

OPEN

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*}

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 pharmacogenic 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 gene-sets enriched with schizophrenia associated common variation. Notable candidates targeting these pathways included vitamins, antioxidants, insulin modulating agents, and cholinergic drugs. Interestingly, elevated PES was also observed in individuals with otherwise low common variant burden. The biological saliency of PES profiles were observed directly through their impact on gene expression in a subset of the cohort with matched transcriptomic data, supporting our assertion that this gene-set orientated approach could integrate an individual's common variant risk to inform personalised interventions, including drug repositioning, for complex disorders such as schizophrenia.

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) using large cohorts assembled by collaborative consortia are revealing important insights into their common variant architecture¹. While collectively this information has been vital to map genes and pathways that are likely to be etiological factors, the small effect size of each variant, 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. They also present as relatively small targets for therapeutic intervention and 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.

While approaches that summate the genomic risk burden in individuals, such as polygenic risk scoring (PRS), have demonstrated some predictive utility for complex traits (such as neuropsychiatric disorders^{2,3}, diabetes^{4,5}, cardiovascular disease^{6,7}, and inflammatory disorders^{8,9}) their composition of heterogeneous risk factors lack the biological salience needed to design a precision treatment strategy. We, however, hypothesized that the biologically supervised enrichment of trait-associated common variants in clinically actionable pathways would provide a means of pharmacologically annotating PRS in individuals. To test this proposal, we devised a statistical framework for scoring polygenic risk (at the multivariable level) in pathways relevant as therapeutic targets for complex traits. This quantitative approach, designated the *pharmacogenic enrichment score* (PES), was developed to provide an indication of an individual's exposure to risk variants that are potentially treatable by existing pharmacological

¹School of Biomedical Sciences and Pharmacy, The University of Newcastle, Callaghan, NSW, Australia. ²Centre for Brain and Mental Health Research, Hunter Medical Research Institute, Newcastle, NSW, Australia. ³School of Psychiatry, University of New South Wales, Randwick, NSW, Australia. ⁴Neuroscience Research Australia, Sydney, NSW, Australia. ⁵Department of Psychiatry, Monash University, Melbourne, VIC, Australia. *email: Murray.cairns@newcastle.edu.au

Pathway	P threshold (P_T)	P value
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. The P thresholds denote the SNPs included in the model, that is, only SNPs below that threshold of association in the GWAS were included. P values refer gene-set association of aggregated SNPs for this pathway from the schizophrenia in the PGC GWAS.

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 endeavour 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 individuals with the complex psychiatric 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 (P_T), to capture different components of the polygenic signal, and thereafter, assembled into genesets. We selected four P -value thresholds (all SNPs, $P_T < 0.5$, $P_T < 0.05$, and $P_T < 0.005$) to represent nominal uncorrected significance ($P < 0.05$), SNPs within an order of magnitude more or less significant, and a model with all SNPs. 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. Schizophrenia is a typical complex trait disorder with a prevalence around 0.7% and heritability in the region of 80%^{10,11}. A substantial proportion of this heritability was accounted for in the 2014 psychiatric genomics consortium (PGC) mega GWAS, which identified over 100 common variant loci at rigorous genome-wide significance level³, making the disorder a suitable candidate to test the implementation of the PES framework. Using the complete summary statistics, we identified eight clinically actionable gene-sets at the different P_T containing known drug targets (Table 1). The most significantly associated of these 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 ($P = 1.4 \times 10^{-4}$, $\beta = 0.433$, $SE = 0.119$, $P_T < 0.05$). Two gene-sets were related to the function of the neurotransmitters GABA and Acetylcholine, whilst other signalling pathways represented were *NOS1* (Nitric Oxide Synthase 1), *Hedgehog* signalling and the semaphorin related *CRMP* (*Collapsin Response Mediator*) proteins in *Sema3A* signalling pathway. In addition, the gene-set *Regulation of Insulin Secretion* passed the threshold for inclusion. The number of genes in each pathway and the number of clumped SNPs used as input for each score are detailed in Supplementary Table 1.

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 2).

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 Target Central Resource Database (TCRD) as T_{Clin} from each of the gene-sets, T_{Clin} genes were annotated by the TCRD as those with a high confidence interaction with an approved pharmacological agent¹². Thereafter, these genes were matched to their known drug-interactions using the drug gene interaction database (DGidb v3.02, Supplementary Table 3). The top FDA approved drug per pathway was

Pathway	Top Drug	Overlap	Adjusted P-value	ATC Code
NOS1	Milnacipran	5	1.63×10^{-7}	Other antidepressants
GABA	Temazepam	14	0	Benzodiazepine derivatives
Insulin	Clodronic acid	3	1.97×10^{-3}	Bisphosphonates
HIF-2	Ascorbic acid (Vitamin C)	3	4.61×10^{-2}	Ascorbic acid (Vitamin C)
Acetylcholine	Nicotine	11	0	Drugs used in nicotine dependence
Folate	Tetrahydrofolic acid	11	0	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, very small *P* values are denoted as zero by WebGestalt. Overlap refers to the number of genes targeted by the drug in the candidate pathway.

