Pharmacological enrichment of polygenic risk for precision medicine in complex disorders

William R. Reay[1,2], Joshua R. Atkins[1,2], Vaughan J. Carr[3,4,5], Melissa J. Green[3,4], Murray J. Cairns[1,2 *]

[1]School of Biomedical Sciences and Pharmacy, The University of Newcastle, Callaghan, NSW, Australia

[2]Centre for Brain and Mental Health Research, Hunter Medical Research Institute, Newcastle, NSW, Australia

[3]School of Psychiatry, University of New South Wales, Randwick, NSW, Australia

[4]Neuroscience Research Australia, Sydney, NSW, Australia

[5]Department of Psychiatry, Monash University, Melbourne, VIC, Australia

To whom correspondence should be addressed:

Professor Murray J. Cairns

School of Biomedical Science and Pharmacy, The University of Newcastle, University Drive, Callaghan, NSW 2308, Australia, Email: Murray.cairns@newcastle.edu.au, Phone: +61 02 4921 8670, Fax: +61 02 4921 7903

Running title: Pharmacological enrichment of polygenic risk

ABSTRACT

Individuals with complex disorders typically have a heritable burden of common variation that can be expressed as a polygenic risk score (PRS). While PRS has some predictive utility, it lacks the molecular specificity to be directly informative for clinical interventions. We therefore sought to develop a framework to quantify an individual's common variant enrichment in clinically actionable systems responsive to existing drugs. This was achieved with a metric designated the *pharmagenic enrichment score* (PES), which we demonstrate for individual SNP profiles in a cohort of cases with schizophrenia. A large proportion of these had elevated PES in one or more of eight clinically actionable genesets enriched with schizophrenia associated common variation. Notable candidates targeting these pathways included vitamins, insulin modulating agents, and protein kinase inhibitors with putative neuroprotective properties. Interestingly, elevated PES was also observed in individuals with otherwise low common variant burden. Our data suggest that the PES framework could integrate an individual's common variant risk to inform personalised interventions, including drug repositioning, for complex disorders such as schizophrenia.

Keywords: complex disorders/drug repurposing/genomics/precision medicine/systems biology

INTRODUCTION

A significant burden of disease is caused by complex traits including psychiatric and neurobehavioural disorders, inflammatory and autoimmune disorders, metabolic and cardiovascular disease, and cancer. Until relatively recently it was difficult to identify the heritable components of these traits, however, the emergence of well powered genome-wide association studies (GWAS) of large cohorts assembled by collaborative consortia are revealing important insights into their common variant architecture [1]. While collectively this information has been vital to map genes and pathways that are likely to be etiological factors, the small effect size of individual variants, and their heterogeneity in the population make their relevance to individuals with the disorder highly variable, relatively specific, and fairly minor with respect to the total variant burden. In other words, the individual variants present as relatively small targets for therapeutic intervention and in many disorders may not attract the investment needed for pharmaceutical development. We therefore need mechanisms for using this vast amount of diverse genetic information to maximise its utility for therapeutic advances. This requires a personalised approach that can capture variant burden in affected individuals with respect to biological components that align with existing medications, and/or pathways of relevance to key pathophysiological processes, to provide sufficient support for the development of new interventions. Approaches that summate the genomic risk burden, such as polygenic risk scoring (PRS), have demonstrated a significant association with many complex traits, including neuropsychiatric disorders [2,3], diabetes [4,5], cardiovascular disease [6,7], and inflammatory disorders [8,9] as just a few examples. In the PRS model, individual alleles are weighted by their association effect size related to a given phenotype generated in well-powered GWAS. Cumulative burden of common variation can therefore be quantified in patients via this methodology. Importantly, PRS has shown the most utility when incorporating variants at less strict thresholds than genome wide significance ($P < 5 \times 10^{-8}$),

that is, a higher degree of captured polygenicity [2]. Recent analysis has shown that individuals in the population with very high PRS can have a greater than three-fold risk increase for a number of complex conditions [6].

A key limitation of genome-wide PRS as an unannotated instrument is its composition of heterogeneous risk factors, which lack biological salience and cannot provide specific information that might help in designing a precision treatment strategy. We therefore hypothesized that investigating common variant enrichment (at the multivariable level) in biological pathways would provide a means of functionally annotating PRS. To test the hypothesis, we devised a statistical framework for scoring polygenic risk in pathways that may be relevant as therapeutic targets for complex traits. This quantitative approach, designated the *pharmagenic enrichment score* (PES), was aimed to provide an indication of an individual's exposure to risk variants that are potentially treatable by existing pharmacological agents, including many that have never been considered or tested previously for the condition/disorder they are experiencing. By focusing on biological pathways with known drug targets, we aimed to enhance the clinical utility of polygenic risk approaches by providing novel and specific opportunities to identify treatment targets and/or repurpose existing drugs. This application of genome-wide common variant genotyping should have particular relevance for the precision treatment of individuals that are resistant to currently indicated medications. In this study we outlined the PES approach and sought to exemplify its utility in the complex neuropsychiatric condition, schizophrenia.

RESULTS

The *pharmagenic enrichment score* approach

The *pharmagenic enrichment score* (PES) leverages the joint effect of common variants in pathways which can be putatively modulated by known pharmacological compounds. Gene-

wise common variant enrichment is calculated at varying *P* value thresholds, to capture different components of the polygenic signal, and thereafter, assembled into genesets. These candidate druggable biological systems are then utilised to deconvolve polygenic common variant risk for the trait at a systems level, thus, providing a pharmacologically directed annotation of this genomic burden. The cumulative effect sizes of variation within that pathway are utilised to construct a PES for an individual. Subjects with elevated PES in a clinically actionable pathway may be candidates for intervention with a drug that is known to modulate it. We propose that the definition of an elevated PES should be relative to a large reference cohort or a cohort with cases and controls for the phenotype in question. A key advantage of this methodology is its capacity to detect unbiased pathway level association, which may reveal novel drug repurposing opportunities.

Clinically actionable pathways enriched with common variant risk in schizophrenia

In schizophrenia, a psychiatric syndrome with approximately 0.7% population prevalence, heritability is traditionally estimated in the region of 80% [10,11]. Global collaboration via the psychiatric genomics consortium (PGC) has amassed large cohorts of schizophrenia cases to define genomic variants associated with the disorder through genome-wide association studies (GWAS). Over 100 loci thus far have been identified at rigorous genome-wide significance levels, with common variants encompassed by GWAS shown to account for a significant portion of schizophrenia heritability [3]. However, its application to treatment is not well defined, particularly as it provides no insight into novel clinical interventions beyond the traditional antipsychotic paradigm.

