De Jong, S., Vidler, L., Mokrab, Y., Collier, D. A., & Breen, G. D. (Accepted/In press). Gene-set analysis based on the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia.
Gene-set analysis based on the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia

Simone de Jong\textsuperscript{1,2}\textsuperscript{*}, Lewis R Vidler\textsuperscript{3}, Younes Mokrab\textsuperscript{4}, David A Collier\textsuperscript{1,3}, Gerome Breen\textsuperscript{1,2}

\textsuperscript{1} Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College London SE5 8AF
\textsuperscript{2} NIHR Biomedical Research Centre for Mental Health, Maudsley Hospital and Institute of Psychiatry, Psychology & Neuroscience, King’s College London, UK
\textsuperscript{3} Discovery Neuroscience Research, Eli Lilly and Company Ltd, Erl Wood, Windlesham, Surrey, United Kingdom GU20 6PH, UK
\textsuperscript{4} Sidra Medical Research Center, P.O. Box 26999, Doha, Qatar

*To whom correspondence should be addressed: Simone de Jong, sdejongwork@gmail.com

\textbf{Running title:} Genetic analysis of drugable pathways in schizophrenia
Abstract

Genome-wide association studies (GWAS) have identified thousands of novel genetic associations for complex genetic disorders, leading to the identification of potential pharmacological targets for novel drug development. In schizophrenia, 108 conservatively defined loci that meet genome-wide significance have been identified and hundreds of additional sub-threshold associations harbor information on the genetic aetiology of the disorder. In the present study, we used gene-set analysis based on the known binding targets of chemical compounds to identify the ‘drug pathways’ most strongly associated with schizophrenia-associated genes, with the aim of identifying potential drug repositioning opportunities and clues for novel treatment paradigms, especially in multi-target drug development. We compiled 9,389 gene sets (2,496 with unique gene content) and interrogated gene-based $p$-values from the PGC2-SCZ analysis. Although no single drug exceeded experiment wide significance (corrected $p<0.05$), highly ranked gene-sets reaching suggestive significance including the dopamine receptor antagonists Metoclopramide and Trifluoperazine and the tyrosine kinase inhibitor Neratinib. This is a proof of principal analysis showing the potential utility of GWAS data of schizophrenia for the direct identification of candidate drugs and molecules that show polypharmacy.
Introduction

Schizophrenia is a common, debilitating neuropsychiatric disorder with a high unmet medical need. Although efficacious, current antipsychotics are not ideal as treatments because a third of patients show treatment resistance, and approximately another third respond only partially (Smith et al. 2009). There has been no novel pharmacological treatment paradigm for schizophrenia for several decades, and new approaches are urgently needed (Hyman 2012; Nutt and Need 2014; Millan et al. 2015). The dopamine D2 receptor remains the primary target of currently marketed multi-target antipsychotic drugs (Ginovart and Kapur 2012), despite attempts to develop alternative treatment strategies. Most recent approaches to drug development in schizophrenia have focused mainly on single-target drugs, which have so far failed in the clinic (Dunlop and Brandon 2015). There is a growing realization that because binding of multiple receptor targets is characteristic of effective antipsychotics that new therapeutic development should try to develop drugs which also show “polypharmacy”, (Wong et al. 2010) and tools to improve the success of multi-target drug development (MTDD) approaches will be important (Carrieri et al. 2013).

Support for the multi-target approach to drug development in schizophrenia comes from GWAS studies, which support have confirmed that schizophrenia is a complex genetic disorder with a strong polygenic component, from hundreds or even thousands of common variants of small effect, as well as rare moderate-risk variants (Neale and Sklar 2015) Specifically, GWAS meta-analysis in schizophrenia by the Psychiatric Genomics Consortium (PGC) schizophrenia group has identified 128 independent associations spanning that meet genome-wide significance in 108 genomic regions notably including the dopamine D2 receptor locus, but also many other neurotransmitter genes, especially those involved in CNS calcium and glutamate signaling (Ripke et al. 2014). In addition, studies have also utilized sub-GWAS significant associations to generate polygenic risk scores, thereby allowing the profiling of genome-wide risk in individuals and across neuropsychiatric disorders (Purcell et al. 2009), and to identify neuronal, immune and histone biological pathways enriched for association with psychiatric disorders using gene-set analyses (O’Dushlaine 2015; Harrison 2015).
In particular, this latter approach has the potential to drive forward new drug discovery paradigms, beyond the identification of novel individual targets, by identifying new multi-target entities aimed at modulating schizophrenia associated pathways or disease networks. Alternatively, this approach could also be used to find indications for drug repositioning to generate new treatment paradigms for schizophrenia, via matching existing clinical compounds’ binding profiles to targets or sets of targets generated from GWAS or other genetic studies (Sanseau et al. 2012; Grover et al. 2014). If compounds already known to be safe in man for other indications can be repositioned, they could be progressed more quickly and cheaply through human experimental medicine approaches aimed at providing proof-of-concept by clinical validation in small efficacy studies. Effective compounds can then be further refined to develop novel effective medications. In addition it may provide leads for MTDD approaches, which are difficult to prosecute from scratch, because of the difficulty in generating structure-activity relationships for a single compound aimed at multiple targets. Repositioning in this way may provide a head start for MTDD efforts.