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 WebGestalt¹³. Of the eight gene-sets tested, six had a significant drug enrichment with a minimum overlap of three genes after multiple testing correction (Table 2, Supplementary Table 4). Nervous system drugs were the most common ATC category (level 1) across all the input pathways. Some interesting repurposing candidates with previous clinical trials in the disorder included the psychostimulant Atomoxetine¹⁴, the $\alpha 4\beta 2$ nicotinic acetylcholine receptor subtype partial agonist Varenicline¹⁵, acetylcysteine (*N*-acetylcysteine) – a precursor to the antioxidant glutathione^{16–18}, ascorbic acid (Vitamin C)^{19,20}, vitamin E²⁰, and memantine^{21,22}. Whilst the results of these trials were mixed, targeting such interventions to specific individuals based on genomic risk is yet to be investigated.

Individual profiling of pharmagenic enrichment scores in a schizophrenia cohort. We profiled PES in a cohort of schizophrenia patients ($N = 425$) and screened healthy controls ($N = 251$) from the Australian Schizophrenia Research Bank (ASRB) and identified members of the cohort with relatively high PES in clinically actionable gene-sets²³. Firstly, we examined individuals in the 99th percentile of the ASRB cohort for each PES, to explore the phenotypic characteristics of those with highly elevated risk scores. There were 55 individuals in the 99th percentile for a PES in the ASRB, including one schizophrenia case who had a PES in the 99th percentile for 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 99th percentile status, that is, having a PES in this quantile, and diagnosis ($z = 0.975$, $P = 0.33$). We investigated clinical characteristics obtained for ASRB participants to prioritise 99th 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^{24,25}. Interestingly, of the 38 schizophrenia cases with an elevated PES in the 99th percentile, 71% of this subset meet at least one of these criteria ($N = 27$): clozapine prescription ($N = 9$), GAF < 50 ($N = 12$), onset age < 18 ($N = 9$). While this does not represent a statistically significant enrichment of these individuals carrying a 99th percentile PES, these patients in particular may benefit from a precision intervention using the PES approach, due to their challenging clinical presentation.

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 quartile 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 $\text{PRS}_{\text{Total}}$, which denotes conventional genome-wide PRS for schizophrenia. As visualised with kernel density estimation in Fig. 1b,c, there is evidence of skew towards high $\text{PRS}_{\text{Total}}$ 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 ($\text{PRS}_{\text{Total}}$, $P > 0.05$).

Relationship between pathway-based annotation and genome wide polygenic risk for schizophrenia. We sought to define the relationship between $\text{PRS}_{\text{Total}}$ and PES in further detail. Pairwise correlation between each of the scores demonstrated there was no significant univariate relationship between any PES, whilst in this sample no PES profiles were correlated with $\text{PRS}_{\text{Total}}$ (Fig. 1a). In addition, top percentile PES individuals were not enriched with $\text{PRS}_{\text{Total}}$ in comparison to the rest of the cohort: $z = 0.819$, $P = 0.413$ (Fig. 1d). This

