the CardiffCOGS study. For the CLOZUK2 sample we thank Leyden Delta for supporting the sample collection, anonymisation and data preparation (particularly Marinka Helthuis, John Jansen, Karel Jollie and Anouschka Colson), Magna Laboratories, UK (Andy Walker) and, for CLOZUK1, Novartis and The Doctor's Laboratory staff for their guidance and cooperation. We acknowledge Lesley Bates, Catherine Bresner and Lucinda Hopkins, at Cardiff University, for laboratory sample management. We acknowledge Wayne Lawrence and Mark Einon, at Cardiff University, for support with the use and setup of computational infrastructures.

### *Control data*

A full list of the investigators who contributed to the generation of the Wellcome Trust Case Control Consortium (WTCCC) data is available from [www.wtccc.org.uk](http://www.wtccc.org.uk). Funding for the

project was provided by the Wellcome Trust (WT) under award 076113. The UK10K project was funded by the Wellcome Trust award WT091310. Venous blood collection for the 1958 Birth Cohort (NCDS) was funded by the UK's Medical Research Council (MRC) grant G0000934, peripheral blood lymphocyte preparation by Juvenile Diabetes Research Foundation (JDRF) and WT and the cell-line production, DNA extraction and processing by WT grant 06854/Z/02/Z. Genotyping was supported by WT (083270) and the European Union (EU; ENGAGE: HEALTH-F4-2007- 201413). The UK Blood Services Common Controls (UKBS-CC collection) was funded by WT (076113/C/04/Z) and by the National Institute for Health Research (NIHR) programme grant to NHS Blood and Transplant authority (NHSBT; RP-PG-0310-1002). NHSBT also made possible the recruitment of the Cardiff Controls, from participants who provided informed consent. Generation Scotland (GS) received core funding from the Chief Scientist Office of the Scottish Government Health Directorates CZD/16/6 and the Scottish Funding Council HR03006. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the WT Clinical Research Facility, Edinburgh, Scotland and was funded by the MRC. The Type 1 Diabetes Genetics Consortium (T1DGC; EGA dataset EGAS00000000038) is a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and JDRF. The People of the British Isles project (POBI; <http://www.peopleofthebritishisles.org>) is supported by WT (088262/Z/09/Z). TwinsUK is funded by WT, MRC, EU, NIHR-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. Funding for the QIMR samples was provided by the Australian National Health and Medical Research Council (241944, 339462, 389875,

389891, 389892, 389927, 389938, 442915, 442981, 496675, 496739, 552485, 552498, 613602, 613608, 613674, 619667), the Australian Research Council (FT0991360, FT0991022), the FP-5 GenomeEUtwin Project (QLG2-CT- 2002-01254) and the US National Institutes of Health (NIH; AA07535, AA10248, AA13320, AA13321, AA13326, AA14041, MH66206, DA12854, DA019951), and the Center for Inherited Disease Research (Baltimore, MD, USA). TEDS is supported by a program grant from the MRC (G0901245-G0500079), with additional support from the NIH (HD044454; HD059215). In the GERAD1 Consortium, Cardiff University was supported by WT, MRC, Alzheimer's Research UK (ARUK) and the Welsh Government. Kings College London acknowledges support from the MRC. The University of Belfast acknowledges support from ARUK, Alzheimer's Society, Ulster Garden Villages, N.Ireland R&D Office and the Royal College of Physicians/Dunhill Medical Trust. Washington University was funded by NIH grants, Barnes Jewish Foundation and the Charles and Joanne Knight Alzheimer's Research Initiative. The Bonn group was supported by the German Federal Ministry of Education and Research (BMBF), Competence Network Dementia and Competence Network Degenerative Dementia, and by the Alfried Krupp von Bohlen und Halbach-Stiftung.