To address this, we profiled PES in a cohort of schizophrenia patients and screened healthy controls [12]. To identify clinically actionable gene-sets and construct PES, we processed the 2014 psychiatric genomics consortium (PGC) schizophrenia mega GWAS [3]. Eight gene-sets were derived from the schizophrenia GWAS at the different P_T with known drug targets

(Table 1). The most significantly associated gene-set was the *HIF-2 pathway* ($P = 3.12 \times 10^{-5}$, $\beta = 0.435$, $SE = 0.109$, $P_T < 0.005$), which is comprised of genes in the hypoxia inducible factor 2 (HIF-2) alpha transcription factor network. *One carbon pool by folate* was the second most significant pathway with a putative drug interaction ($P = 1.4 \times 10^{-4}$, $\beta = 0.433$, $P_T < 0.05$). Two gene-sets were related to the function of the neurotransmitters GABA and Acetylcholine, whilst other signalling pathways represented were *NOS1* (Nitric Oxide Synthase I), *Hedgehog* signalling and the semaphorin related *CRMP (Collapsin Response Mediator) proteins in Sema3A signalling* pathway. In addition, the geneset *Regulation of Insulin Secretion* passed the threshold for inclusion.

Pathway	***P* threshold (P_T)**	***P***
NOS1 pathway	All SNPs	6.3×10^{-4}
Regulation of insulin secretion	$P < 0.5$	3.9×10^{-4}
CRMPs in Sema3A signalling	$P < 0.5$	9.4×10^{-4}
GABA synthesis, release, reuptake and degradation	$P < 0.5$	5.8×10^{-4}
One carbon pool by folate	$P < 0.05$	1.4×10^{-4}
Hedgehog signalling	$P < 0.05$	1.9×10^{-4}
HIF-2 pathway	$P < 0.005$	3.1×10^{-5}
Acetylcholine binding and downstream events	$P < 0.005$	3.8×10^{-4}

Table 1. Pathways enriched with common variation associated with schizophrenia with putative clinical actionability. Pathways with putative clinical actionability by virtue of having targets for existing drugs with potential for repurposing. Enrichment P values refer gene-set association aggregated SNPs associated with schizophrenia in the PGC GWAS.

The genes which constitute these eight pathways had upregulated expression in the brain relative to the rest of the protein coding genome, with the anterior cingulate cortex the most highly enriched region after multiple testing correction, $P_{Adj} = 6.45 \times 10^{-13}$. Conversely, they were downregulated ($P_{Adj} < 0.05$) in several peripheral tissues including the stomach and skin (Supplementary Fig. 2). These genes were also overrepresented in the GWAS catalogue for traits relevant to psychiatry including schizophrenia, post-traumatic stress disorder, nicotine dependence and cognitive performance ($P_{Adj} < 0.05$, Supplementary Table 1).

The eight gene-sets prioritised by our pipeline are indicative of a diverse range of drug classes. We sought to investigate a selection of candidate pharmacological agents which may be utilised for each PES input pathway. Firstly, we extracted the genes classified in the TCRD as T_{Clin} from each of the gene-sets and matched them to their known drug-interactions using the drug gene interaction database (DGidb v3.02, Supplementary Table 2). The top FDA approved drug per pathway was selected based on the DGidb score of interaction confidence between a T_{Clin} gene and drug. After annotation via the anatomical chemical (ATC) classification system two candidate drugs were anti-neoplastic and immunomodulating agents (ATC = L), two were classified as nervous system (ATC = N), whilst the remaining encompassed one of the following: blood and blood forming organs (ATC = B), musculoskeletal system (ATC = M), sensory organs (ATC = S) and alimentary tract and metabolism (ATC = A). Clinical trials for schizophrenia, either completed or in the recruiting phase, were registered for three of these compounds – glycine, varenicline and exenatide.

Drugs which target a statistically significant number of genes in each pathway were derived using over-representation analysis in WebGestalt [13]. Of the eight gene-sets tested, six had a significant drug enrichment with a minimum overlap of three genes after multiple testing correction (Table 3, Supplementary Table 3). Nervous system drugs were the most common

ATC category (level 1) across all the input pathways. This accorded with previous reports in the literature of schizophrenia associated common variant enrichment in the targets of antiepileptics [14]. Some interesting repurposing candidates with previous clinical trials in the disorder included the psychostimulant Atomoxiene [15], which targets the NOS1 pathway, the α4β2 nicotinic acetylcholine receptor subtype partial agonist Varenicline [16], and ascorbic acid (Vitamin C) [17]. Whilst the results of these trials were mixed, targeting such interventions to specific individuals based on genomic risk is yet to be investigated.

Pathway	**Top Drug***	**Overlap#**	**ATC Code**
NOS1	Milnacipran	5	Other antidepressants
GABA	Temazepam	14	Benzodiazepine derivatives
Insulin	Clodronic acid	3	Bisphosphonates
HIF-2	Ascorbic acid (Vitamin C)	3	Ascorbic acid (Vitamin C)
Acetylcholine	Nicotine	11	Drugs used in nicotine dependence
Folate	Tetrahydrofolic acid	11	Folic acid and derivatives

Table 2. The top enriched drug target for each *pharmagenic enrichment score* pathway with at least three interacting genes after multiple testing correction. Most significantly associated drug after multiple testing correction was selected, when corrected *P* values were equal, the drug with the highest geneset overlap was selected. Overlap refers to the number of genes targeted by the drug in the candidate pathway.

Individual profiling of pharmagenic enrichment scores in a schizophrenia cohort

In this study, a tiered system was implemented to define members of the cohort with a high PES in a clinically actionable geneset. Firstly, we examined individuals in the top percentile of the ASRB cohort for each PES, to explore the phenotypic characteristics of an elevated risk score with high confidence. There were 55 individuals with a top percentile PES, as one schizophrenia case had elevated PES in both the *One Carbon Pool by Folate* and the *GABA synthesis, release, reuptake and degradation* pathways. From this subset, the majority were schizophrenia patients (N=38), however, there was no significant association between top percentile status and diagnosis ($z = 0.975$, $P = 0.33$). We investigated clinical characteristics obtained for ASRB participants to prioritise top percentile PES carriers who may benefit most from a personalised treatment regime. Three variables were selected as a proxy of a more clinically challenging presentation of the disorder: clozapine prescription (as a surrogate for treatment resistance), a global assessment of functioning (GAF) score < 50, and an adolescent onset of the disorder before the age of 18 [18,19]. Interestingly, of the 38 schizophrenia cases with an elevated PES, 71% of this subset meet at least one of these criteria (N =27): clozapine prescription (N = 9), GAF < 50 (N = 12), onset age < 18 (N = 9).

