medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Linking GWAS to pharmacological treatments for psychiatric disorders

Aurina Arnatkevičiūtė[1]✉, Alex Fornito[1], Janette Tong[1], Ken Pang[2], Ben D. Fulcher[*3], and Mark A. Bellgrove[*1]

[*]These authors contributed equally
[1]The Turner Institute for Brain and Mental Health, School of Psychological Sciences, and Monash Biomedical Imaging, Monash University, Victoria, Australia.
[2]Murdoch Children's Research Institute, Department of Paediatrics, University of Melbourne, Victoria, Australia.
[3]School of Physics, The University of Sydney, Camperdown NSW 2006, Australia.

Large-scale genome-wide association studies (GWAS) have identified multiple disease-associated genetic variations across different psychiatric disorders raising the question of how these genetic variants relate to the corresponding pharmacological treatments. Here we investigated whether functional information from a range of open bioinformatics datasets can elucidate the relationship between GWAS-identified genetic variation and the genes targeted by current drugs for psychiatric disorders. We introduce a novel measure of weighted similarity between gene targets for pharmacological treatments and GWAS risk variants for psychiatric disorders according to SNP position, gene distance on the protein interaction network (PPI), brain eQTL, and gene expression pattern across the brain. Focusing on four psychiatric disorders—ADHD, bipolar disorder, schizophrenia, and major depressive disorder—we assess relationships between the gene targets of drug treatments and GWAS hits across these weighted similarity metrics. Our results indicate that while incorporating information derived from functional bioinformatics data, such as the PPI network and spatial gene expression, revealed links for bipolar disorder, the overall correspondence between treatment targets and GWAS-implicated genes in psychiatric disorders rarely exceeds null expectations. This relatively low degree of correspondence across modalities suggests that the genetic mechanisms driving the risk for psychiatric disorders may be distinct from the pathophysiological mechanisms used for targeting symptom manifestations through pharmacological treatments and that novel approaches for understanding and treating psychiatric disorders may be required.

GWAS | genetics | treatment | pharmacology | psychiatry | PPI | eQTL | transcriptomics

Correspondence: ***aurina.arnatkeviciute@monash.edu.au***

Introduction

Advances in genotyping technologies and statistical bioinformatics over the last decade have led to the rapid growth of genome wide association studies (GWASs) across a broad range of complex diseases (1, 2). Like most complex traits, psychiatric disorders are influenced by a large number of genetic variants, each contributing only a small fraction of the overall genetic liability (3, 4). Large international collaborative efforts increased the sample sizes of GWASs to boost power for identifying robust associations, yielding multiple disease-associated genetic variations related to biologically relevant genes involved in neuronal development, neurotransmission, and plasticity (5–12).

One of the original core goals of GWAS was to inform the development of pharmacological treatments (13), under the assumption that if select genes are important in driving risk for a disorder, treatments that influence the function of these genes should mitigate risk and/or alleviate symptom expression. Current drugs for psychiatric disorders, however, were developed independently from these genomic findings, and it remains unclear whether the mechanisms of action in alleviating their symptoms relate to the genetic variation identified through GWAS. Despite advances in uncovering the genetic architecture of psychiatric disorders, the direct application of GWAS findings in developing novel pharmacological treatments remains challenging (14–17). Moreover, some psychiatric disorders—such as autism spectrum disorder, attention deficit hyperactivity disorder (ADHD), and schizophrenia (18)—are regarded as neurodevelopmental conditions. In these cases, risk genes may influence brain development in ways that create a vulnerability to illness, but these processes may be temporally separated, and distinct from the pathophysiological mechanisms that influence symptom onset and severity, and which may be more proximal and effective treatment targets. Thus, if GWAS findings implicate pathways that are targeted by existing pharmacological treatments, we can conclude that genetic mechanisms of disease risk are closely related to the mechanisms underlying their symptomatology. On the other hand, a lack of correspondence between GWAS findings and drug targets could imply either a mismatch between genetic risk mechanisms and required interventions, or the fact that the relationship between genetic liability and disease pathophysiology is more complex and involves additional biological mechanisms.

Understanding the alignment between GWAS findings and treatments is of critical importance in psychiatry, given that the design of current drugs is mainly based on neurochemical hypotheses developed in the mid-20th century, which lack supporting evidence from genetic

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

studies (19). In fact, whereas most pharmacological treatments in psychiatry work by modulating neurotransmission (20–22), GWAS studies of psychiatric disorders generally implicate genes involved in development, synaptic regulation, and plasticity (5, 7, 9). This apparent discrepancy has been proposed as a potential explanation for the relatively low efficacy of current psychiatric treatments (22–27), suggesting that the genetic mechanisms leading to the vulnerability to the disorder might differ from the pathophysiological mechanisms that influence symptom onset and severity.