### ***CRESTAR Consortium Members***

Evanthia Achilla<sup>1</sup>, Esben Agerbo<sup>2,3</sup>, Cathy L. Barr<sup>4</sup>, Theresa Wimberly Böttger<sup>2</sup>, Gerome Breen<sup>5,6</sup>, Dan Cohen<sup>7</sup>, David A. Collier<sup>5,8</sup>, Sarah Curran<sup>9,10</sup>, Emma Dempster<sup>11</sup>, Danai Dima<sup>5</sup>, Ramon Sabes-Figuera<sup>1</sup>, Robert J. Flanagan<sup>12</sup>, Sophia Frangou<sup>13</sup>, Josef Frank<sup>14</sup>, Christiane Gasse<sup>2,3</sup>, Fiona Gaughran<sup>15</sup>, Ina Giegling<sup>16</sup>, Jakob Grove<sup>18,19</sup>, Eilis Hannon<sup>11</sup>, Annette M. Hartmann<sup>16</sup>, Barbara HeiBerer<sup>19</sup>, Marinka Helthuis<sup>20</sup>, Henriette Thisted Horsdal<sup>2</sup>, Oddur Ingimarsson<sup>21</sup>, Karel Jollie<sup>20</sup>, James L. Kennedy<sup>22</sup>, Ole Köhler<sup>23</sup>, Bettina Konte<sup>16</sup>, Maren Lang<sup>14</sup>, Sophie Legge<sup>24</sup>, Cathryn Lewis<sup>5</sup>, James MacCabe<sup>15</sup>, Anil K. Malhotra<sup>25</sup>, Paul

McCrone<sup>1</sup>, Sandra M. Meier<sup>2</sup>, Jonathan Mill<sup>5,11</sup>, Ole Mors<sup>23</sup>, Preben Bo Mortensen<sup>2</sup>, Markus M. Nöthen<sup>26</sup>, Michael C. O'Donovan<sup>24</sup>, Michael J. Owen<sup>24</sup>, Antonio F. Pardiñas<sup>24</sup>, Carsten B. Pedersen<sup>2,3</sup>, Marcella Rietschel<sup>14</sup>, Dan Rujescu<sup>16</sup>, Ameli Schwalber<sup>19</sup>, Engilbert Sigurdsson<sup>21</sup>, Holger J. Sørensen<sup>27</sup>, Benjamin Spencer<sup>28</sup>, Hreinn Stefansson<sup>29</sup>, Henrik Støvring<sup>3</sup>, Jana Strohmaier<sup>14</sup>, Patrick Sullivan<sup>30,31</sup>, Evangelos Vassos<sup>5</sup>, Moira Verbelen<sup>5</sup>, James T. R. Walters<sup>24</sup>, Thomas Werge<sup>27</sup>.

<sup>1</sup>Centre for Economics of Mental and Physical Health, Health Service and Population Research Department, Institute of Psychiatry, King's College London, London, UK.

<sup>2</sup>National Centre for Register-Based Research, Department of Economics and Business, School of Business and Social Sciences, Aarhus University, Aarhus, Denmark. <sup>3</sup>Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark.

<sup>4</sup>Toronto Western Research Institute, University Health Network, Toronto, Canada. <sup>5</sup>MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. <sup>6</sup>NIHR Biomedical Research Centre for Mental Health, Maudsley Hospital and Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. <sup>7</sup>Department of Community Mental Health, Mental Health Organization North-Holland North, Heerhugowaard, The Netherlands.

<sup>8</sup>Discovery Neuroscience Research, Eli Lilly and Company Ltd, Lilly Research Laboratories, Erl Wood Manor, Surrey, UK. <sup>9</sup>Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. <sup>10</sup>Brighton and Sussex Medical School, University of Sussex, Brighton, UK. <sup>11</sup>University of Exeter Medical School, RILD, University of Exeter, Exeter, UK. <sup>12</sup>Toxicology Unit, Department of Clinical Biochemistry, King's College Hospital NHS Foundation Trust, London, UK.

