bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Pathways analyses of schizophrenia GWAS focusing on known and novel drug targets

Authors: H. A. Gaspar[1,2] and G. Breen[1,2*]

Affiliations:

[1] King's College London, Institute of Psychiatry, Psychology and Neuroscience, MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, UK

[2] National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, UK

*Email: gerome.breen@kcl.ac.uk

Abstract

Schizophrenia is a common, heritable and highly complex psychiatric disorder for which genome-wide association studies (GWASs) have discovered >100 loci. This, and the progress being made in other complex disorders, leads to the questions of how efficiently GWAS can be used to identify novel drug targets and druggable pathways. Taking a series of increasingly better powered GWASs for schizophrenia, we analyse genetic data using information about drug targets and drug therapeutical classes to assess the potential utility of GWAS for drug discovery. As sample size increases, schizophrenia GWAS results show increasing enrichment for known antipsychotic drugs, psycholeptics, and antiepileptics. Drugs targeting calcium channels or nicotinic acetylcholine receptors also show significant association. We conclude that current schizophrenia GWAS results may hold potential therapeutic leads given their power to detect existing treatments.

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Genome-wide association studies (GWAS) have been performed on numerous human disorders and traits [1], uncovering thousands of associations between disorders or quantitative phenotypes and common genetic variants, usually single nucleotide polymorphisms (SNPs), that 'tag' or identify specific genetic loci. Summary statistics from hundreds of GWASs are freely available online, including those from the Psychiatric Genetics Consortium (PGC), which regularly releases GWAS summary statistics for major psychiatric disorders. These include schizophrenia, a complex disorder with a lifetime prevalence of ~1%, significant environmental risk factors, and a heritability of 65%-85%[2] that has been suggested to be highly polygenic in nature [3]. As with other complex genetic disorders, the application of GWAS to schizophrenia has identified multiple disease susceptibility loci. In 2014, over 100 robustly associated loci were identified in a GWAS meta-analysis by the PGC [4]. Similar progress is underway in other psychiatric disorders, with new GWAS reports expected for attention deficit hyperactivity disorder, autism, major depressive disorder, anorexia nervosa, and bipolar disorder in the next year. A key question is how the emergence of new and well powered GWAS data will inform the development of new therapeutics.

Most attention on the therapeutic utility of GWAS has focused on the identification of individual drug targets [5]. Nelson et al. recently demonstrated an increase in the proportion of drug mechanisms with genetic support from 2.0% at the preclinical stage to 8.2% among approved drugs [6]. Results from genetic studies can also be harnessed for repurposing, which aims to find new indications for known drugs [7–9]. Recent studies have also shown how pathway analysis on GWAS data could help discover new drugs for schizophrenia [10,11]; however, these studies, as well as studies focused on single genes or targets, have generally lacked a validation step to show if a GWAS has sufficient power to reliably identify known drugs: a crucial indication that would lend confidence to the discovery of novel drug associations in GWAS data.

Mining of data available on drug-gene interactions (**Fig. 1**) allows the combination of individual drug targets into "drug pathways" represented by sets of genes that encode all targets of a given drug or potential novel therapeutic. Any drug can be represented by such a gene-set derived from its drug activity profile, and assigned a p-value generated by pathway analysis assessing the association of a given drug gene-set with the phenotype. An enrichment curve can be drawn for

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

any particular group of drugs using the entire dataset of drugs ranked by p-value. The associated area under the enrichment curve (AUC) provides a simple way to assess the enrichment of any class of drug for a specific disorder.

In this article, we use MAGMA [12] for pathway analysis, accounting for confounders like linkage disequilibrium, gene size, and gene density, to generate p-values for drugs using the latest schizophrenia GWAS from the PGC (SCZ-PGC2) [4]. Drug-gene interactions from the Drug-Gene Interaction database (DGIdb) [13] and the Psychoactive Drug Screening Program K_i DB [14] were used to assess the enrichment of schizophrenia drug classes. We computed the AUCs using three successively larger schizophrenia studies from the PGC Schizophrenia working group: SCZ-PGC1 [15], SCZ-PGC1+SWE [16], and SCZ-PGC2 [4]. After testing this, we proceeded to analyse the SCZ-PGC2 GWAS for the associations of druggable genes, druggable gene families and known biological pathways with schizophrenia.

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS

Drug classes enrichment: Comparing different drug classes

The enrichment of several ATC (Anatomical Therapeutic Chemical) drug classes in the latest schizophrenia GWAS (SCZ-PGC2) is reported in **Fig. 2a**. The enrichment is assessed using the AUC, where AUC = 100% indicates optimal enrichment and AUC = 50% a random result. AUC p-values were computed using Wilcoxon-Mann-Whitney's test and converted to false discovery rate (FDR) adjusted p-values or *q-values* to account for multiple testing. Psycholeptics (ATC code N05), which include antipsychotics, are significantly enriched (AUC = 70%, q-value = 6.66×10^{-11}) as well as antiepileptics (AUC = 71%, q-value = 0.002). There is also a weak enrichment of immunosuppressants (AUC = 62%) and anesthetics (63%); however, the corresponding q-values are not significant (q-value = 0.051 and 0.063).