Converging evidence across complex disorders supports the overall relevance of genetic data in understanding and developing pharmacological treatments. For instance, drugs with genetic support move further along the development pipeline and are more likely to be clinically successful albeit the absolute size of the effect is relatively small (28, 29). Moreover, genes that are less tolerant to mutations are more likely to be drug targets (28). GWAS-identified genes are also enriched in gene targets for drugs (30) and demonstrate enrichment for relevant treatment-related gene categories such as antipsychotics relating to schizophrenia genes (31, 32) and antidepressants, anxiolytics, and antipsychotics linked to major depression genes (33). However, only a very small fraction of current drug targets for complex disorders can be identified through GWAS studies (16, 17, 28, 34–36), indicating that the exact correspondence between GWAS findings and pharmacological treatments is limited. For instance, Finan et al. (17) demonstrated that unique GWAS associations map to treatment target genes based on SNP position, but the vast majority of those associations were discordant, meaning that treatment indication and genetic associations did not match. Even a relatively liberal approach based on pathway analysis indicated that GWAS genes for only 97 out of 182 diseases were enriched in pathways with at least one drug target (37). Most strikingly, only 29 of these 97 diseases (30%) were enriched for drug targets for the same disorder and none of those associations were identified based on GWAS alone (37).

Whereas the correspondence between GWAS findings and the respective treatment targets across a range of disorders is relatively low, incorporating functional information about gene action can provide a bridge between them. The main examples of integrating functional information to link pharmacological treatments and GWAS, however, come from studies of non-psychiatric disorders. Integrating immune-related annotations together with functional genomics, such as interactions through chromatin conformation and modulation of gene expression through eQTLs, recovered experimentally and clinically verified drug targets for immune traits and has been used to prioritize other biologically relevant genes that are not currently used as treatment targets (38). Network-based approaches quantifying functional links between genes or their products can also be used to investigate indirect associations between genes (39). For instance, GWAS-implicated genes and drug target genes have been found to couple through a protein functional interaction network that combines information from curated biological pathways, protein–protein interaction (PPI) networks, and gene co-expression, among others (34). That is, GWAS genes for a disorder tend to exhibit stronger functional interactions with each other and with treatment target genes, indicating the relative similarity of their biological function. Extensive investigations of the genetics of type 2 diabetes (40) and rheumatoid arthritis (41) also support the use of functional gene information, including interactions within the PPI network, in identifying associated drug targets as well as potential candidates for drug repositioning. Together, these findings suggest that, although the overall correspondence between the genetic origins of complex disorders and their treatment targets is relatively low, incorporating functional information has the potential to identify these links (34, 38, 40, 41), motivate new treatments (38) and prioritize existing treatments for repositioning (38, 42) [for a review see (43)]. Attempts to extend these functional genomics approaches to psychiatric disorders have been mainly limited to the analyses of transcriptomic data, with only a handful of studies identifying correspondence between GWAS and treatment targets at the level of individual genes (42, 44). Therefore, the utility of functional information derived from bioinformatic datasets to provide the link between the GWAS findings and corresponding treatments for psychiatric disorders remains unclear.

Here we introduce a novel method to evaluate the relationships between genes implicated in GWAS for psychiatric disorders and their pharmacological treatments according to several bioinformatics datasets. This allows us to determine the degree to which different sources of functional genomic information facilitate links between GWAS findings and drug targets; identify which functional data provide the most accurate correspondence; and assess whether treatments for phenotypically and genetically similar disorders show correspondence to their respective GWASs. To evaluate these associations we introduce a strong null hypothesis-testing framework that compares identified associations for a particular disorder treatments to a randomly selected set of treatments, therefore quantifying the specificity of the match. We show that, for most psychiatric disorders, the correspondence between GWAS-implicated genes and treatment targets does not exceed null expectation, suggesting that treatments for psychiatric disorders may operate via distinct pathways in alleviating symptoms, relative to the genetic causes identified using GWAS. Incorporating functional information derived from independent databases, such as protein–protein interaction networks, can uncover stronger links for some disorders, demonstrating the potential of exploiting publicly avail-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

able datasets to help understand the mechanisms underlying current treatments.

Results

Our main analysis focused on four psychiatric disorders: ADHD, schizophrenia, bipolar disorder, and major depressive disorder. We additionally used type 2 diabetes (T2D) as a non-psychiatric benchmark for comparison, as it has a well-known pathogenesis and effective treatments (45). First, we present a brief introduction to our method, followed by an investigation of different bioinformatic data modalities in identifying associations between GWAS-implicated genes and treatment targets for a given disorder. We then extend the analysis to explore whether treatments for some psychiatric disorders demonstrate associations with GWAS-implicated genes in phenotypically and genetically similar disorders and assess how data-processing choices influence the correspondence between GWAS and treatment targets.