<sup>13</sup>Clinical Neurosciences Studies Center, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, USA. <sup>14</sup>Department of Genetic Epidemiology in

Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany. <sup>15</sup>Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. <sup>16</sup>Department of Psychiatry, University of Halle, Halle, Germany. <sup>17</sup>Department of Biomedicine, Aarhus University, Denmark. <sup>18</sup>Bioinformatics Research Centre (BiRC), Aarhus University, Aarhus, Denmark. <sup>19</sup>Concentris Research Management GmbH, Fürstfeldbruck, Germany. <sup>20</sup>Leyden Delta B.V., Nijmegen, The Netherlands. <sup>21</sup>Department of Psychiatry, Landspítali University Hospital, Reykjavik, Iceland. <sup>22</sup>Centre for Addiction and Mental Health, Toronto, Ontario, Canada. <sup>23</sup>Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark. <sup>24</sup>MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK. <sup>25</sup>Division of Psychiatry Research, The Zucker Hillside Hospital, Northwell Health System, Glen Oaks, New York, USA. <sup>26</sup>Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany. <sup>27</sup>Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Denmark. <sup>28</sup>Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. <sup>29</sup>deCODE Genetics, Reykjavik, Iceland. <sup>30</sup>Center for Psychiatric Genomics, Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. <sup>31</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

### ***GERAD1 Consortium Members***

Denise Harold<sup>1,9</sup>, Rebecca Sims<sup>1</sup>, Amy Gerrish<sup>1</sup>, Jade Chapman<sup>1</sup>, Valentina Escott-Price<sup>1</sup>, Richard Abraham<sup>1</sup>, Paul Hollingworth<sup>1</sup>, Jaspreet Pahwa<sup>1</sup>, Nicola Denning<sup>1</sup>, Charlene Thomas<sup>1</sup>, Sarah Taylor<sup>1</sup>, John Powell<sup>2</sup>, Petroula Proitsi<sup>2</sup>, Michelle Lupton<sup>2</sup>, Simon

Lovestone<sup>2,10</sup>, Peter Passmore<sup>3</sup>, David Craig<sup>3</sup>, Bernadette McGuinness<sup>3</sup>, Janet Johnston<sup>3</sup>, Stephen Todd<sup>3</sup>, Wolfgang Maier<sup>4</sup>, Frank Jessen<sup>4</sup>, Reiner Heun<sup>4</sup>, Britta Schurmann<sup>4,5</sup>, Alfredo Ramirez<sup>4</sup>, Tim Becker<sup>6</sup>, Christine Herold<sup>6</sup>, André Lacour<sup>6</sup>, Dmitriy Drichel<sup>6</sup>, Markus Nothen<sup>7</sup>, Alison Goate<sup>8</sup>, Carlos Cruchaga<sup>8</sup>, Petra Nowotny<sup>8</sup>, John C. Morris<sup>8</sup>, Kevin Mayo<sup>8</sup>, Peter Holmans<sup>1</sup>, Michael O'Donovan<sup>1</sup>, Michael Owen<sup>1</sup> and Julie Williams<sup>1</sup>.

<sup>1</sup>Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Neurosciences and Mental Health Research Institute, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, UK. <sup>2</sup>King's College London, Institute of Psychiatry, Department of Neuroscience, De Crespigny Park, Denmark Hill, London. <sup>3</sup>Ageing Group, Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, UK. <sup>4</sup>Department of Psychiatry, University of Bonn, Sigmund-Freud-Straße 25, 53105 Bonn, Germany. <sup>5</sup>Institute for Molecular Psychiatry, University of Bonn, Bonn, Germany. <sup>6</sup>Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn. <sup>7</sup>Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany. <sup>8</sup>Departments of Psychiatry, Neurology and Genetics, Washington University School of Medicine, St Louis, MO 63110, US. <sup>9</sup>Present address: Neuropsychiatric Genetics Group, Department of Psychiatry, Trinity Centre for Health Sciences, St James's Hospital, Dublin. <sup>10</sup>Present address: Department of Psychiatry, University of Oxford.

### **Competing financial interests**

D. A. C. is a full-time employee and stockholder of Eli Lilly and Company. The remaining authors declare no conflicts of interest.