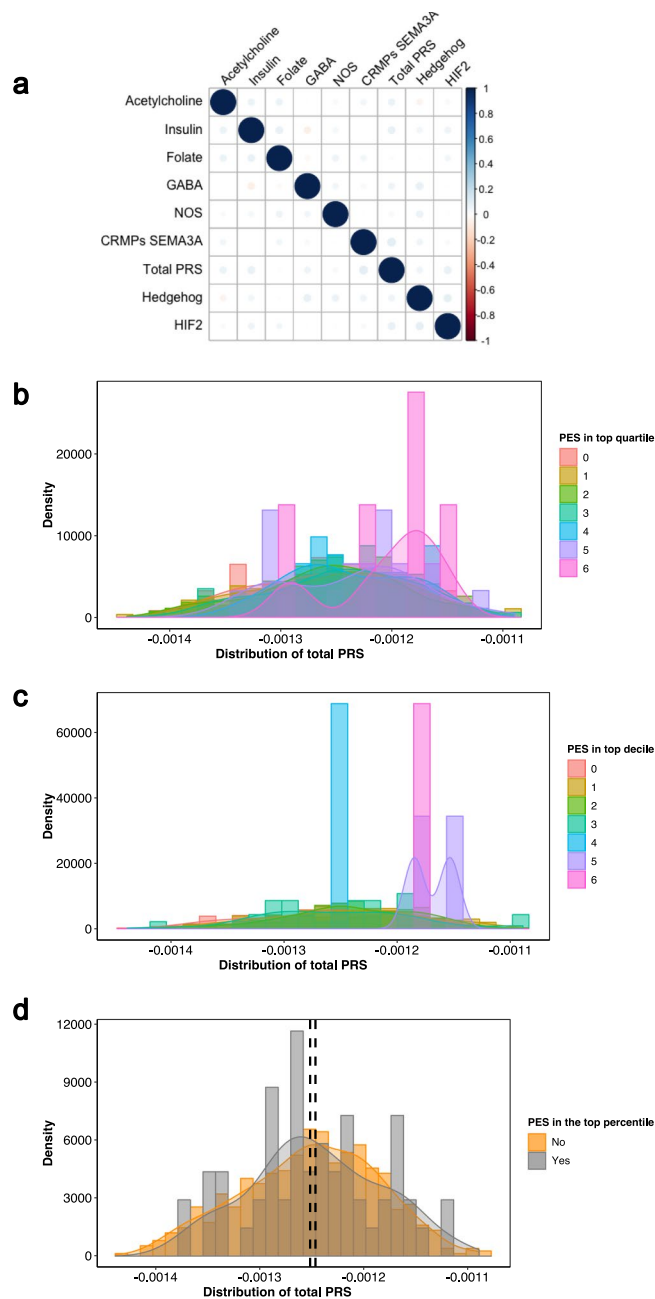


Figure 1. Relationship between genome wide schizophrenia PRS and *pharmagenic enrichment scores* in the ASRB cohort. **(a)** Pairwise univariate correlation between each of the PES and total PRS. Scale represents strength of relationship in the positive or negative direction. Kernel density estimation of the distribution of total schizophrenia PRS in the entire ASRB cohort, stratified by the number of PES profiles in the top quartile **(b)** and decile **(c)** respectively. These plotted distributions range from individuals with no elevated PES, denoted zero on the scale, to those with six elevated PES, 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 99th percentile of the cohort (grey, $N = 55$) or without (orange, $N = 621$). Black dashed line represents the mean PRS_{Total} for the cohort with a top percentile PES (right) and without (left).

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 ($N = 10$, 9.3% of the bottom quartile of PRS_{Total} in the schizophrenia cohort [$N = 107$], and 1.5% of the entire ASRB sample [$N = 676$]) with a 99th percentile PES but reduced 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,

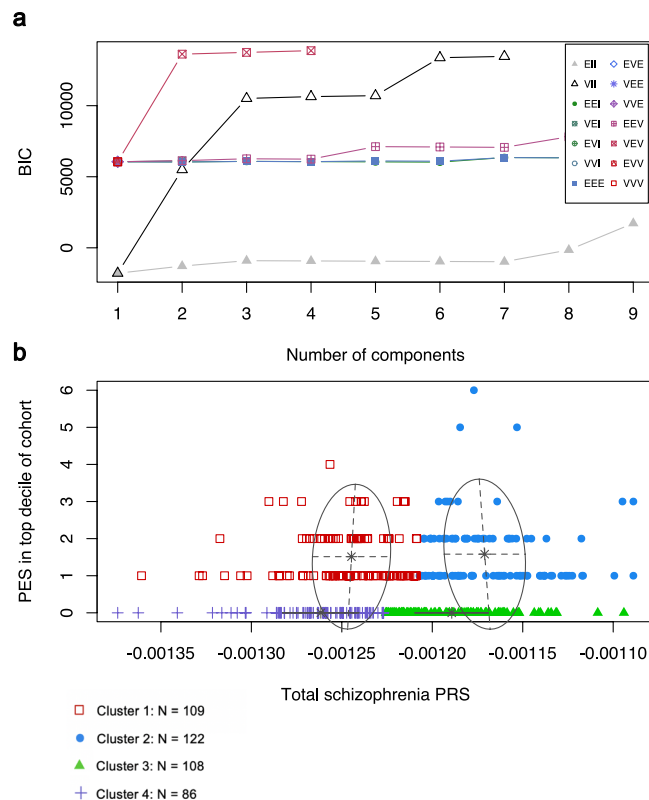


Figure 2. Clustering of genome wide PRS and the individual count of elevated *pharmagenic enrichment score* using finite Gaussian mixture modelling. **(a)** Selection of the number of clusters using the Bayesian information criterion (BIC). The scale represents the fourteen different Gaussian models (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*.

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.

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 higher 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*: $P = 0.977$).