Method Overview. To assess the correspondence between GWAS-implicated genes and drug targets, we considered a candidate set of 2155 genes that are listed as targets for all approved treatments for all disorders in the *DrugBank* database (46) [see Online methods]. For each disorder we manually curated a list of pharmacological treatments which included 14 drugs for ADHD, 29 for schizophrenia, 22 for bipolar disorder, 48 for major depressive disorder, and 45 for diabetes. Each drug had a corresponding set of gene targets based on the *DrugBank* database (version 5.1.7) (46), ranging from 63 gene targets (for ADHD) to 126 targets (for major depressive disorder) from the total candidate set of 2155 genes [see Online methods]. We then independently assigned each gene two scores for each disorder—one based on its involvement as a drug target for the disorder (as depicted in Fig. 1a), and the other based on the mapping between GWAS-implicated SNPs and genes according to one of four bioinformatic data modalities (as depicted in Figs. 1b–e).

First, each gene was scored according to the specificity of each drug, where genes targeted by high specificity drugs were assigned proportionally higher values [see Online methods]. Gene weights across all drugs for a disorder were then summed, producing a single treatment-based score vector, s^{drug}, that independently measured the involvement of each gene as a treatment target for that disorder (Fig. 1a). As a result, genes that are targeted by more drugs with high specificity were assigned higher values compared to genes that are targets for fewer drugs or drugs with low specificity [see Online methods]. Genes that do not act as targets for that disorder were assigned a score of zero and consequently did not contribute to the similarity between GWAS-implicated genes and treatment targets.

The GWAS-based score, denoted s^{GWAS}, quantified a gene's correspondence to variants implicated in a given GWAS, where similarity was defined with respect to one of four definitions: (i) the base-pair distance from the SNP on the DNA; (ii) its topological proximity to the GWAS-implicated genes in the PPI network; (iii) its brain eQTLs; and (iv) its spatial pattern of gene expression across the brain (47) (Fig. 1b-e). Each of these methods captures a distinct biological mechanism through which SNPs identified in a GWAS might influence genes. For instance, the simplest approach was based on mapping SNPs to their corresponding genes according to their DNA position, such that genes with the highest number of significant SNPs get higher scores (Fig. 1b) [see Online methods]. This method relies on the assumption that SNPs located within protein-coding regions of a gene can potentially affect that gene by changing its functional products whereas SNPs in non-coding regions or in linkage disequilibrium (LD) with a particular gene can be involved in its regulatory mechanisms (48, 49). Proteins encoded by specific genes exert their function through interactions with one another, therefore, indirect relationships between genes can be identified using functional information such as PPI networks (39). In this case, we also investigated the direct neighbors of those 2155 genes in the PPI network and calculated the proportion of these neighbors that were implicated in a GWAS (Fig. 1c) [see Online methods]. As a result, we could quantify the degree to which a particular gene is associated with the GWAS-implicated genes through functional interactions. SNPs that influence the expression of one or more genes are referred to as expression quantitative trait loci (eQTL) and can be located in close proximity or at a distance from that gene. Therefore, evaluating the extent to which GWAS-identified SNPs act as eQTLs in brain tissue allows quantification of the functional impact of these SNPs on the genes of interest (Fig. 1d) [see Online methods]. Considering that psychiatric disorders are associated with altered gene expression in the brain (50), and assuming that spatial gene-expression patterns are related to the mechanism of drug action (51), we also incorporated spatial gene-expression data derived from the Allen Human Brain Atlas (AHBA) (47). Each gene was scored according its spatial expression similarity to GWAS-implicated genes, giving higher scores to genes whose expression patterns were more strongly correlated to genes identified through GWAS (Fig. 1e) [see Online methods].