The aim of the current study is to identify potential drug repositioning opportunities and small molecule clues by leveraging genetic mega-analysis results for schizophrenia, utilizing knowledge of the pharmacological action of compounds to test ‘drug pathways’ in GWAS data. Formally this is a gene-set analysis (GSA) (Mooney et al. 2014) where instead of using biological pathway information, gene-sets are defined on the basis of pharmacological profiles of chemical compounds, based on their target binding affinities. For each compound a drug gene set (DGS) is composed of the protein targets of that molecule, which meet a selected binding threshold. Compounds which bind significant targets from within the GWAS data and show significant statistical association as a DGS and could be considered potential multi-target drugs for that disorder, with the caveat that GWAS data is agnostic as to the direction of effect of the drug as genetic pathway results lack direction of effect (O’Dushlaine 2015). However, after suitable expert or experimental examination candidate, drugs indicated by this approach drugs then be repositioned to the disorder under investigation if it is clinically approved, or may provide clues to novel treatment strategies in multi-target drug development. We present a proof of principle analysis applying this analytical approach to schizophrenia GWAS data.
Materials and methods

Genome wide filtered (INFO score ≥0.8 & MAF ≥1%) SNP results for schizophrenia were taken from the PGC2 mega analysis (resulting in summary statistics for 7,865,159 variants)(Ripke et al. 2014). SNP p-values were GC corrected prior to pathway analysis (Dadd et al. 2009). Pathways sets were compiled using the ChEMBL database (v18, 2014) (Bento et al. 2014) containing ligands with an activity of better than 100 µM against proteins. We required a minimum of 5 genes per ligand in order to avoid bias in pathway analyses and in line with our aim to investigate multi-drug targets in schizophrenia. This resulted in 9,389 pathways of which 2,496 have a unique gene content. Many of the drugs nevertheless had very similar sets of targets and/or chemical structures. There are many ways to address this but we choose to use spectral decomposition of a 2,496x2,496 matrix of the Tanimoto similarity indices (Kristensen et al. 2011) using the ECFP4 fingerprint (Rogers and Hahn 2010) of the drugs and small molecules. We ascertained the number of independent tests performed as 417.85 for multiple testing corrections (http://neurogenetics.qimrberghofer.edu.au/matSpDlite/matSpDlite.R (Cheverud 2001; Nyholt 2004; Li and Ji 2005) setting the 5% significance threshold to $p=0.0001196$.

We used FORGE to combine p-values per gene and subsequently pathways (Pedroso et al. 2012; Pedroso et al. 2015). In order to combine SNP p-values per gene (Forge.pl) we set a maximum of 1,000,000 permutations and the algorithm was run with a fuzzy border option (5’ 35kb, 3’ 10kb) to attempt to capture gene promoter and 3’ UTR regulatory elements. This results in a Z statistic and corresponding p-value (fixed effects estimate after permutations) per gene adjusted for LD, or the raw SNP p-value in the case of only one SNP being mapped to a gene (Supplementary Table S1). These statistics are subsequently fed into the gsa.pl script to combine p-values of the genes per pathway (Supplementary Table S2). The SNP2Gene mapping file was constructed to contain all protein coding genes in hg19 build with at least one SNP in PGC2 (19,492 genes). To assess LD structure we used the 1000 Genomes Phase 1.v3 data after liftover to hg19 and lightly pruned with a $R^2=0.9$ threshold using Priority Pruner (http://prioritypruner.sourceforge.net), prioritizing significant SNPs within the PGC2 SCZ SNP results. The top resulting pathways were inspected for presence of gene-clusters, and, if
detected those drug gene sets were rerun with the entire gene cluster represented as one gene.