PES profiles impact the expression of genes within the candidate pathways. Each PES profile was tested for association with the peripheral blood expression of genes within their respective pathways for a subset of schizophrenia cases with expression data available – $N = 75$ ($N_{Male} = 45$, $N_{Female} = 30$), mean age of 42.21 (s.d. = 10.47), (Fig. 3, Supplementary Table 6). We tested this association using linear regression and report a standardised regression coefficient ($t = \hat{\beta} / SE$) and a FDR corrected P -value (q -value), with $FDR < 0.1$ used as the exploratory cut-off for multiple testing correction. In total, we tested 191 genes with expression data available in this sample considering all of the PES. After covariation for sex, age, and PRS_{Total} , the *NOS1* PES was associated with downregulated expression of the calcineurin subunit gene *PPP3CC* ($t = -3.08$, $P = 2.9 \times 10^{-3}$, $q = 0.05$). This was followed by the *regulation of insulin secretion* PES, which was associated with decreased expression of the syntaxin gene *STX1A* ($t = -3.5$, $P = 8.1 \times 10^{-4}$, $q = 0.055$); and the *One carbon pool by folate* gene PES, which was

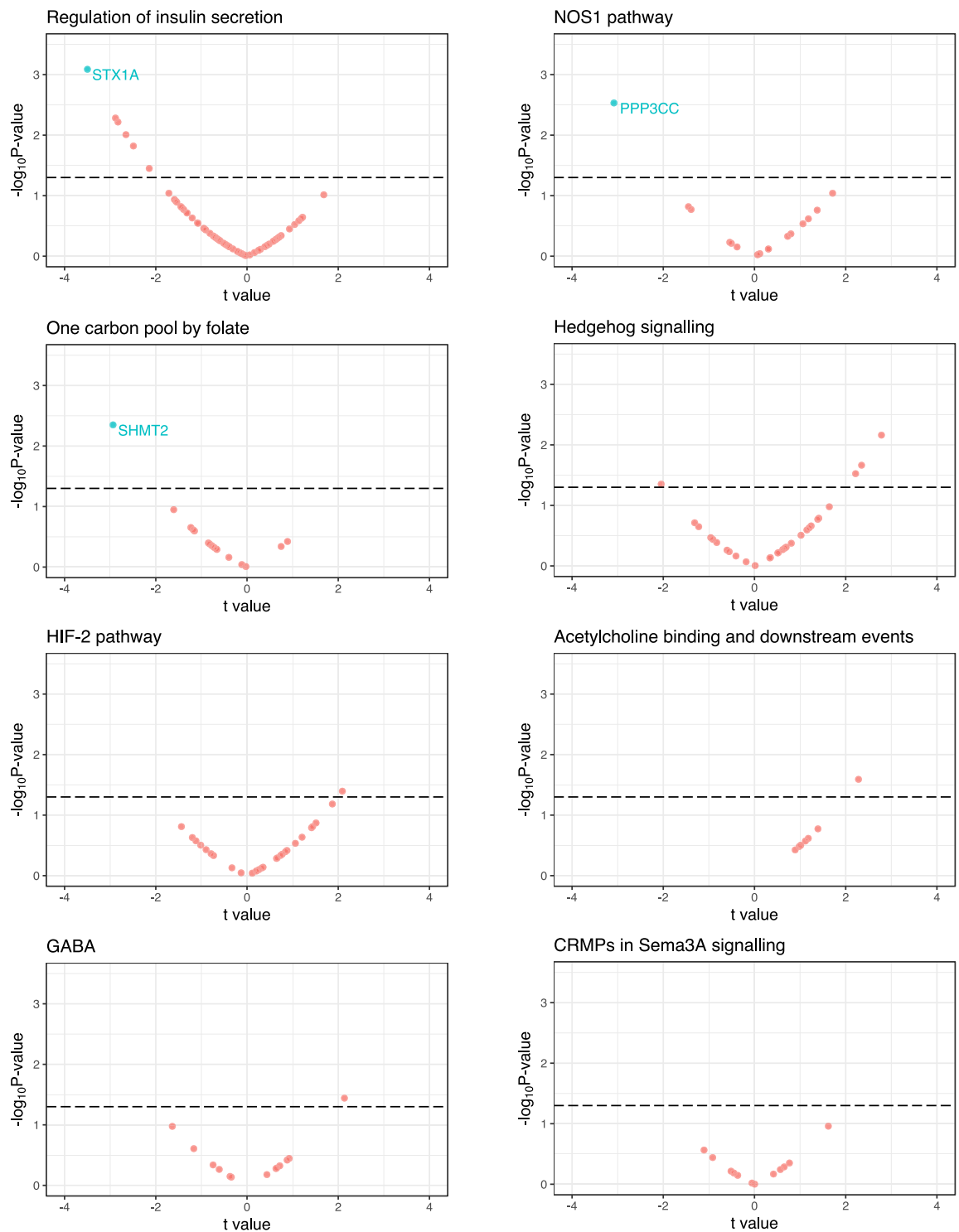


Figure 3. The relationship between PES profiles and the expression of pathway genes in peripheral blood mononuclear cells. Plot for each PES of the results of a model which investigated the association of PES with the expression of genes which comprise the PES pathway in a sample of 75 individuals with schizophrenia. The t values on the x-axes were derived from the regression model for each PES-gene pair ($t = \hat{\beta} / SE$), with the $-\log_{10} P$ value of association on the y-axes. The dotted line is indicative of uncorrected $P < 0.05$, genes highlighted blue were significant after correcting for the number of genes in the set with a liberal false discovery rate cut-off ($q < 0.1$).

associated with downregulation of serine hydroxymethyltransferase 2 (*SHMT2*; $t = -2.94$, $P = 4.5 \times 10^{-3}$, $q = 0.072$). Excluding these three genes ($q < 0.1$), there were eleven others with a nominally significant (Raw $P < 0.05$) relationship with a PES, with all PES profiles except *CRMPs in Sema3A signalling* having at least one such nominally significant model.

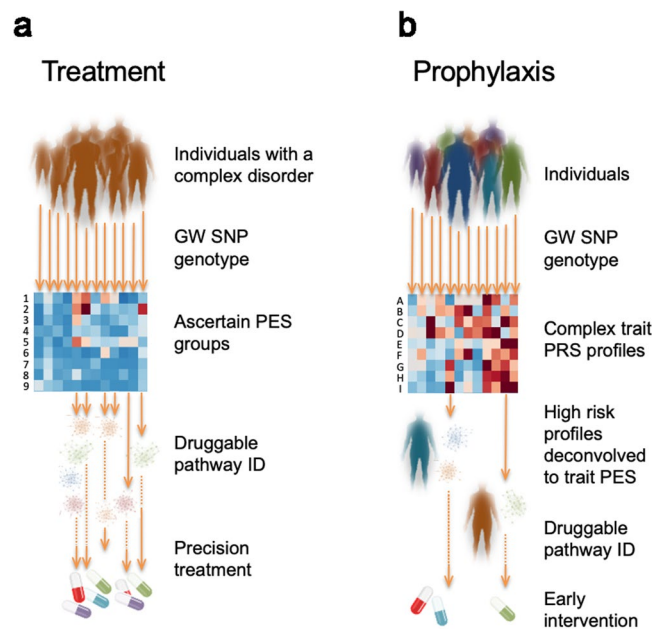


Figure 4. Implementation of *pharmagenic enrichment score* (PES) in precision treatment and prophylaxis of complex disorders. **(a)** Using the PES framework individuals with a complex disorder provide DNA for common variant SNP genotyping, which is used to ascertain individuals with a high PES. The 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.

Discussion