Effect of sample size on therapeutic drug class enrichment

Antipsychotics enrichment curves were generated for SCZ-PGC1, SCZ-PGC1+SWE and SCZ-PGC2 (**Fig. 2b**), using only SNPs present in all three studies (whereas results in **Fig. 2a** use all SNPs present in SCZ-PGC2). In the ATC system, antipsychotics (code N05A) are a subset of the psycholeptics class (N05). The p-values associated to the AUC were not corrected for multiple testing, since only three planned comparisons were made. For SCZ-PGC1, the antipsychotics enrichment is equal to a random result (AUC = 50%, p = 0.516); the enrichment is moderate for SCZ-PGC1+SWE (64%, $p = 2.27\times10^{-4}$), and high (82%, $p = 2.31\times10^{-15}$) for SCZ-PGC2. As the sample size used in schizophrenia GWAS increases (and consequently the statistical power), so does the enrichment for antipsychotics.

Druggable genes, gene families and pathways

An analysis of druggable genes, gene families, biological pathways and drugs was conducted using SCZ-PGC2. A druggable gene Manhattan plot is presented in **Supplementary Fig. 1a**. We define druggable genes as genes with known drug interactions and genetic variations. After applying a Bonferroni correction, 124 out of 3048 are significant for schizophrenia, and 403 have an FDR q-value < 5%. Among significant genes, several are related to the major histocompatibility complex (MHC), calcium voltage-gated channels, potassium channels and cholinergic receptors (**Supplementary Fig. 1b**).

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The gene families of the main targets of antipsychotic drugs are shown in **Fig. 3a** - most antipsychotics target dopamine and G protein-coupled 5-hydroxytryptamine (5-HT) receptors. The top druggable gene families, drugs and biological pathways from SCZ-PGC2 pathway analysis are shown in **Fig. 3b**, **3c,** and **3d**, respectively (associated data in **Supplementary Tables 3-5**); entities with a q-value < 5% are considered significant. Most of the top ranked drugs (**Fig. 3c**) for SCZ-PGC2 are calcium channel blockers and fourteen drugs exceed the significance threshold. The top ranked biological pathways (**Fig. 3d**) are *voltage-gated calcium channel complex* (from Gene Ontology or "GO"), *gated channel activity* (Reactome) and *neuronal system* (GO); however, they are not significant. The top-ranked druggable gene families (**Fig. 3b**) are *C1-set domain containing*, *calcium voltage-gated channel subunits* (CACN) and *cholinergic receptors nicotinic subunits* (CHRN). The *C1-set domain containing* family includes many MHC genes; although MHC results can be inflated due to high gene density and LD, the method chosen for the analysis (MAGMA) directly corrects for these confounders.

Detailed associations for the top gene families are given in **Supplementary Fig. 2**. Among CHRN genes, controlling for LD, the cluster of genes *CHRNA3-CHRNA5-CHRNB4* as well as *CHRNA4* are strongly associated with schizophrenia. All together seven CACN genes show a significant association with schizophrenia, with *CACNA1I*, *CACNA1C* and *CACNB2* most highly ranked. *DRD2* and *HTR5A* are the antipsychotic target genes with the strongest association amongst the dopamine and 5-HT receptors families. The top drugs targeting CACN receptors (with at least five targets) are nitrendipine and felodipine; the top drugs targeting CHRN receptors are varenicline and galantamine (cf. **Supplementary Table 6** for a complete list). Potassium voltage-gated channels and cholinergic receptors are also targets of antipsychotics and are enriched in our analysis; however, most cholinergic receptors targeted by antipsychotics are muscarinic. Complete output of gene-wise and gene-set wise analyses can be found in excel format in **Supplementary Tables 7-17**.

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DISCUSSION

We find that the targets of antipsychotics, the primary drug class used to treat schizophrenia, are enriched for association in current schizophrenia GWAS results. We also show that the enrichment for known antipsychotics increases with the number schizophrenia cases included in the GWAS, the largest being SCZ-PGC2 (PGC Schizophrenia working group phase 2). In addition, our results show significant enrichment for two broad drug classes: psycholeptics and antiepileptics. Some antiepileptics have been investigated for treatment-resistant schizophrenia and may have GABAergic and antiglutamatergic action [17]. These results suggest that, in schizophrenia (and other complex disorders), current well-powered GWAS results hold potential new therapeutic leads given their power to detect existing treatments.