In addition, two less stringent partitions of elevated PES were implemented, specifically, a decile and quartile cut-off for PES in the entire cohort was used to triage patients at elevated risk of dysfunction in that pathway. The highest number of PES in the top decile or quartile respectively for an individual was six (Supplementary Fig. 3). An increasing number of PES in both the top quartile (OR = 1.1493 [95% CI: 1.016 – 1.303], $P = 0.0287$) and decile (OR = 1.207 [95% CI: 1.013 – 1.447], $P = 0.0384$) was associated with schizophrenia. However, this signal was not significant after adjustment for PRS_{Total}. As visualised with kernel density estimation in figure 1b-c, there is evidence of skew towards high PRS_{Total} for those individuals with at least four top quartile or decile PES categories. Whilst the aim of this

study was not to find association with cases for this cohort, two PES were nominally associated with schizophrenia in the ASRB - *Regulation of Insulin Secretion* ($z = 2.262$, $P = 0.0237$) and the *Acetylcholine Binding and Downstream Events* pathways ($z = 2.167$, $P = 0.0303$). However, significance was diminished when covaried for total schizophrenia PRS (PRS_{Total}, $P > 0.05$).

Relationship between pathway-based annotation and genome wide polygenic risk for schizophrenia

We sought to define the relationship between PRS_{Total} and PES in further detail. Pairwise correlation between each of the scores demonstrated no significant univariate relationship between any PES or with PRS_{Total} (Fig. 1a). In addition, top percentile PES individuals were not enriched with PRS_{Total} in comparison to the rest of the cohort: $z = 0.819$, $P = 0.413$ (Fig. 1d). This presented a clinically significant subset of schizophrenia cases with a less polygenic phenotype, that is, low PRS_{Total} relative to the schizophrenia cohort but high heritable risk in one or more pathways. Analysis of the bottom quartile of PRS_{Total} in the ASRB schizophrenia cohort revealed cases (N=10) with top percentile PES but depleted PRS_{Total}. The pathways encompassed in these individuals were: *Acetylcholine* (N=3), *Hedgehog signalling* (N=2), *CRMPs in Sema3A* (N=2), *GABA* (N=1), *HIF-2* (N=1) and *Insulin secretion* (N=1). This information may be of great clinical value as these cases have less marked common variant burden genome-wide, but localised risk in a geneset. Furthermore, three of these patients were prescribed clozapine (surrogate for treatment resistance), a further three had low global functioning (GAF < 50), along with two adolescent onset cases – potentially highlighting a heightened need for precision intervention in these individuals.

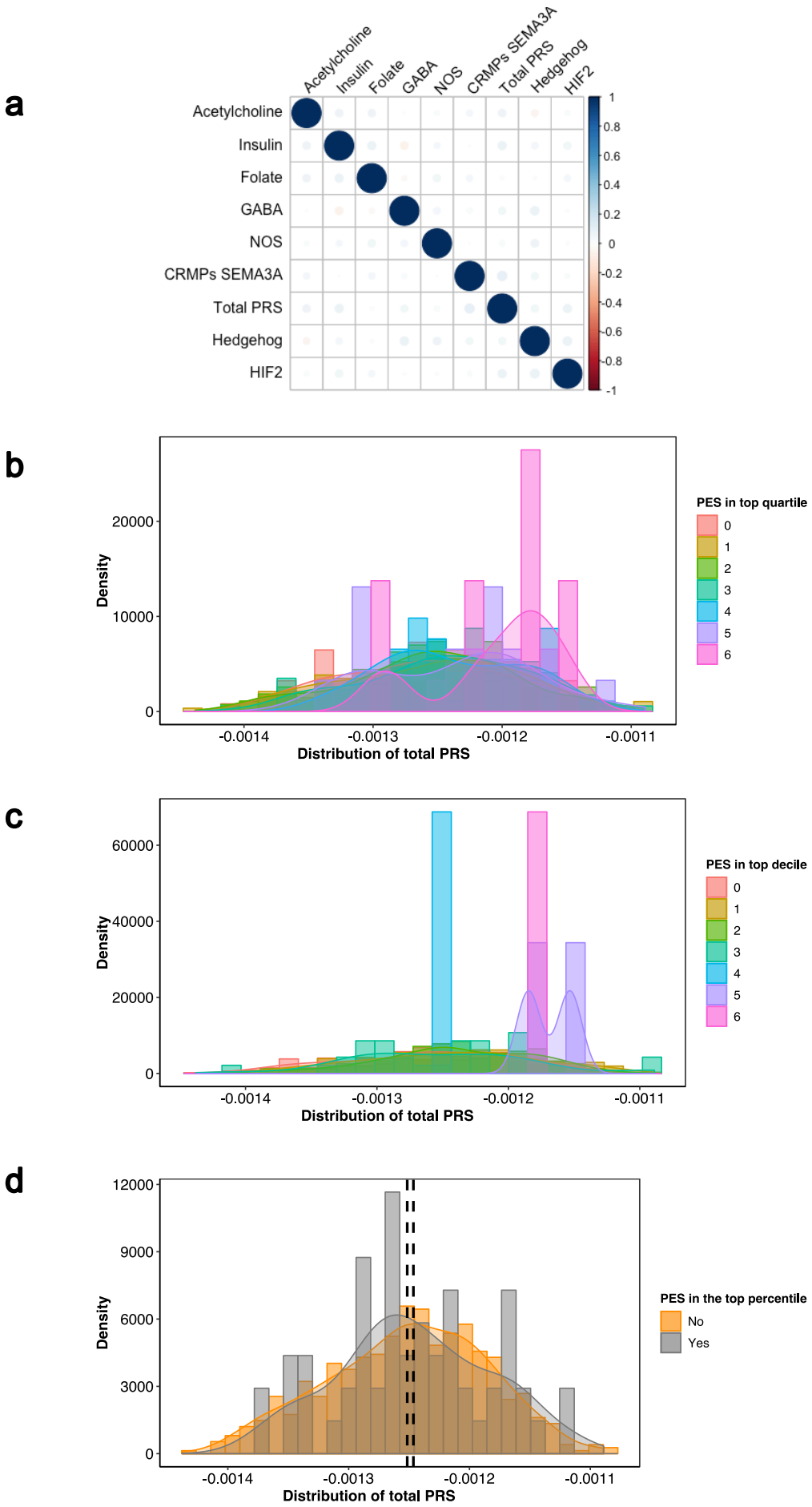
a
Acetylcholine
Insulin
Folate
GABA
NOS
CRMPs SEMA3A
Total PRS
Hedgehog
HIF2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
b
Density
20000
10000
0
-0.0014
-0.0013
-0.0012
-0.0011
Distribution of total PRS
PES in top quartile
0
1
2
3
4
5
6
c
Density
60000
40000
20000
0
-0.0014
-0.0013
-0.0012
-0.0011
Distribution of total PRS
PES in top decile
0
1
2
3
4
5
6
d
Density
12000
9000
6000
3000
0
-0.0014
-0.0013
-0.0012
-0.0011
Distribution of total PRS
PES in the top percentile
No
Yes