After scoring genes both according to their involvement in current treatments for a disorder (s^{drug}) and to their implication by GWAS (s^{GWAS}), we developed a method to assess the correspondence between the the two sets of scores (and hence between the GWAS-implicated genes for a disorder and the genes targeted by currently approved pharmaceutical treatments for it). We first normalized each of the score vectors across genes and calculated the inner product between these two unit-normalized vectors, resulting in a *weighted similarity score*, ρ. As a result, only genes that were assigned non-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

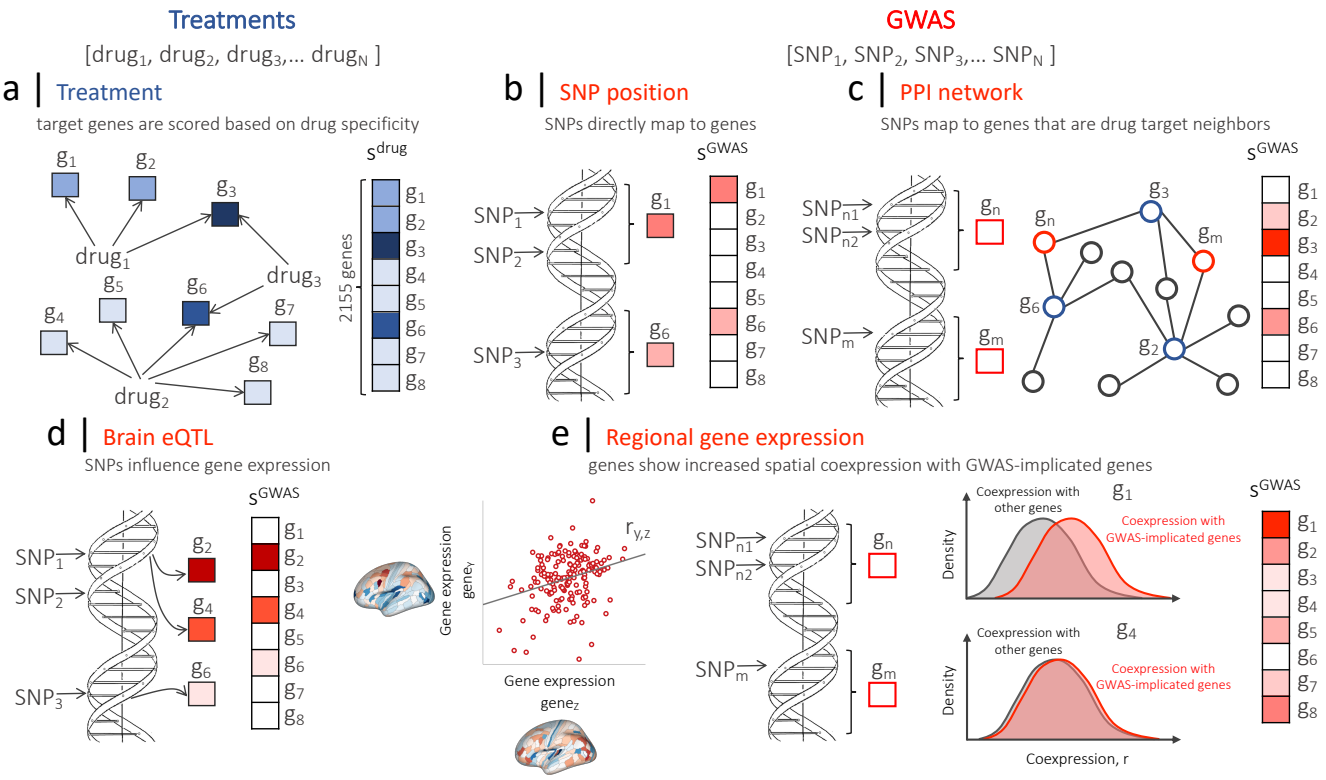

Fig. 1. We assess relationships between drug targets and GWAS by scoring genes based on: (i) their involvement in approved pharmaceutical treatments; and (ii) GWAS, for a given psychiatric disorder. (a) The analysis is based on a set of for all clinically approved treatments. For each disorder, each of 2155 target genes derived from the *DrugBank* database (46) are assigned scores corresponding to their role as a target across drugs approved for use in treating that disorder. A gene is scored based on its specificity for each drug by assigning a value of $w = 1/L$, where L is the total number of targets for that drug. The total score for a gene is quantified as a sum of scores across drugs. For each disorder, the same set of 2155 genes is also scored based on their similarity to variants implicated in corresponding GWASs using four main methods for mapping SNPs to genes: **(b)** SNP position, where the SNPs are mapped to genes based on their position on the DNA and the gene score consequently proportional to the cumulative effect size of all SNPs mapped to that gene; **(c)** PPI network, where genes are scored based on the proportion of their direct neighbors that are implicated in GWAS (as mapped based on SNP position); **(d)** brain eQTL, where genes are scored based on the cumulative impact of GWAS-identified SNPs on influencing their expression; **(e)** regional gene-expression similarity calculated from the Allen Human Brain Atlas (AHBA) (47). First, for each pair of genes in the AHBA, the similarity of their spatial gene-expression profiles is quantified using a measure of gene–gene coexpression. Then, GWAS-implicated genes are identified based on SNP position and for each of 2155 target genes, coexpression value distributions are compared between GWAS-implicated and all other genes. As a result, a gene receives a high score if its coexpression with GWAS-implicated genes is, on average, higher than its coexpression with all other genes.

zero scores for both s^{drug} and s^{GWAS} contributed to this similarity score.