We also demonstrated that both calcium channel drugs and nicotinic acetylcholine receptor drugs show significant association in PGC2-SCZ. The top drugs are verapamil and cinnarizine, two calcium channel blockers. Verapamil has been reported to be as efficient as lithium for the treatment of mania [18]. Cinnarizine, which has atypical antipsychotic properties in animal models [19], is prescribed for vertigo because of its antihistamine properties and is also an antagonist of dopamine D2 receptors. Varenicline and galantamine are the two top ranked drugs in our analysis that target nicotinic acetylcholine receptors. Varenicline is a nicotinic agonist used for smoking cessation [20] while galantamine is an allosteric modulator of nicotinic receptors and an acetylcholinesterase inhibitor, and has been investigated for the treatment of cognitive impairment in schizophrenia [21].

We also tested the association of known biological pathways and known druggable gene families. The top biological pathways in PGC2-SCZ are consistent with previous knowledge of schizophrenia and mirror our drug pathway results. Voltage-gated channels have been widely studied in psychiatric disorders [22], and L-type calcium channels have been associated with schizophrenia in numerous studies [23]. For druggable gene families, our results show the strongest association signal for calcium voltage-gated channels, and a weaker signal for nicotinic acetylcholine receptors. However, individually, the nicotinic receptor *CHRNA4* as well as the *CHRNA3-CHRNA5-CHRNB4* cluster are strongly associated with schizophrenia. *CHRNA4* encodes the α4 subunit found in the α4β2 receptors, which are widely expressed in the brain, including the thalamus, brainstem, and cerebellum, and are particularly sensitive to nicotine [24].

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The *CHRNA3-CHRNA5-CHRNB4* cluster consists of genes in high LD with each other and has been linked to nicotine dependence [25]. Some studies indicate that nicotine could have a positive effect on psychotic symptoms and cognitive function in schizophrenic patients [26]; the expression of both the α4β2 and α7 nicotinic receptor subunits has been reported to be altered in the brains of patients in post-mortem binding studies [27]. Our results are consistent with a recent study by Won et al. that also highlights the enrichment of acetylcholine receptor activity in schizophrenia [28].

Our analyses primarily focused on drugs with multiple targets (>5). To assess drugs with fewer targets, we specifically tested the association of individual genes within the druggable genome. Overall, 403 druggable genes have an FDR q-value < 5%. Many of these are clustered in the high LD region of the major histocompatibility complex (MHC) and are thus difficult to interpret. However, we also see significance with individual gene loci encoding calcium voltage-gated channels, potassium channels and cholinergic receptors (**Supplementary Fig. 1b**) and multiple novel genes. The top druggable genes outside the MHC region and with 10 to 100 associated ligands are *DPYD*, *CACNA1I*, *CACNA1C*, *CACNB2*, *CHRNA3*, *AKT3*, *NOS1*, *MCHR1*, *CYP2D6*, and *DPP4*. The associations with metabolic enzymes such as *CYP2D6*, in which variability may influence antipsychotic plasma levels, is difficult to interpret as the large number of treatment resistant cases included in the PGC2-SCZ GWAS may influence the results; however, recent studies suggest that *CYP2D6* is not associated with treatment resistant cases [29]. Compounds targeting proteins encoded by *MCHR1* and *DPP4* are of particular interest. *MCHR1* antagonists include high affinity ligands such as ATC0175 or ATC0065, which exhibit antidepressive and anxiolytic effects in mouse and rat behavioral models [30]. *DPP4* inhibitors include gliptins (dutogliptin, alogliptin, etc.) that are used to treat type 2 diabetes, and atorvastatin, which is prescribed for its cholesterol-lowering properties [31]. Current antipsychotics can induce insulin resistance [32], and drugs which do not or would reverse these effects would be a welcome addition to the pharmacopoeia.

In summary, our approach may be used to validate the power of a given GWAS and to identify new drug targets. This approach is primarily a way to generate new therapeutic hypotheses from (hypothesis-neutral) polygenic genetic data. It is suitable for use as a filtering process in the first

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

stages of drug discovery, with detailed target qualification analyses and validation experiments necessary for individual genes and molecules. We conclude that sufficiently powerful GWASs can be examined with increased confidence for drug target identification and repurposing opportunities across complex disorders, by investigating the ranking of biological pathways, drug gene-sets and druggable genes. In disorders that have few known drug treatments, such as eating disorders and obesity, our validation step may be impossible, but once well-powered GWASs with multiple significant signals are available this approach could still be effective to generate much needed therapeutic hypotheses.