Fig 1. Relationship between genome wide schizophrenia PRS and *pharmagenic enrichment scores* in the ASRB cohort. (**a**) Pairwise univariate correlation between each of the PES and total PRS. Scale represents strength of relationship in the positive or negative direction. Kernel density estimation of distribution of total PRS amongst individuals with multiple PES in the top quartile (**b**) or decile (**c**). Scale refers to the number of PES over the threshold in an individual, that is, a score of six represents an individual with six PES categories in the top quartile or decile of the ASRB cohort. (**d**) Distribution of total PRS between ASRB participants with at least one PES in the top percentile of the cohort (grey) or without (orange). Black dashed line represents the mean PRS_{Total} for the cohort with a top percentile PES (right) and without (left).

To investigate the relationship between low polygenic load and elevated PES, genome wide PRS and the PES category count in the top decile per individual were clustered using finite Gaussian mixture modelling (GMM). The optimal number of clusters was selected based on parametrisation of the covariance matrix utilising the Bayesian Information Criterion (BIC), with the highest BIC value used for selection of the number of clusters (Fig. 2a). Four clusters were derived from the data (*BIC* = 13871.21, VEV: variable volume, equal shape, variable orientation) (Fig. 2b). Clusters were ellipsoidal, with the volume of the ellipsoid, shape of the density contours, and orientation of the corresponding ellipsoid also determined by the covariance matrix. The first two clusters were comprised of schizophrenia patients with at least one PES in the top decile of the ASRB cohort, with *Cluster 1* having low PRS_{Total} relative to *Cluster 2*. The third and fourth clusters had no elevated PES but *Cluster 3* represents patients with greater polygenic load, that is PRS_{Total}, than the *Cluster 4*. The distribution of PRS_{Total} in *Cluster 1* reinforces the concept that a subset of schizophrenia patients with lower polygenic risk may have concentrated elevation in one or more specific

biological systems. Analogous to its univariate relationship with PRS_{Total} individuals with extreme PES in the top percentile of the ARSB cohort were not enriched in any of the GMM clusters relative to the largest cluster, *Cluster 2* (*Cluster 2* vs *Cluster 1*: $P = 0.668$; *Cluster 2* vs *Cluster 3*: $P = 0.492$; *Cluster 2* vs *Cluster 4*: $P = 0.977$).