The drug development pipeline continues to be prohibitively expensive and time consuming in the translation of novel compounds for clinical practice^{26,27}. 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^{28,29}. 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. 4a). 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^{30,31}, 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 signalling³². 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^{19,20}. 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^{33,34}. 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 properties³⁵. Enrichment in these actionable pathways is consistent with longstanding hypotheses of deficits in neurodevelopment contributing to the aetiology of schizophrenia³⁶. There is also evidence for aetiological overlap between schizophrenia and diabetes beyond what is attributable to metabolic effects of antipsychotic treatment, which supports our identification of an insulin related pathway as a candidate PES^{37–39}.

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, N-acetylcysteine, vitamin C, Atomoxetine, and Varenicline which were identified using PES in this study^{14–16,19}. 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. However, further work is required to fully appreciate the relationship between PES profiles and PRS_{total} with both replication required for schizophrenia observations in additional cohorts and the investigation of this phenomenon for other complex traits.

In order to better understand how PES profiles could be leveraged for treatment, the effect of sequence variation which comprises the PES needs to be investigated. We outlined an example of this approach in this study by testing the effect of PES on the expression of genes which comprise each pathway in schizophrenia. Several associations between PES and mRNA expression were observed after correcting for the number of genes within the tested set. These effects may arise due to direct *cis*-acting loci within the PES and/or the downstream biologic effects of variation which effect genes with interrelated functionality. This was exemplified by the *Regulation of insulin secretion* PES, which was negatively correlated with *STX1A* expression, a syntaxin postulated to play a role in insulin homeostasis. Interestingly, *STX1A* has been shown to be positively correlated with glucose stimulated insulin secretion, suggesting downregulation conferred by the PES may have an important effect through this biological pathway^{40,41}. Similarly, downregulation of the calcineurin subunit gene *PPP3CC* was associated with the *NOS1 pathway* PES. Previous research suggests there is a bidirectional relationship between nitric oxide signalling and calcineurin, where the calcineurin subunit is both regulated by redox products and able to induce nitric oxide synthesis^{42,43}. Several other genes had suggestive association with PES profiles which may be established with a larger cohort. We suspect that this approach to validation would be particularly informative for schizophrenia in genotyped expression cohorts from brain tissue. In addition, we believe that larger cohorts need to be analysed to ascertain with sufficient power which candidate PES display the strongest association with case-status, as the variation within such pathways may be more biologically salient.