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ONLINE METHODS

Methods: Pathway analysis in MAGMA

The pathway analysis software MAGMA v. 1.03 [12] was used to generate p-values for genes and gene-sets representing drugs, gene families and biological pathways. GWAS summary statistics are available as SNP p-values, which MAGMA combines to produce gene and gene-set p-values. Brown's method, implemented in MAGMA, is an extension of Fisher's method that combines dependent SNP p-values into a single gene p-value [33] using information on SNP correlations. These gene p-values are converted to Z-values, which are used as the response variable in a regression model, solved using a generalized least squares approach accounting for LD. Two types of regression analyses can be conducted: self-contained or competitive. The self-contained approach tests whether the pathway is associated to a trait of interest, whereas the competitive approach tests whether genes in the pathway are more strongly associated than genes outside the pathway. The self-contained approach is more powerful, but it is sensitive to the polygenic nature of observed GWAS statistics inflation and may lead to a higher Type I error [12,34]. Therefore, we used competitive p-values. In MAGMA, the competitive analysis corrects for gene size and density and takes into account gene-gene correlations such as those observed in gene clusters. The SNP positions and frequencies were extracted from the European subset of 1000 genomes phase III v.5a [35] with genome assembly hg19. We used Ensembl release 75 [36] for the gene positions. The gene window was set to 35 kb upstream and 10 kb downstream in MAGMA to include gene regulatory regions. We generated FDR-adjusted p-values or *q-values* for genes and gene-sets, using Benjamini and Hochberg's method to account for multiple testing [37].

Methods: Enrichment measure for groups of gene-sets

Pathway analysis approaches generate gene-set p-values for a trait of interest taking into account LD and other confounders. Instead of investigating individual gene-sets, we focused on *groups* of gene-sets. For example, a class of drugs can be represented by a group **G** of drugs (gene-sets). To determine whether **G** is significantly enriched, we can draw enrichment curves, widely used in virtual screening [38]. The curves display the percentage of hits found when decreasing the value of a scoring function. Here, the scoring function is the gene-set association with the trait of interest in $-\log_{10}$(p-value) units, and the hits are the gene-sets in **G**. The area

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

under this enrichment curve (AUC) provides a quantitative assessment of the enrichment of **G** in a GWAS and is computed using the trapezoidal approximation of an integral. The expected random result is $= 50\%$ and the maximum value is $= 100\%$.

The AUC significance was assessed using Wilcoxon-Mann-Whitney test (WMW), which tests whether the data distribution is the same within two different groups (e.g., gene-sets in **G** and not in **G**) [39] - also, the AUC could be directly calculated from the Wilcoxon-Mann-Whitney U statistic [40].

Materials: Schizophrenia GWAS summary statistics

In this paper, we used three GWASs conducted in 2011 [15], 2013 [16] and 2014 [4] with increasing sample sizes (cf. **Fig. 2b** and **Supplementary Table 1**). The three studies were coined SCZ-PGC1, SCZ-PGC1+SWE and SCZ-PGC2, respectively. The three studies mainly contain individuals of European ancestry [4,15,16]; SCZ-PGC2 is the only study including the X chromosome and individuals with East Asian ancestry. Only SNPs present in the European subset of 1000 genomes phase 3 v.5a [35] with minor allele frequency (MAF) ≥ 1% were kept. The genomic inflation factor as well as the LD score intercept were computed for each set using the LDSC software v. 1.0.0 [41]. All p-values were subsequently corrected using the LD score intercept - a score based on linkage disequilibrium that should provide a better way to control for inflation than the genomic inflation factor [42]. Only the 1,123,234 SNPs shared among SCZ-PGC1, SCZ-PGC1+SWE and SCZ-PGC2 were considered when comparing the three studies. The latest and most powerful GWAS (SCZ-PGC2) was used to investigate the enrichment of drug classes, biological pathways and gene families, with an additional filter (MAF ≥ 5%) leaving 5,739,569 SNPs.

Materials: Drug gene-sets

Drug/gene interactions are mainly derived from drug/target activity profiles. The data was drawn from two sources: the Drug-Gene Interaction database DGIdb v.2 [13], and the Psychoactive Drug Screening Database K_i DB [14] downloaded in June 2016. DGIdb is a new resource that integrates drug/gene interactions from 15 databases, amongst which DrugBank and ChEMBL; the data is directly available as drug/gene pairs and genes are identified by their HGNC (HUGO Gene Nomenclature Committee) names [43]. K_i DB provides K_i values for drug/target pairs and is particularly relevant for psychoactive drugs. Only human assays were

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

considered and drug/target pairs were kept if $7 \leq pK_i < 14$. More details on the filtering procedure can be found in **Supplementary Text 1**. Gene-sets were produced by merging both DGIdb and K_i DB drug/gene data and by converting HGNC names to Ensembl release 75 [36] identifiers. The number of unique gene-sets was 3,940 at the end of the filtering process, but only 2736 gene-sets had more than 2 genes and could be used for pathway analysis. On average, 3.8 molecules shared the same gene-set. We annotated groups of drugs using ATC classes, listed in **Figure 2.a** and containing at least 10 drugs. The "validation" set for schizophrenia GWASs is the set of antipsychotics with ATC code N05A - all schizophrenia drugs belong to this class (cf. **Supplementary Table 2** for the list of prescription drugs). Druggable gene families were defined using the HGNC nomenclature downloaded in July 2016 [43] and keeping only the ~3000 genes present in our drug/gene interaction dataset.