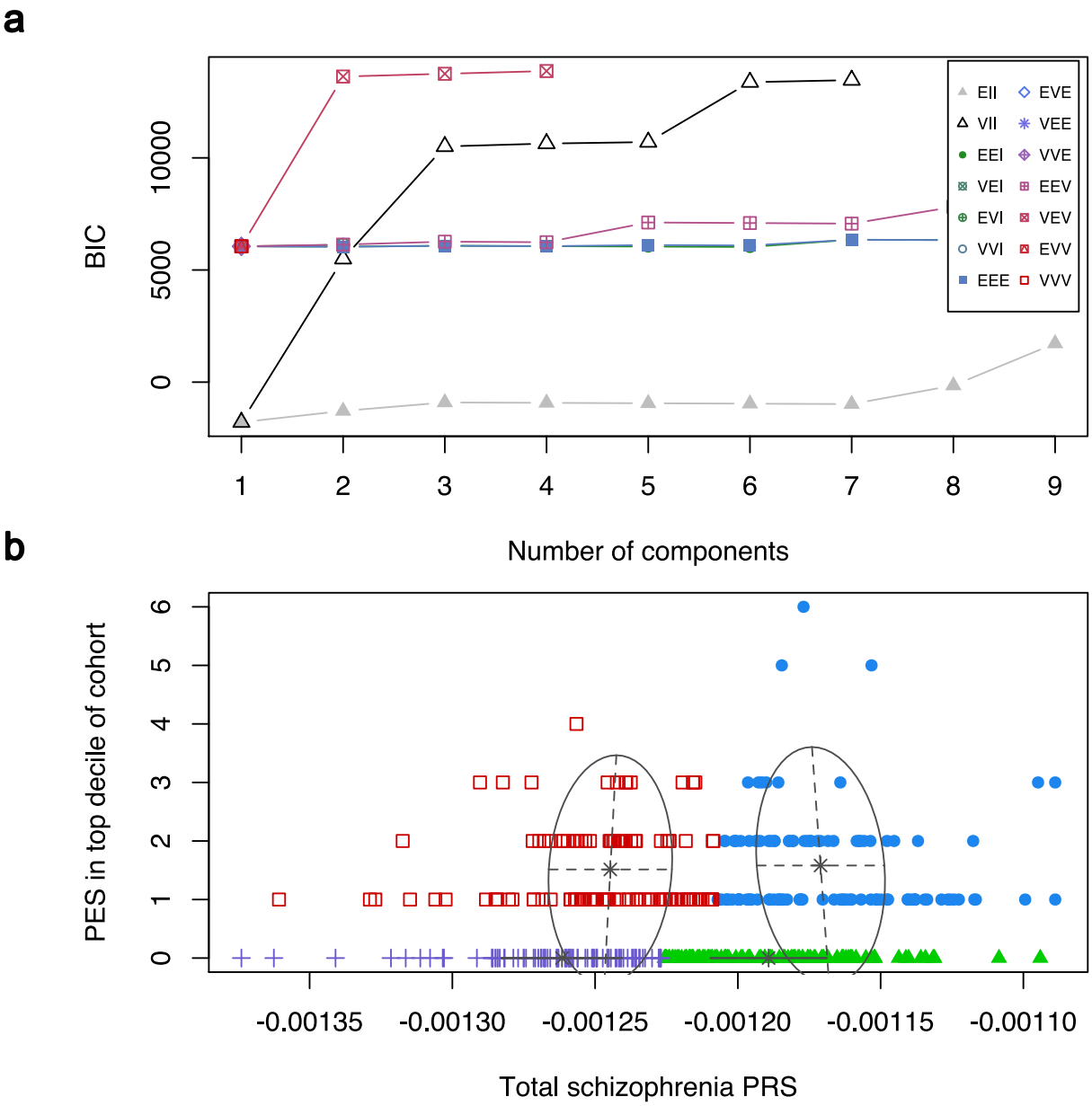

Fig 2. Clustering of genome wide PRS and the individual count of elevated *pharmagenic enrichment score* using finite Gaussian mixture modelling. (**a**) Selection of the number of clusters using the Bayesian information criterion (BIC). The scale represents the fourteen different Gaussian models (see Supplementary Table 5 for definitions) tested for parametrisation of the within-group covariance matrix. (**b**) Derived clusters of genome wide schizophrenia PRS (Total PRS) and the number of PES in the top decile of the ASRB cohort per individual. Red boxes = *Cluster 1*, blue circles = *Cluster 2*, green triangles = *Cluster* 3, purple crosses = *Cluster 4*.

DISCUSSION

The drug development pipeline continues to be prohibitively expensive and time consuming in the translation of novel compounds for clinical practice [20,21]. Repositioning of previously approved drugs for other human health conditions can be a more readily achievable action, particularly for rare disorders where a causal factor can be identified. However, in complex disorders, such as schizophrenia, this approach is hindered by the complexity of the pathophysiology and heterogeneity of genomic risk, along with inter-individual variability in illness onset and clinical course [22,23]. Annotation of the individually-relevant (personalised) genetic components associated with complex syndromes, for the purpose of delineating clinically meaningful biological systems, will both better target existing treatments and reveal new opportunities for drug repurposing (Fig. 3a). In this study, we developed a novel method for capturing common variant risk in biological networks with known drug interactions – *pharmagenic enrichment scores* (PES) – to facilitate precision treatment design relevant to individuals with a particular set of risk variants. A distinct advantage of our PES approach is that it can capture latent enrichment of polygenic signal in pathways relevant to pharmaceutical actions, among individuals whose overall trait PRS is low relative to others

with a shared phenotype. Even in cases where polygenic burden is high, genome-wide PRS (as a biologically unannotated instrument) does not necessarily provide insight into pathways suitable for pharmacological intervention in individuals. Our approach for selection of putative drug targets exemplified in schizophrenia GWAS has revealed potential targets for drug repurposing with substantial clinical utility.

Aggregation of common variation from schizophrenia GWAS into biological pathways with known drug interactions revealed a diverse array of systems relevant to eight distinct PES categories. These candidate pathways displayed common variant enrichment at a range of P_T, indicative of the degree of polygenicity, ranging from using all SNPs as input, to a significance threshold below 0.005 ($P_T < 0.005$). While two of these pathways included GABAergic and cholinergic neurotransmission, both of which are intuitive candidates that have been extensively implicated in schizophrenia with associated drugs already in common practice for neuropsychiatric disorders [24,25], many others were more surprising. The most significantly associated gene-set pertained to the HIF-2 transcription factor network, an important mediator in response to decreases in available cellular oxygen. This has clear significance for biological mechanisms involved in psychiatric disorders, for example in dopaminergic signalling [26]. Enrichment of ascorbic acid (vitamin C) targets in this pathway is notable from a therapeutic perspective because of its antioxidant capabilities, along with preliminary evidence for its efficacy as an adjuvant in the treatment of the disorder [17,27]. The interaction between HIF-2 signalling and NOS1 signalling, another candidate pathway with pharmagenic enrichment in schizophrenia, is supported by previous evidence of redox dysfunction in the disorder [28,29]. The activity of glutamate receptors in the NOS1 system suggests that psycholeptics and psychoanaleptics are likely to modulate this pathway. We also observed common variant enrichment in two developmental pathways that can be pharmacologically modulated: CRMPs in semaphorin 3a signalling and Hedgehog signalling.

The former is able to interact with the tyrosine kinase inhibitor Dasatinib, which is postulated to have neuroprotective properties [30]. Enrichment in these actionable pathways is consistent with longstanding hypotheses of deficits in neurodevelopment contributing to the aetiology of schizophrenia [31].

The breadth of drugs which target these pathways used to construct PES suggests that individual level treatment formulation can become highly specific depending on which systems genomic risk is localised. This would include the stratification of individuals for precision treatment with compounds previously tested on undifferentiated schizophrenia cohorts, including vitamin C, Atomoxiene, and Varenicline which were identified using PES in this study [15-17]. Repurposing drugs for individuals informed by their genetic liability may assist in the reduction of response heterogeneity, which hinders the implementation of novel treatments in very complex phenotypes like schizophrenia. We suggest that the individuals with PES in the top percentile of any pathway, particularly those with low genome wide PRS, present as the most tractable candidates for this approach; whereas the clinical significance of particular sets of common variant burden would be missed by an unannotated genome wide association indexed by total PRS alone.

While we expect that in most circumstances the aggregate of variation constituting high PES represent pathology of the target pathways, a current limitation of this methodology is that it does not integrate the direction of effect. While this may be possible as more functional annotations become available, this would present an immensely complex paradigm to predict *in silico* due to the vast array of factors which influence the penetrance of genomic risk. To overcome this, candidate drugs could be modelled on patient derived cell lines to, firstly, establish the extent of dysregulation conferred by elevated PES and, secondly, investigate the interaction with the drug in question. In addition, whilst approaches like MAGMA approximate joint variant effects at gene level, future advancements in annotation

(particularly of non-coding regions) will enhance the identification of true functional signals within each GWAS locus. The current analysis also used very stringent criteria for drug pathway gene inclusion and there are likely to be many more genes and pathways that may be implicated in future ontologies and other investigator curated gene sets. In a recent example of this we observed an enrichment of retinoid gene variation in schizophrenia [32].

New GWAS summary statistics are emerging daily on ever larger samples and these too will further enrich the substrate for PES determination and increase its clinical utility. Despite the aforementioned challenges, we believe that this methodology provides a useful framework to better utilise the breadth of available GWAS data for personalised treatment formulations. Particularly, as there remains a largely unmet need to translate polygenic risk for complex disorders into tractable treatment outcomes for affected individuals. Whilst we have demonstrated here the potential utility in schizophrenia, there is clearly scope to adapt this analytical approach to other complex disorders with summary statistics from well-powered GWAS. This methodology may also be applicable to prophylactic intervention for individuals at high risk for a complex phenotype (Fig. 3b). This could be implemented conservatively with lifestyle or dietary measures implicated by clusters of enrichment captured within the PES framework. For example, in schizophrenia we identified multiple actionable pathways quantified by PES that are modulated by vitamins, which represent a relatively uncomplicated intervention for individuals at high genetic risk for this disorder.