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. We believe that an analysis of the effect of PES profiles on gene expression in larger cohorts will be an important future direction of this work. The impact of candidate molecules could also be modelled in patient derived cell lines to, firstly, establish the extent of dysregulation conferred by elevated PES and, secondly, investigate the interaction with the compounds implicated. Discerning the directionality of effect may also reveal drugs which potentiate the trait-associated effect of PES, thus revealing adverse effects which may increase the risk of the disorder. Furthermore, given the observed genomic correlations amongst many complex traits it would be valuable to better understand how trait-associated risk as indexed by PES profiles also impacts closely related traits and how repurposed drugs would affect phenotypes with shared genetic architecture. 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 schizophrenia⁴⁴. In future work it may also be possible to further optimise polygenic risk scoring in the PES by implement alternative thresholding or aggregating approaches. We also believe that pathways without currently available drug targets, warrant further investigation as these may represent an unmet therapeutic niche and therefore provide a rationale for new drug development.

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. 4b). 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.

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 Fig. 1. We exploit the results of gene-set association analysis to aggregate variants from GWAS into gene-sets which may be candidates for pharmacological intervention. These gene-sets were high quality canonical and hallmark pathways sourced from the molecular signatures database (MSigDB)⁴⁵. 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})¹².

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 distribution⁴⁶. We mapped SNPs to protein coding genes (hg19, NCBI) with the genic boundaries encompassing 5 kb upstream and 1.5 kb downstream to capture variation in regulatory regions, we used a longer boundary upstream of the gene to more capture the more extensive regulatory elements before the transcription start site. Genes within the 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^{2,7,47}. 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 gene-sets 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 comprise the candidate pathway are extracted and cumulative genomic risk calculated in an analogous fashion to PRS, assuming an additive model². 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 above³. Variants were selected based on their significance (P value) for inclusion 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 gene-sets were identified after the application of multiple testing correction ($\text{FDR} < 0.05$)¹³. A minimum overlap of at least three targets overlapping the geneset for each pharmaceutical agent was also implemented. Drugs were also mapped to gene-sets 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 drug⁴⁸. Further, the tissue specificity of expression for all genes from the eight pathways was investigated using GENE2FUNC application of FUMA⁴⁹. 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)^{23,44}. Detailed descriptions of consent procedures along with inclusion and exclusion criteria for the ASRB have been extensively described elsewhere²³. The use of these data was approved by the University of Newcastle Human Ethics Research Committee (HREC) and the Australian Schizophrenia Research Bank. The Illumina Infinium Human 610 K (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 previously⁴⁴. 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 7). The use of these data was approved by the University of Newcastle Human Research Ethics 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 .

$$PES_i = \sum_j^M (\widehat{\beta}_j \times G_{ij}) \quad (1)$$

This was calculated for the individuals using the PRSice2 package⁵⁰. 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 , this PRS was an additive risk score as described in Eq. 1, with 56761 SNPs included after clumping weighted by their effect size in the GWAS. 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; including those in the 99th, 90th and 75th percentiles 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 5.4⁵¹. 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 principal components as above. After dividing the regression coefficients by their standard error to derive z , P values were calculated using the Wald Test.

Investigation of the effect of PES profiles on gene expression. We sought to investigate the relationship between PES profiles and the expression of genes which comprise their pathways in individuals with schizophrenia. A subset of schizophrenia cases in this cohort ($N = 75$) had mRNA expression data available from a previous study⁵². These participants had a mean age of 42.21 (s.d. = 10.47), whilst the majority of the subcohort was male ($N_{\text{Male}} = 45$, $N_{\text{Female}} = 30$). RNA extracted from peripheral blood mononuclear cells was profiled using Illumina HT-12_V3 BeadChips and normalised as described in Gardiner *et al.*⁵². Genes which comprise each PES pathway were extracted if they were available on the array with normalised expression values which survived quality control. The relationship between PES and the expression of each gene in that pathway as the outcome was assessed using a linear model covaried for age, sex, and PRS_{Total} for schizophrenia. These models were constructed in R version 3.4.4 using the *lm* function. Multiple testing correction was applied to each PES model individually to account for the number of genes tested in each pathway using Benjamini-Hochberg method via the *p.adjust* function in R.