Materials: Biological pathways

We refer to both gene ontologies and canonical pathways as "biological pathways". Pathway gene-sets were extracted from MSigDB v5.1 [44], encompassing canonical (CP) and Gene Ontology (GO) gene-sets. MSigDB is a regularly updated resource gathering pathways and ontologies from the main online databases. CP sets were curated from: BioCarta, KEGG, Matrisome, Pathway Interaction Database, Reactome, SigmaAldrich, Signaling Gateway, Signal Transduction KE and SuperArray. Only pathways containing between 10-1000 genes were included, for a total of 2714 gene-sets (1405 GO, 1309 CP). The 10-1000 cut-off was set to limit the number of tested pathways, but also to avoid the case of a single gene driving the association (too few genes) and noisy results (too many genes). These "pathways" provide a practical way to investigate the function of a subnetwork without accounting for the complexity of biological networks.

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Materials

PDF File:

Supplementary Figures 1-2, Supplementary Tables 1–6, and Supplementary Texts 1-2

Excel File:

Supplementary Tables 7-17

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

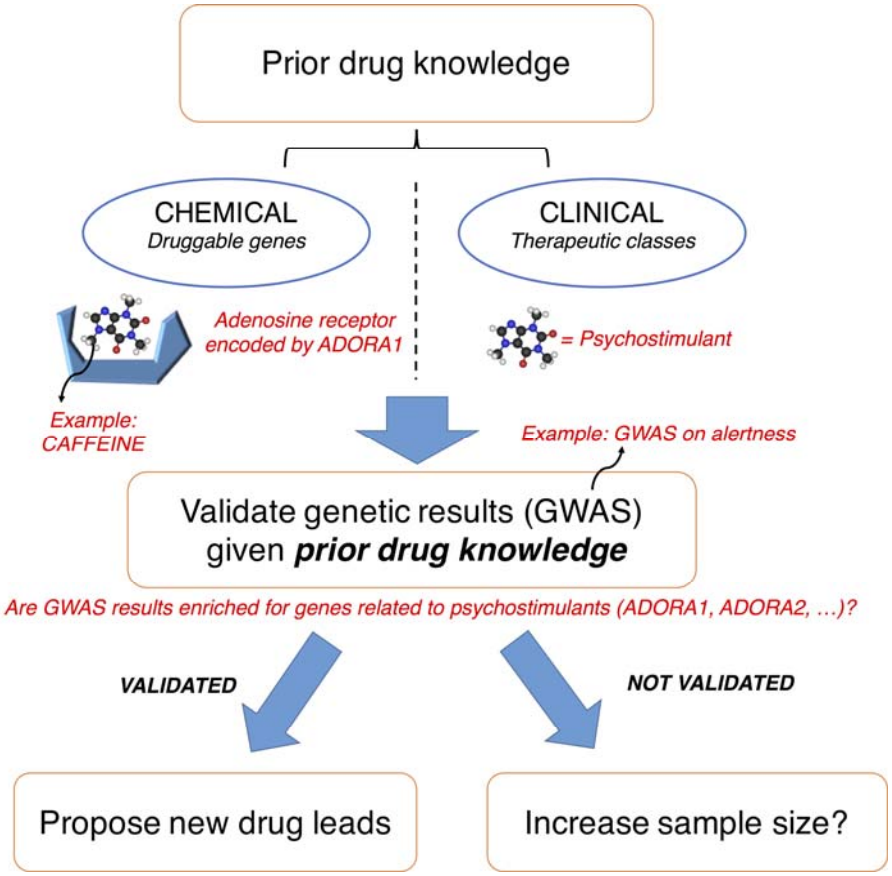
References and Notes:

Acknowledgments: K_i determinations were generously provided by the National Institute of Mental Health's Psychoactive Drug Screening Program, Contract # HHSN-271-2013-00017-C (NIMH PDSP). The NIMH PDSP is Directed by Bryan L. Roth MD, PhD at the University of North Carolina at Chapel Hill and Project Officer Jamie Driscoll at NIMH, Bethesda MD, USA. **Funding**: National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, UK. **Author contributions**: H.A.G. produced the results, G.B. provided advice and guidelines, both H.A.G. and G.B. contributed to the methodology and the writing of the paper. **Competing interests**: There is no conflict of interest. **Data and materials availability**: All data used in this paper are freely available online (cf. references and supplementary materials).