In conclusion, we propose that enrichment of common variation within pathways with known drug targets could be clinically significant. The *pharmagenic enrichment score* framework described in this study, thus, provides a mechanism to both identify novel drug repurposing candidates from GWAS and prioritise their use in genetically susceptible individuals.

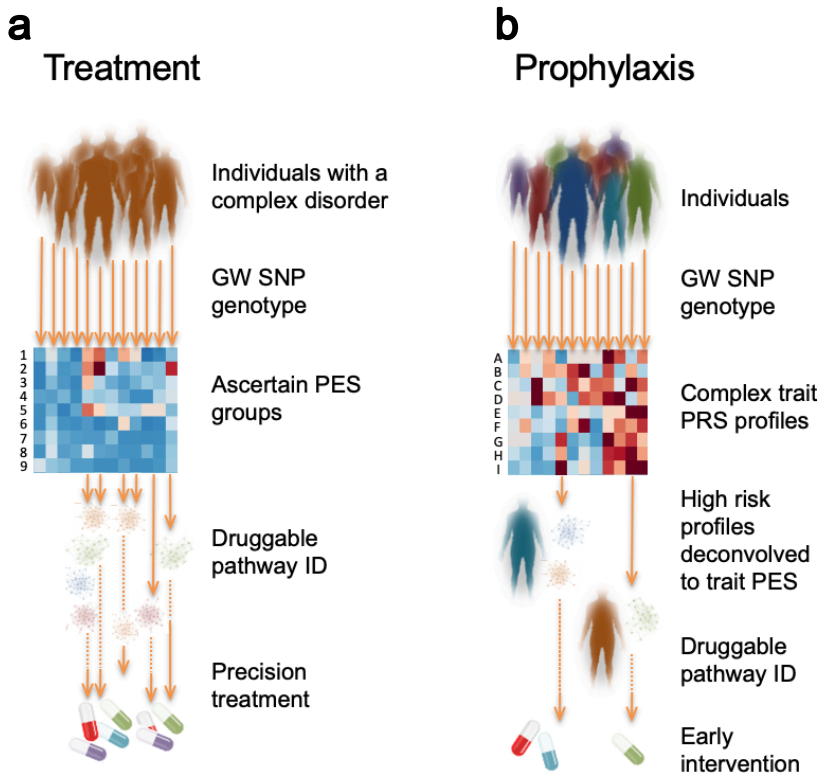

Fig. 3. Implementation of *pharmagenic enrichment score* (PES) in precision treatment and prophylaxis of complex disorders. (**a**) Using the PES framework individuals with a complex disorder provide DNA for common variant SNP genotyping, which is used to ascertain individuals with a high PES. The PES groups (heatmap rows 1-9, with enrichment denoted by darker colour) intrinsically identify precision treatment options tailored to the individual's biological enrichment (of pathways with known drug targets) for polygenic risk in that pathway. (**b**) A more advanced implementation of PES could be achieved for prophylactic intervention for individuals in the population at very high polygenic risk for a variety of complex traits with clinical actionability (heatmap rows A-I, with high PRS denoted by darker colour). Conservative prophylactic measures in this context may account for environmental risk exposure and focus on lifestyle interventions, such as diet and exercise, rather than pharmaceutical treatments that may not be justified without symptom presentation because of their side effect and/or cost.

MATERIALS AND METHODS

Pipeline for the derivation of *pharmagenic enrichment scores*

The methodology developed for constructing *pharmagenic enrichment scores* (PES) is outlined in a schematic presented in Supplementary Figure 1. We exploit the results of gene-set association analysis to aggregate variants from GWAS into genesets which may be candidates for pharmacological intervention. These genesets were high quality canonical and hallmark pathways sourced from the molecular signatures database (MSigDB) [33]. Pathways were designated as clinically relevant if they contained at least one gene annotated to interact with an approved pharmacological agent in the DrugCentral database as classified by the Target Central Resource Database (TCRD, genes annotated as T_{Clin}) [34].

Firstly, this process tests the combined effect of variants at the gene level. We utilised an omnibus *P* value test to achieve this, whereby a linear combination of variant-wise *P* values in a gene is compared to a null distribution to derive a combined *P* value for that gene. This was performed using the MAGMA package to account for linkage disequilibrium between variants in the approximation of the null χ^2 distribution [35]. We mapped SNPs to protein coding genes (hg19, NCBI) with the genic boundaries encompassing 5kb upstream and 1.5 kb downstream to capture potential regulatory variation. Genes within the extended major histocompatibility complex (MHC) on chromosome 6 were excluded from this study due to the complexity of haplotypes in that region. Traditionally, the full breadth of variants available in the summary statistics are utilised in these models. This, however, may miss important aspects of the biological interpretation of the genomic signal. PRS derived from large GWAS cohorts are tested at different P value thresholds (P_T) to exploit a model which explains the most variance between the case and control groups (in a dichotomous construct). For instance, a $P_T < 0.05$ means that only *P* values in the GWAS with a stronger association (*P* value) than this threshold are included. A range of P_T have been suggested to be optimal

for PRS in several disorders depending on their genomic architecture [2,7,36]. Aggregating the combined effects of variants in genes at different P_T may therefore capture the biological complexity of the signal at varying degrees of polygenicity.

Once variants are aggregated in genes at varying significance thresholds (P_T), we conduct gene-set association of the pre-defined druggable genesets with the trait of interest. This was a competitive association test, which tests the null hypothesis of genes in the set being no more strongly associated than all other genes. This model was implemented with the MAGMA package. Enriched pathways with a known drug target uncovered from this pipeline then form the basis of calculating the *pharmagenic enrichment score* (PES). To profile individuals for a PES, genes which compromise the candidate pathway are extracted and cumulative genomic risk calculated in an analogous fashion to PRS, assuming an additive model [2]. Each individual profiled thus has a genomic risk score within each clinically actionable pathway.

Application of *pharmagenic enrichment score* to identify drug repurposing candidates for schizophrenia

To identify clinically actionable gene-sets and construct PES, we processed the 2014 psychiatric genomics consortium GWAS for the complex neuropsychiatric disorder schizophrenia with the pipeline described above [3]. Variants were selected based on their significance (P value) for inclusion in the at four different thresholds – all SNPs, $P < 0.5$, $P < 0.05$, and $P < 0.005$. Geneset association was conducted on 1012 MSigDB pathways with at least one druggable (T_{Clin}) gene. To capture a wide variety of pathways for a complex phenotype like schizophrenia, we used a nominal significance threshold of $P < 0.001$ to select candidate pathways for PES.