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.

Received: 12 September 2019; Accepted: 3 January 2020;

Published online: 21 January 2020

References

1. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics* **101**, 5–22 (2017).
2. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–52 (2009).
3. Schizophrenia Working Group of the Psychiatric Genomics, C. *et al.* Biological Insights From 108 Schizophrenia-Associated Genetic Loci. *Nature* **511**, 421–427 (2014).
4. Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications* **9**, 2941 (2018).
5. Patel, K. A. *et al.* Type 1 Diabetes Genetic Risk Score: A Novel Tool to Discriminate Monogenic and Type 1 Diabetes. *Diabetes* **65**, 2094–2099 (2016).
6. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**, 1219–1224 (2018).
7. Abraham, G. *et al.* Genomic prediction of coronary heart disease. *European Heart Journal* **37**, 3267–3278 (2016).
8. Belsky, D. W. *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* **1**, 453–461 (2013).
9. Cleyneen, I. *et al.* Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *The Lancet* **387**, 156–167 (2016).

10. Hilker, R. *et al.* Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biological Psychiatry* **83**, 492–498 (2018).
11. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Archives of General Psychiatry* **60**, 1187–1192 (2003).
12. Oprea, T. I. *et al.* Unexplored therapeutic opportunities in the human genome. *Nat Rev Drug Discov* **17**, 317–332 (2018).
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* **45**, W130–W137 (2017).
14. Kelly, D. L. *et al.* A randomized double-blind trial of atomoxetine for cognitive impairments in 32 people with schizophrenia. *J Clin Psychiatry* **70**, 518–25 (2009).
15. Smith, R. C. *et al.* Varenicline Effects on Smoking, Cognition, and Psychiatric Symptoms in Schizophrenia: A Double-Blind Randomized Trial. *PLoS one* **11**, e0143490–e0143490 (2016).
16. Berk, M. *et al.* N-Acetyl Cysteine as a Glutathione Precursor for Schizophrenia—A Double-Blind, Randomized, Placebo-Controlled Trial. *Biological Psychiatry* **64**, 361–368 (2008).
17. Breier, A. *et al.* Effects of 12-month, double-blind N-acetyl cysteine on symptoms, cognition and brain morphology in early phase schizophrenia spectrum disorders. *Schizophrenia Research* **199**, 395–402 (2018).
18. Rapado-Castro, M. *et al.* Cognitive effects of adjunctive N-acetyl cysteine in psychosis. *Psychological Medicine* **47**, 866–876 (2017).
19. Dakhale, G. N., Khanzode, S. D., Khanzode, S. S. & Saoji, A. Supplementation of vitamin C with atypical antipsychotics reduces oxidative stress and improves the outcome of schizophrenia. *Psychopharmacology (Berl)* **182**, 494–8 (2005).
20. Bentsen, H., Osnes, K., Refsum, H., Solberg, D. K. & Bohmer, T. A randomized placebo-controlled trial of an omega-3 fatty acid and vitamins E+C in schizophrenia. *Transl Psychiatry* **3**, e335 (2013).
21. Veerman, S. R., Schulte, P. F., Smith, J. D. & de Haan, L. Memantine augmentation in clozapine-refractory schizophrenia: a randomized, double-blind, placebo-controlled crossover study. *Psychol Med* **46**, 1909–21 (2016).
22. Lieberman, J. A. *et al.* A randomized, placebo-controlled study of memantine as adjunctive treatment in patients with schizophrenia. *Neuropsychopharmacology* **34**, 1322–9 (2009).
23. Loughland, C. *et al.* Australian Schizophrenia Research Bank: a database of comprehensive clinical, endophenotypic and genetic data for aetiological studies of schizophrenia. *Aust N Z J Psychiatry* **44**, 1029–35 (2010).
24. Samara, M. T. *et al.* Efficacy, acceptability, and tolerability of antipsychotics in treatment-resistant schizophrenia: A network meta-analysis. *JAMA Psychiatry* **73**, 199–210 (2016).
25. Clemmensen, L., Vernal, D. L. & Steinhausen, H.-C. A systematic review of the long-term outcome of early onset schizophrenia. *BMC psychiatry* **12**, 150–150 (2012).
26. Smietana, K., Siatkowski, M. & Møller, M. Trends in clinical success rates. *Nature Reviews Drug Discovery* **15**, 379 (2016).
27. McNamee, L. M., Walsh, M. J. & Ledley, F. D. Timelines of translational science: From technology initiation to FDA approval. *Plos One* **12**, e0177371 (2017).
28. McGrath, J. Dissecting the Heterogeneity of Schizophrenia Outcomes. *Schizophrenia Bulletin* **34**, 247–248 (2008).