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

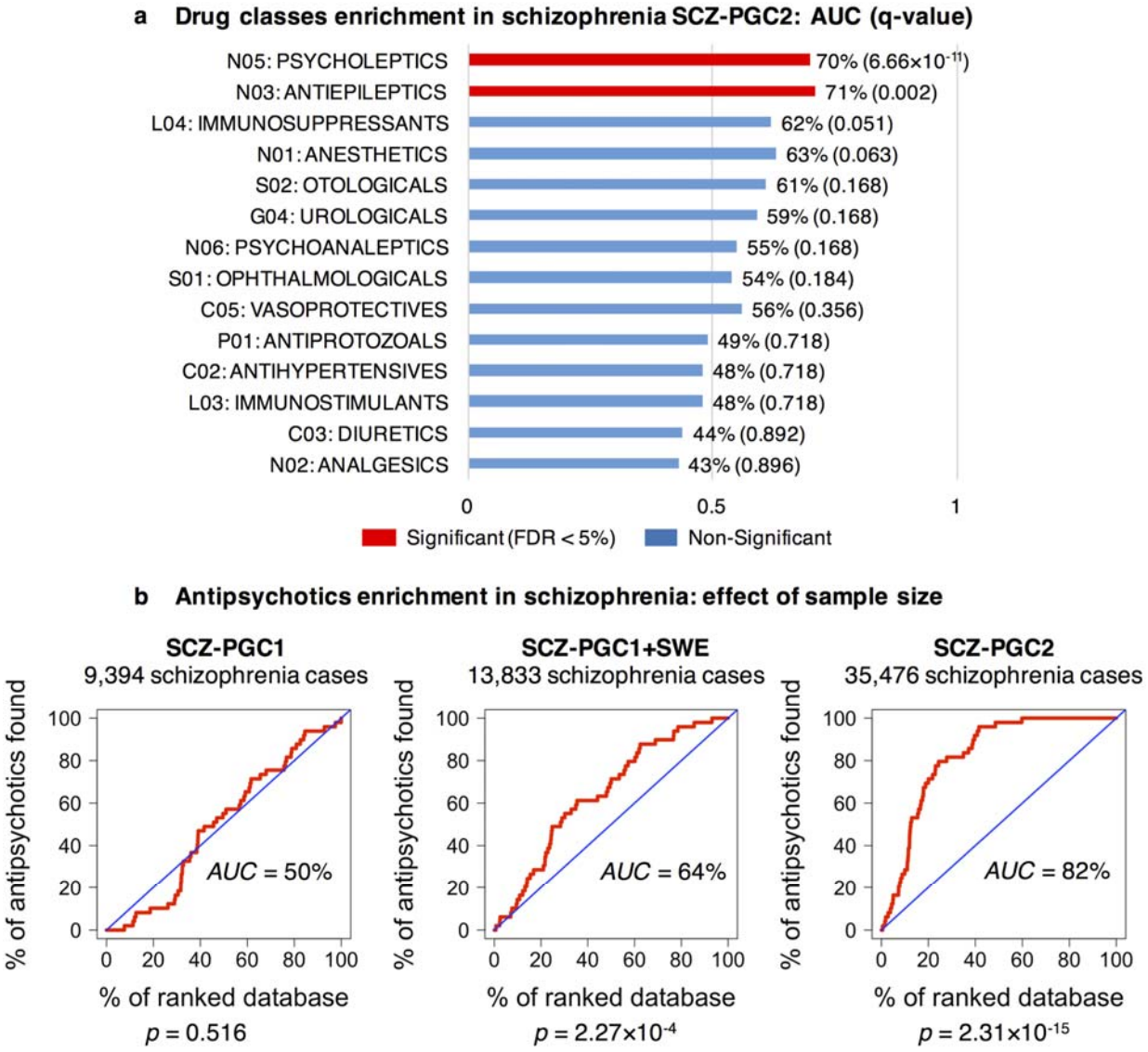
Figures:

Fig. 1. Using drug knowledge to validate genetic results. Drug knowledge, encompassing therapeutic classes and druggable genes (*e.g.*, caffeine is a psychostimulant targeting adenosine receptors), may be used to validate the ability of a GWAS to find known drugs for a given trait (*e.g.*, alertness). Novel targets and potential drugs could then be found in validated genetic results.