With the aim of validation, the INRICH algorithm (Lee et al. 2012) was also run using a fuzzy border option (5’ 35kb, 3’ 10kb) and PGC2 SNP data clumped according to protocol for the top three pathways. Finally, we compiled a list of 100 psychoactive drugs prescribed for SCZ, MDD, BPD, ADHD or ASD and looked up corresponding FORGE p-values in our pathway results (Supplementary table S3).
Results

Using the FORGE algorithm (Pedroso et al. 2012; Pedroso et al. 2015) we combined p-values per gene for genes with 5’ 35kb, 3’ 10kb borders. Results are given in Table S1. Subsequently we combined p-values for genes per gene-set. This resulted in 1,056 nominally significant pathways (139 with unique gene content) that are given in Table S2. The significance threshold required to set the family wise error rate (FWER) significance threshold for 5% is uncorrected \( p=0.0001196 \) using matrix spectral decomposition to derive the number of independent tests (n=417, see Methods). After correction for gene clusters no single drug exceeded experiment wide significance (corrected \( p<0.05 \)). However, three known schizophrenia drugs showed highly ranked suggestive pathways and their gene content are listed in Table 1. Of these, trifluoperazine and metoclopramide also reach nominal significance in INRICH (\( p=0.02 \)), while neratinib is non-significant in INRICH results (\( p=0.29 \)).

We then examined drugs commonly prescribed for SCZ, MDD, BPD, ADHD and ASD (Supplementary Table S3). We find that 41 of these drugs are represented in our pathway results, with 22 (non-exclusively) prescribed for SCZ. 7 out of 41 drugs reach a \( p \)-value <0.05. Of these, 5 (71%, Fisher’s exact \( p=0.42 \)) are prescribed for SCZ (trifluoperazine, chlorpromazine, fluphenazine, pimozide and aripiprazole (the latter also prescribed for BPD)), 1 for ADHD (bupropion) and 1 for ADHD and MDD (amitriptyline).
Discussion

The aim of the present study is to identify potential drug repositioning opportunities and clues to novel treatment in schizophrenia by interrogating genome-wide association data with gene-sets derived from knowledge of the pharmacological action of drug. The development of new pharmaceutical medicines is a long, complex and expensive process, with about 14 years required to take a new molecular entity to market (Pammolli et al. 2011). Two of the most costly aspects are the identification of novel small molecules with drug-like properties and good pharmacological and safety profiles through iterative medicinal chemistry, and assessment of human safety and efficacy in clinical trials. However the attrition rate of compounds in drug development is high with the predominant reason for failure in clinical trials not toxicity or pharmacology, but lack of efficacy, a result of inadequate target validation: even though the drug acts on the target in the desired way, it does not alter the disease process (Bunnage et al. 2015). Thus there is a growing pool of highly characterized investigational compounds, which have known safety and pharmacology profiles but no medicinal use. These have been making their way into drug repositioning efforts (Frail et al. 2015). There are some classical success stories in drug repositioning, including the erectile dysfunction drug sildenafil, an inhibitor of PDE5 originally developed for the treatment of coronary artery disease by Pfizer in 1980s (Shim and Liu 2014).

However, there are a number of difficulties that need to be overcome for successful drug repositioning. Firstly, the available chemical space is very large (Reymond et al. 2010; Reymond 2015) and the proportion taken up by the existing set of preclinical or clinical compounds is very small. Secondly even if one is able to match a drug with a target implicated in a given disease, considerable resources and time will still be required to move the drug into the new indication, depending on the level of development it has reached. Thirdly, for a brain disorder such as schizophrenia, the drug would also need sustained bioavailability in the brain, which will not often be the case, and some optimization will therefore be required.

Despite these caveats, there is increasing effort in this area because of developments in methodology and the availability of data, together with the promise that pre-clinical and clinical compounds have for shortening the drug development process and providing new therapeutic
leads (Shineman et al. 2014; Shameer et al. 2015). The topic of the present study, *in silico* drug repositioning uses publically available data on drugs in combination with bioinformatics tools to systematically identify interactions between compounds and targets (Dubus et al. 2009; Wilkinson and Pritchard 2015), in our case generated using disease genetics. This approach has the potential to efficiently identify compounds for target validation studies and reduce the time-to-market.

For each individual drug we created a gene-set using the ChEMBL database resulting in 9,389 (2,496 with unique gene content) gene-sets that are, in essence, drug signatures based on known pharmacological properties. We assessed the significance of these gene sets in the PGC2 Schizophrenia GWAS summary statistics (Ripke et al. 2014) to identify drugs and molecules showing increased evidence in the genetic data.

We found 1,056 nominally significant pathways (139 with unique gene content). The top gene sets include the gene target sets of several plausible and several novel compounds. A set of three drugs and compounds were suggestively associated ($p<0.001196$). The first was Metoclopramide, is a typical antipsychotic of the phenothiazine class currently used an anti-nausea medication. The CheMBL targets of Metoclopramide are *ACHE*, *CYP2D6*, *DRD2*, *DRD3*, *HTR3A*, *HTR3B* of which *DRD2* shows GWAS association with schizophrenia, and the *HTR3A/HTR3B* locus shows association with SNPs with GWAS $p<5x10^{-5}$). The third drug, Trifluoperazine (also known as the antipsychotic Stelazine) is also a D2 agonist with P-glycoprotein (P-gp) blocking activity. P-gp is encoded by the *ABCB1* gene which shows sub-GWAS association (best SNP $p = 4x10^{-7}$) with schizophrenia (Ripke et al. 2014). In addition, Neratinib, the second ranked drug, is a tyrosine kinase inhibitor under investigation for the treatment breast cancer and other solid tumours with a very large spectrum of activity against kinases, notably including multiple kinases within loci showing GWAS significant association with schizophrenia (*FES*, *PRKD1*, *PAK6*, *PTK2B*, *TIE1*) (Ripke et al. 2014) However, with such a promiscuous kinase, it is difficult to draw conclusions on precise therapeutic leads.