29. Deng, C. & Dean, B. Mapping the pathophysiology of schizophrenia: interactions between multiple cellular pathways. *Frontiers in Cellular Neuroscience* **7**(2013).
30. de Jonge, J. C., Vinkers, C. H., Hulshoff Pol, H. E. & Marsman, A. GABAergic Mechanisms in Schizophrenia: Linking Postmortem and *In Vivo* Studies. *Frontiers in Psychiatry* **8**, 118 (2017).
31. Scarr, E., Gibbons, A., Neo, J., Udawela, M. & Dean, B. Cholinergic connectivity: its implications for psychiatric disorders. *Frontiers in Cellular Neuroscience* **7**(2013).
32. Smeyne, M., Sladen, P., Jiao, Y., Dragatsis, I. & Smeyne, R. J. HIF1 α is Necessary for Exercise-Induced Neuroprotection while HIF2 α is Needed for Dopaminergic Neuron Survival in the Substantia Nigra pars compacta. *Neuroscience* **295**, 23–38 (2015).
33. Schmidt-Kastner, R., van Os, J., Esquivel, G., Steinbusch, H. W. M. & Rutten, B. P. F. An environmental analysis of genes associated with schizophrenia: hypoxia and vascular factors as interacting elements in the neurodevelopmental model. *Molecular Psychiatry* **17**, 1194 (2012).
34. 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* **25**, 125–137 (2011).
35. Wrasidlo, W. *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* **171**, 5757–5773 (2014).
36. Rapoport, J. L., Giedd, J. N. & Gogtay, N. Neurodevelopmental model of schizophrenia: update 2012. *Molecular psychiatry* **17**, 1228–1238 (2012).
37. Venkatasubramanian, G. *et al.* Insulin and insulin-like growth factor-1 abnormalities in antipsychotic-naïve schizophrenia. *Am J Psychiatry* **164**, 1557–60 (2007).
38. Guest, P. C. *et al.* Increased levels of circulating insulin-related peptides in first-onset, antipsychotic naïve schizophrenia patients. *Molecular Psychiatry* **15**, 118 (2010).
39. Hackinger, S. *et al.* Evidence for genetic contribution to the increased risk of type 2 diabetes in schizophrenia. *Translational Psychiatry* **8**, 252 (2018).
40. Andersson, S. A. *et al.* Reduced insulin secretion correlates with decreased expression of exocytotic genes in pancreatic islets from patients with type 2 diabetes. *Molecular and Cellular Endocrinology* **364**, 36–45 (2012).
41. Liang, T. *et al.* New Roles of Syntaxin-1A in Insulin Granule Exocytosis and Replenishment. *The Journal of biological chemistry* **292**, 2203–2216 (2017).
42. Namgaladze, D., Hofer, H. W. & Ullrich, V. Redox control of calcineurin by targeting the binuclear Fe(2+)–Zn(2+) center at the enzyme active site. *J Biol Chem* **277**, 5962–9 (2002).
43. Obasanjo-Blackshire, K. *et al.* Calcineurin regulates NFAT-dependent iNOS expression and protection of cardiomyocytes: co-operation with Src tyrosine kinase. *Cardiovasc Res* **71**, 672–83 (2006).
44. Reay, W.R. *et al.* Polygenic disruption of retinoid signalling in schizophrenia and a severe cognitive deficit subtype. *Molecular Psychiatry* (2018).
45. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems* **1**, 417–425 (2015).
46. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *Plos Computational Biology* **11**, e1004219 (2015).
47. The International Multiple Sclerosis Genetics, C. Evidence for Polygenic Susceptibility to Multiple Sclerosis—The Shape of Things to Come. *American Journal of Human Genetics* **86**, 621–625 (2010).
48. Cotto, K.C. *et al.* DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Research* **46**, D1068–D1073 (2018).
49. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).
50. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).
51. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R journal* **8**, 289–317 (2016).
52. Gardiner, E. J. *et al.* Gene expression analysis reveals schizophrenia-associated dysregulation of immune pathways in peripheral blood mononuclear cells. *Journal of Psychiatric Research* **47**, 425–437 (2013).

Acknowledgements

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). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

W.R. and M.C. designed the study and wrote the first draft of the manuscript. W.R. performed the genomic and statistical analyses. J.A. performed the variant imputation. J.A., V.C. and M.G. contributed to the interpretation of the results and the preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-57795-0>.

Correspondence and requests for materials should be addressed to M.J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020