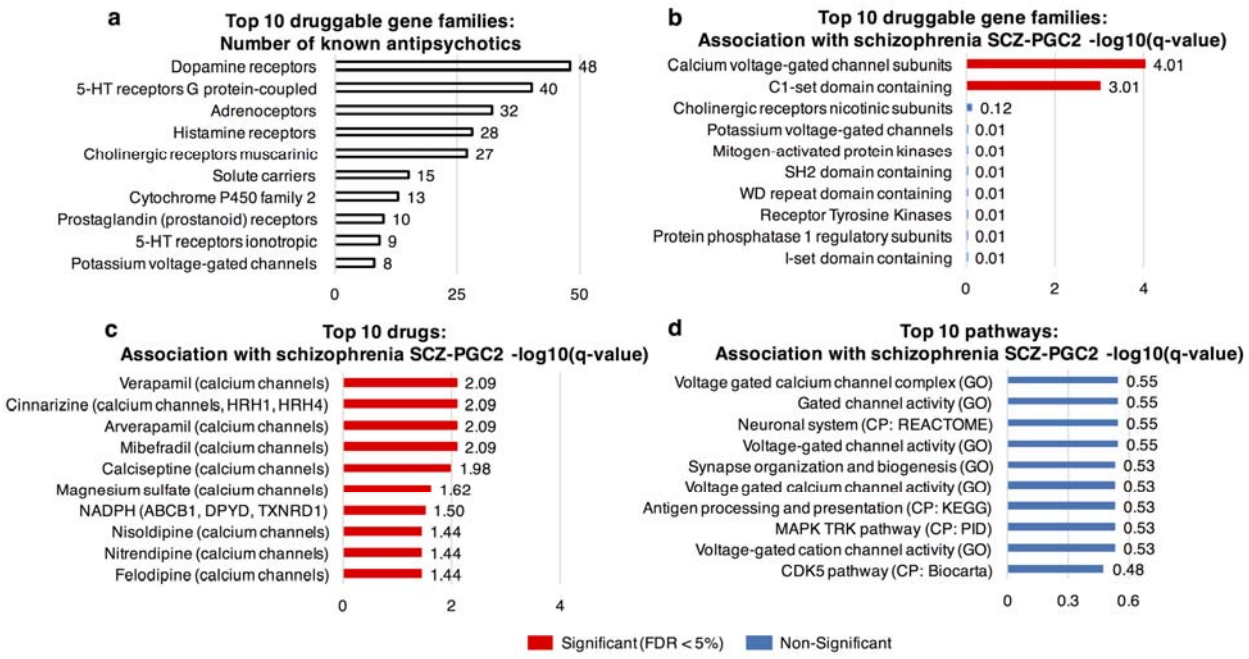
bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 2. **(a)** Enrichment of several ATC drug classes in SCZ-PGC2 GWAS. Significant classes (FDR < 5%) are highlighted; AUC is the area under the curve, with associated FDR-adjusted p-values or *q-values* derived from Wilcoxon-Mann-Whitney's test to account for multiple testing. **(b)** Antipsychotic enrichment in schizophrenia GWASs as a function of sample size. The figure shows enrichment curves in red and corresponding areas under the curve (AUC with p-values *p*) for antipsychotics (ATC code N05A), using three GWASs with increasing sample sizes. The expected "random" enrichment curve is indicated in blue.



bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 3. (**a**) Top 10 druggable gene families in schizophrenia, ranked by the number of antispychotics with interaction data in DGIdb or K_i DB. (**b**) Pathway analysis results: top 10 druggable gene families in PGC-SCZ2, ordered by FDR-adjusted p-values (q-values). (**c**) Top 10 drugs in PGC-SCZ2. (**d**) Top 10 GO and canonical pathways in PGC-SCZ2.



bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90,** 7–24 (2012).
2. Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373,** 234–239 (2009).
3. Gottesman, I. I. & Shields, J. A polygenic theory of schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **58,** 199–205 (1967).
4. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
5. Lencz, T. & Malhotra, A. K. Targeting the schizophrenia genome: a fast track strategy from GWAS to clinic. *Mol. Psychiatry* **20,** 820–826 (2015).
6. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47,** 856–860 (2015).
7. Sanseau, P. *et al.* Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* **30,** 317–320 (2012).
8. Wang, Z.-Y. & Zhang, H.-Y. Rational drug repositioning by medical genetics. *Nat. Biotechnol.* **31,** 1080–1082 (2013).
9. Breen, G. *et al.* Translating genome-wide association findings into new therapeutics for psychiatry. *Nat. Neurosci.* **19,** 1392–1396 (2016).
10. Ruderfer, D. M. *et al.* Polygenic overlap between schizophrenia risk and antipsychotic response: a genomic medicine approach. *Lancet Psychiatry* **3,** 350–357 (2016).
11. de Jong, S., Vidler, L. R., Mokrab, Y., Collier, D. A. & Breen, G. Gene-set analysis based on the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia. *J. Psychopharmacol.* **30,** 826–830 (2016).
12. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of GWAS data. *PLoS Comput. Biol.* **11,** e1004219 (2015).