We annotated each of these candidate pathways for their drug interactions, tissue specificity, and phenotypic associations. Using WebGestalt, drugs which target a statistically significant number of genes in the PES genesets were identified after the application of multiple testing correction (FDR < 0.05) [13]. A minimum overlap of at least three targets overlapping the geneset for each pharmaceutical agent was also implemented. Drugs were also mapped to genesets using DGidb v3.02, with the top FDA approved drug per pathway was selected based on the DGidb score of interaction confidence between a T_{Clin} gene and drug [37]. Further, the tissue specificity of expression for all genes from the eight pathways was investigated using GENE2FUNC application of FUMA [38]. Transcript expression in each of the 53 tissue types in the GTEx v7 dataset for the input pathway genes was tested for upregulation, downregulation and two-sided differential expression in comparison to the entire protein-coding genome. Enrichment of input genes for associated traits in GWAS catalogues was also tested in the FUMA framework.

Individual profiling of *pharmagenic enrichment scores* in a genotyped schizophrenia cohort

We sought to generate PES for the schizophrenia candidate pathways in a cohort of diagnosed schizophrenia cases and non-neuropsychiatric controls sourced from the Australian Schizophrenia Research Bank (ASRB) [12,32]. Detailed descriptions of consent procedures along with inclusion and exclusion criteria for the ASRB have been extensively described elsewhere [12]. The Illumina Infinitium Human 610K (610-Quad) BeadChip platform was used to genotype genomic DNA extracted from peripheral blood mononucleocytes as per standard manufacturer protocols. Variant and individual level quality control, along with imputation using the 1000 genomes phase 3 European reference panel, has been outlined in detail for this cohort previously [32]. High quality autosomal sites with low missingness (< 2%) and an imputation score greater than 0.8 ($R^2 > 0.8$) were retained for analysis in this study. After the

removal of individuals in the post-genotyping quality control, 425 schizophrenia cases and 251 controls were analysed in this study; cases were 67% male, whilst males comprised 44% of the control cohort (Supplementary Table 5). The use of these data was approved by the University of Newcastle Human Ethics Research Committee (HREC).

PES is calculated from SNPs mapped to genes which form the candidate pharmacologically actionable geneset. This comprises the following model (1) which sums the statistical effect size of each variant in the geneset multiplied the allele count (dosage) for said variant– for individual i, let $\widehat{\beta_j}$ denote the statistical effect size from the GWAS for each variant j in the candidate gene-set, multiplied by the dosage (G) of j in i .

$$PES_i = \sum_{j}^{M} \left(\widehat{\beta_j} \times G_{ij}\right) \qquad (1)$$

This was calculated for the individuals using the PRSice2 package [39]. For each PES, the P_T used to derive the score was selected based on which P_T the corresponding geneset was derived from the GWAS, when a geneset was associated at multiple P_T, the most significant was chosen. A genome-wide PRS (PRS_{Total}) was also constructed for this cohort, with the P_T which explained the most variance between cases and controls selected using Nagelkerke's R^2. Association for each of the scores with cases was conducted using binomial logistic regression, adjusted for sex and the first three principal components using R 3.4.4. *P* values were derived using the Wald test with and without the total PRS score at the optimum threshold as a covariate in the model.

Each PES was ranked for individuals within the ASRB cohort, with three metrics used to define a person with an 'elevated PES' score: the top percentile, decile, and quartile of the study population. The number of pathways which pass these thresholds were totalled for each individual and the association between these totals and schizophrenia assessed using the same

model as for univariate PRS as described above. To investigate the relationship between PRS_{Total} and PES, genome wide PRS_{Total} and the count of PES in the top decile per individual were clustered using finite Gaussian mixture modelling (GMM) with the mclust package version 5.4 [40]. The optimal number of clusters was selected based on parametrisation of the covariance matrix utilising the Bayesian Information Criterion (BIC), with the highest BIC value used for selection of the number of clusters. Clusters were ellipsoidal, with the volume of the ellipsoid, shape of the density contours, and orientation of the corresponding ellipsoid also determined by the covariance matrix. We tested whether schizophrenia cases in each of the four GMM derived clusters were overrepresented for carriers of a top percentile PES using multinomial logistic regression with the nnet package (https://cran.r-project.org/web/packages/nnet/index.html). The largest cluster, *Cluster 2*, was used as the reference for the other clusters, with the model covaried for sex and prinicipal components as above. After dividing the regression coefficients by their standard error to derive *z*, *P* values were calculated using the Wald Test.

ACKNOWLEDGMENTS

This study was supported by the New South Wales Health, Collaborative Genomics grant program (MC, MG, VC), a NARSAD Independent Investigator Grant (MC) and National Health and Medical Research Council (NHMRC) project grants (1067137, 1147644). Data and samples were also collected by the Australian Schizophrenia Research Bank (ASRB), supported by the Australian NHMRC, the Pratt Foundation, Ramsay Health Care, and the Viertel Charitable Foundation. The ASRB were also supported by the Schizophrenia Research Institute (Australia), utilising infrastructure funding from NSW Health and the Macquarie Group Foundation. WR is supported by an Australian Postgraduate Award. JA was supported by University of Newcastle RHD and an Emlyn and Jennie Thomas

Postgraduate Medical Research Scholarship. MC is supported by an NHMRC Senior Research Fellowship (1121474).

AUTHOR CONTRIBUTIONS

WR and MC designed the study and wrote the first draft of the manuscript. WR performed the genomic and statistical analyses. JA performed the variant imputation. JA, VC, and MG contributed to the interpretation of the results and the preparation of the manuscript.

DATA AVAILABILITY

Schizophrenia GWAS data is available from the Psychiatric Genomics Consortium (https://www.med.unc.edu/pgc/results-and-downloads). SNP array data used in this study is available upon application to the Australian Schizophrenia Research Bank, URL: https://www.neura.edu.au/discovery-portal/asrb/. Command line arguments for the bioinformatics tools in this study are available upon reasonable request to the authors.