A set of nicotinic compounds just failed to reach suggestive significance after for the effects on statistics of the CHRNA gene cluster on chromosome 15. Top amongst these compounds were Altinicline (SIB-1508Y; (S)-(2)-5-ethynyl-3-(1-methyl-2-pyrrolidinyl)pyridine HCl), a neural
nicotinic acetylcholine receptor (nAChR) agonist that displaces nicotine, with high selectivity for the α4β2 subtype, and Dihydro-Beta-Erythroidine, a competitive nicotinic acetylcholine receptor antagonist with moderate selectivity for the neuronal α4 receptor subunit. Nicotinic acetylcholine receptors are well-established therapeutic targets in CNS disorders, including schizophrenia where agonists, partial agonists, or PAMs for α7 nAChRs or partial agonists for α4β2 nAChRs have been developed (Dineley et al. 2015). The nAChR partial agonist, varenicline, improves cognition in schizophrenic patients (Hong et al. 2011; Wing et al. 2013). The rate of tobacco use among schizophrenic patients is very high and indeed nicotine has been proposed to ameliorate some of the sensory deficits in schizophrenia (Leonard et al. 2002). The 15q13.3 region has been implicated in schizophrenia before with rare microdeletions being more prevalent in patients (Stefansson et al., 2008; International Schizophrenia Consortium, 2008). Even though the current analysis was limited to common variation, the analysis still indicated a nominally significant signal in the CHRNA gene cluster.

What was striking was the enrichment for known therapeutic and mechanisms of action among the top hits. This has also been described in a recent paper, where authors focused on SCZ GWAS hits above a certain significance threshold (Lencz and Malhotra 2015) or when examining both rare and common variation (Ruderfer et al. 2016). What sets our approach apart is a genome wide approach, by which we also identified other possible drug targets. Some of the compounds identified by this approach may provide clues to novel treatment strategies in multi-target drug development or potentially be repurposed for the treatment of schizophrenia. However, while GWAS can provide lists of genes from within associated loci, which can be used to identify potential drug targets, it does not provide a therapeutic hypothesis, i.e. whether an agonist or antagonist is required for a particular protein target. This is especially important for multi-target drug development where different modalities may exist for different targets within a multi target set. Therefore, further experimental data is required for each potential target on its alteration and role in disease and finally target validation that will tell us whether manipulation will alter disease.

Additional data on the functional effects of variants is needed to ensure that the direction of effects of drugs is in the therapeutic direction. Secondly, the actions of these drugs are on a set
of individual targets, but in reality they may act on a disease network, so an understanding of their concerted effect on disease biology may be needed, for example by analysis of their biological effects on gene networks in disease models. It may be possible to circumvent these approaches by performing direct proof-of-concept studies in humans where drugs are known to be safe in man, or at least have toxicology and side-effect profiles that are acceptable to patients with schizophrenia. Alternatively, where a drug is commonly prescribed, is brain penetrant and has sufficient affinity for the desired target, pharmaco-epidemiological studies could be performed to see if there is a efficacy signal, such as those performed for calcium channel blockers and Parkinson’s disease (Becker et al. 2008).

Our approach is essentially a drug repositioning and molecule drug discovery method for large scale genetic data, i.e. a “systematic or targeted evaluation of pharmaceutical compound libraries or compound to identify new indications for diseases other than the primary diseases for which the drug was originally designed” (Chong and Sullivan 2007; Dudley et al. 2010; Collins 2011; Shameer et al. 2015). The drug-pathway gene sets we identified act on multiple gene products identified by GWAS of schizophrenia, and thus may target the underlying aetiology of disease. The drugs we identified with formal or suggestive significance include antipsychotic medications acts on the dopamine D2 receptor DRD2 among other targets. Further down nominally significant there are multiple highly ranked drugs and molecules affecting neural nicotinic receptors, calcium channels, and opioid receptors. Notably, all of which (except DRD2) have previously been considered as therapeutic targets for schizophrenia but have never been properly tested. This type of repositioning study, and its refinements, may provide the impetus to test these mechanisms in the clinical treatment of schizophrenia.
**Funding and Disclosure**

This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. GB is a consultant in preclinical genomics and has grant funding from Eli Lilly ltd.
References