13. Wagner, A. H. *et al.* DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res.* **44,** D1036–44 (2016).
14. Roth, B. L., Lopez, E., Patel, S. & Kroeze, W. K. The Multiplicity of Serotonin Receptors: Uselessly Diverse Molecules or an Embarrassment of Riches? *Neuroscientist* **6,** 252–262 (2000).
15. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* **43,** 969–976 (2011).
16. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45,** 1150–1159 (2013).
17. Hosák, L. & Libiger, J. Antiepileptic drugs in schizophrenia: a review. *Eur. Psychiatry* **17,** 371–378 (2002).
18. Dubovsky, S. L. & Buzan, R. The Role of Calcium Channel Blockers in the Treatment of Psychiatric Disorders. *CNS Drugs* **4,** 47–57 (1995).
19. Dall'Igna, O. P., Tort, A. B. L., Souza, D. O. & Lara, D. R. Cinnarizine has an atypical antipsychotic profile in animal models of psychosis. *J. Psychopharmacol.* **19,** 342–346 (2005).
20. Evins, A. E. *et al.* Maintenance treatment with varenicline for smoking cessation in patients with schizophrenia and bipolar disorder: a randomized clinical trial. *JAMA* **311,** 145–154 (2014).
21. Buchanan, R. W. *et al.* Galantamine for the treatment of cognitive impairments in people with schizophrenia. *Am. J. Psychiatry* **165,** 82–89 (2008).
22. Imbrici, P., Camerino, D. C. & Tricarico, D. Major channels involved in neuropsychiatric disorders and therapeutic perspectives. *Front. Genet.* **4,** 76 (2013).
23. Berger, S. M. & Bartsch, D. The role of L-type voltage-gated calcium channels Cav1.2 and Cav1.3 in normal and pathological brain function. *Cell Tissue Res.* **357,** 463–476 (2014).
24. Brody, A. L. *et al.* Cigarette Smoking Saturates Brain α4β2 Nicotinic Acetylcholine Receptors. *Arch. Gen. Psychiatry* **63,** 907 (2006).
25. Greenbaum, L., Rigbi, A., Teltsh, O. & Lerer, B. Role of genetic variants in the CHRNA5-

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CHRNA3-CHRNB4 cluster in nicotine dependence risk: importance of gene-environment interplay. *Mol. Psychiatry* **14,** 828–830 (2009).

26. Chen, J. *et al.* Genetic Relationship between Schizophrenia and Nicotine Dependence. *Sci. Rep.* **6,** 25671 (2016).
27. Ripoll, N., Bronnec, M. & Bourin, M. Nicotinic receptors and schizophrenia. *Curr. Med. Res. Opin.* **20,** 1057–1074 (2004).
28. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* (2016). doi:10.1038/nature19847
29. van de Bilt, M. T. *et al.* Cytochrome P450 genotypes are not associated with refractoriness to antipsychotic treatment. *Schizophr. Res.* **168,** 587–588 (2015).
30. Chaki, S. Anxiolytic- and Antidepressant-Like Profile of ATC0065 and ATC0175: Nonpeptidic and Orally Active Melanin-Concentrating Hormone Receptor 1 Antagonists. *J. Pharmacol. Exp. Ther.* **313,** 831–839 (2004).
31. Taldone, T., Zito, S. W. & Talele, T. T. Inhibition of dipeptidyl peptidase-IV (DPP-IV) by atorvastatin. *Bioorg. Med. Chem. Lett.* **18,** 479–484 (2008).
32. Teff, K. L. *et al.* Antipsychotic-induced insulin resistance and postprandial hormonal dysregulation independent of weight gain or psychiatric disease. *Diabetes* **62,** 3232–3240 (2013).
33. Brown, M. B. 400: A Method for Combining Non-Independent, One-Sided Tests of Significance. *Biometrics* **31,** 987–992 (1975).
34. Maciejewski, H. Competitive and self-contained gene set analysis methods applied for class prediction. in *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on* 55–61 (2011).
35. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
36. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res.* **44,** D710–6 (2016).
37. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful

bioRxiv preprint doi: https://doi.org/10.1101/091264; this version posted January 6, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57,** 289–300 (1995).

38. Varin, T., Schuffenhauer, A., Ertl, P. & Renner, S. Mining for bioactive scaffolds with scaffold networks: improved compound set enrichment from primary screening data. *J. Chem. Inf. Model.* **51,** 1528–1538 (2011).
39. Wilcoxon, F. in *Breakthroughs in Statistics* (eds. Kotz, S. & Johnson, N. L.) 196–202 (Springer New York, 1992).
40. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143,** 29–36 (1982).
41. Brendan Bulik-Sullivan, H. F. *LD Score Regression (LDSC)*. (Broad Institute of MIT and Harvard / MIT Department of Mathematics, 2015).
42. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
43. Gray, K. A., Yates, B., Seal, R. L., Wright, M. W. & Bruford, E. A. Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res.* **43,** D1079–85 (2015).
44. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–15550 (2005).