REFERENCES

1. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics.* 2017;101(1):5-22.
2. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009;460(7256):748-752.
3. Schizophrenia Working Group of the Psychiatric Genomics C, Ripke S, Neale BM, et al. Biological Insights From 108 Schizophrenia-Associated Genetic Loci. *Nature.* 2014;511(7510):421-427.
4. Xue A, Wu Y, Zhu Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications.* 2018;9(1):2941.
5. Patel KA, Oram RA, Flanagan SE, et al. Type 1 Diabetes Genetic Risk Score: A Novel Tool to Discriminate Monogenic and Type 1 Diabetes. *Diabetes.* 2016;65(7):2094-2099.
6. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics.* 2018;50(9):1219-1224.
7. Abraham G, Havulinna AS, Bhalala OG, et al. Genomic prediction of coronary heart disease. *European Heart Journal.* 2016;37(43):3267-3278.

8. Belsky DW, Sears MR, Hancox RJ, et al. Polygenic risk and the development and course of asthma: an analysis of data from a four-decade longitudinal study. *The Lancet Respiratory medicine.* 2013;1(6):453-461.

9. Cleynen I, Boucher G, Jostins L, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *The Lancet.* 2016;387(10014):156-167.

10. Hilker R, Helenius D, Fagerlund B, et al. Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biological Psychiatry.* 2018;83(6):492-498.

11. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Archives of General Psychiatry.* 2003;60(12):1187-1192.

12. Loughland C, Draganic D, McCabe K, et al. Australian Schizophrenia Research Bank: a database of comprehensive clinical, endophenotypic and genetic data for aetiological studies of schizophrenia. *The Australian and New Zealand journal of psychiatry.* 2010;44(11):1029-1035.

13. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Research.* 2017;45(W1):W130-W137.

14. Gaspar HA, Breen G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Scientific Reports.* 2017;7(1):12460.

15. Kelly DL, Buchanan RW, Boggs DL, et al. A randomized double-blind trial of atomoxetine for cognitive impairments in 32 people with schizophrenia. *The Journal of clinical psychiatry.* 2009;70(4):518-525.

16. Smith RC, Amiaz R, Si T-M, et al. Varenicline Effects on Smoking, Cognition, and Psychiatric Symptoms in Schizophrenia: A Double-Blind Randomized Trial. *PloS one.* 2016;11(1):e0143490-e0143490.

17. Dakhale GN, Khanzode SD, Khanzode SS, Saoji A. Supplementation of vitamin C with atypical antipsychotics reduces oxidative stress and improves the outcome of schizophrenia. *Psychopharmacology (Berl).* 2005;182(4):494-498.

18. Samara MT, Dold M, Gianatsi M, et al. Efficacy, acceptability, and tolerability of antipsychotics in treatment-resistant schizophrenia: A network meta-analysis. *JAMA Psychiatry.* 2016;73(3):199-210.

19. Clemmensen L, Vernal DL, Steinhausen H-C. A systematic review of the long-term outcome of early onset schizophrenia. *BMC psychiatry.* 2012;12:150-150.

20. Smietana K, Siatkowski M, Møller M. Trends in clinical success rates. *Nature Reviews Drug Discovery.* 2016;15:379.

21. McNamee LM, Walsh MJ, Ledley FD. Timelines of translational science: From technology initiation to FDA approval. *PLOS ONE.* 2017;12(5):e0177371.

22. McGrath J. Dissecting the Heterogeneity of Schizophrenia Outcomes. *Schizophrenia Bulletin.* 2008;34(2):247-248.

23. Deng C, Dean B. Mapping the pathophysiology of schizophrenia: interactions between multiple cellular pathways. *Frontiers in Cellular Neuroscience.* 2013;7(238).

24. de Jonge JC, Vinkers CH, Hulshoff Pol HE, Marsman A. GABAergic Mechanisms in Schizophrenia: Linking Postmortem and In Vivo Studies. *Frontiers in Psychiatry.* 2017;8:118.

25. Scarr E, Gibbons A, Neo J, Udawela M, Dean B. Cholinergic connectivity: it's implications for psychiatric disorders. *Frontiers in Cellular Neuroscience.* 2013;7(55).

26. Smeyne M, Sladen P, Jiao Y, Dragatsis I, Smeyne RJ. HIF1α is Necessary for Exercise-Induced Neuroprotection while HIF2α is Needed for Dopaminergic Neuron Survival in the Substantia Nigra pars compacta. *Neuroscience.* 2015;295:23-38.

27. Bentsen H, Osnes K, Refsum H, Solberg DK, Bohmer T. A randomized placebo-controlled trial of an omega-3 fatty acid and vitamins E+C in schizophrenia. *Transl Psychiatry.* 2013;3:e335.

28. Schmidt-Kastner R, van Os J, Esquivel G, Steinbusch HWM, Rutten BPF. An environmental analysis of genes associated with schizophrenia: hypoxia and vascular factors as interacting elements in the neurodevelopmental model. *Molecular Psychiatry.* 2012;17:1194.

29. Olson N, van der Vliet A. Interactions between Nitric Oxide and Hypoxia-Inducible Factor Signaling Pathways in Inflammatory Disease. *Nitric oxide : biology and chemistry / official journal of the Nitric Oxide Society.* 2011;25(2):125-137.

30. Wrasidlo W, Crews LA, Tsigelny IF, et al. Neuroprotective effects of the anti-cancer drug sunitinib in models of HIV neurotoxicity suggests potential for the treatment of neurodegenerative disorders. *British Journal of Pharmacology.* 2014;171(24):5757-5773.

31. Rapoport JL, Giedd JN, Gogtay N. Neurodevelopmental model of schizophrenia: update 2012. *Molecular psychiatry.* 2012;17(12):1228-1238.

32. Reay WR, Atkins JR, Quidé Y, Carr VJ, Green MJ, Cairns MJ. Polygenic disruption of retinoid signalling in schizophrenia and a severe cognitive deficit subtype. *Molecular Psychiatry.* 2018.

33. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov Jill P, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems.* 2015;1(6):417-425.

34. Oprea TI, Bologa CG, Brunak S, et al. Unexplored therapeutic opportunities in the human genome. *Nature reviews Drug discovery.* 2018;17(5):317-332.

35. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology.* 2015;11(4):e1004219.

36. The International Multiple Sclerosis Genetics C. Evidence for Polygenic Susceptibility to Multiple Sclerosis—The Shape of Things to Come. *American Journal of Human Genetics.* 2010;86(4):621-625.

37. Cotto KC, Wagner AH, Feng Y-Y, et al. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Research.* 2018;46(D1):D1068-D1073.

38. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications.* 2017;8(1):1826.

39. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics (Oxford, England).* 2015;31(9):1466-1468.

40. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R journal.* 2016;8(1):289-317.