In contrast to previously used approaches that have applied relatively lenient significance testing (e.g. comparing a set of GWAS-implicated genes to all available genes) (17, 28, 34, 41), here we evaluated the significance of a given weighted similarity score, ρ_{emp}, using a permutation test comparing it to an ensemble of 5000 null values, $\boldsymbol{\rho}_{\text{rand}}$, generated by independently selecting a set of random drugs (while preserving the number of drugs associated with the disorder) [see Online methods]. As a result, a significant association indicates that GWAS-implicated genes match treatment targets for a selected disorder over and above that is expected by a random set of currently approved drugs as defined in the DrugBank database.

Correspondence between GWAS-implicated genes and treatment targets. We first investigated whether current treatments for a given psychiatric disorder have gene targets that are more similar to their corresponding GWAS than a random set of drugs for each of the four mapping methods separately (Fig. 1b–e). Figure 2 summarizes the results for all four scoring methods, where the degree of correspondence between s^{drug} and s^{GWAS} for each disorder is quantified using a p-value computed as a permutation test from our random-drug null model [see Online methods], plotted as $-\log_{10}(p)$.

Here we find that mapping GWAS results to genes via the PPI network identified a significant correspondence for both diabetes ($p^{\text{perm}} = 0.003$) and bipolar disorder ($p^{\text{perm}} = 0.008$) (Fig. 2b, permutation test against randomly selected drugs, statistically significant following Bonferroni correction for $n = 5$ tests), but no associations for schizophrenia, ADHD, or major depression. To shed light on the potential reasons why bipolar disorder was the only psychiatric disorder that displayed an GWAS–drug association, we performed a gene ontology (GO) enrichment analysis on both treatment and GWAS-based scores. It indicated that treatment targets for all four psychiatric disorders are consistently enriched for GO categories related to synaptic signaling ([refer to the supplementary files]). In contrast, the number of synaptic signaling-related categories for

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

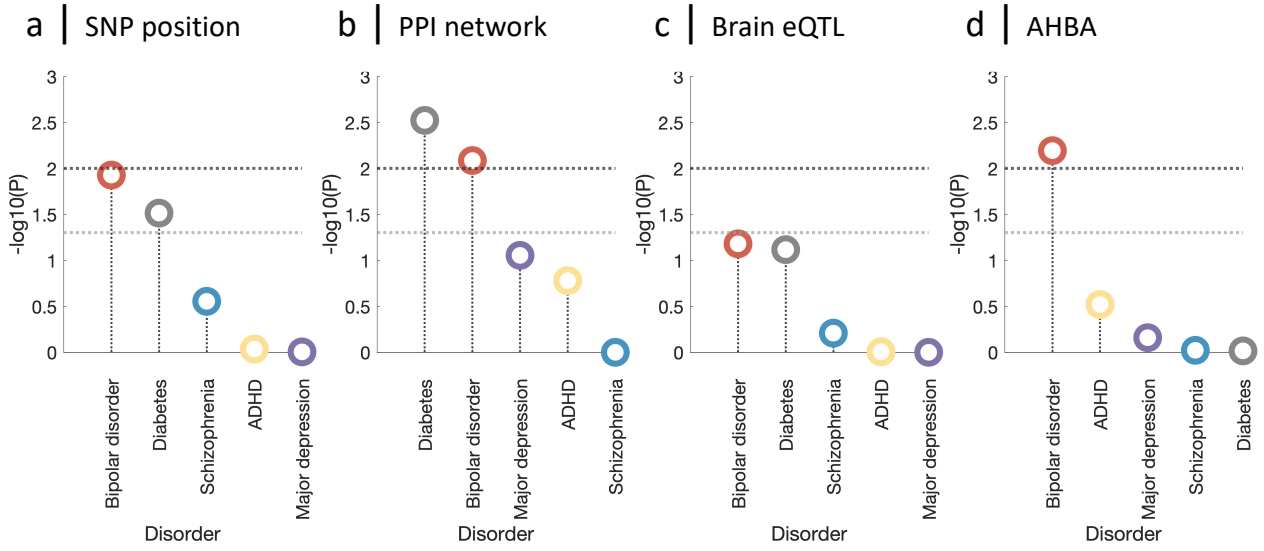

Fig. 2. Bipolar disorder and diabetes demonstrate correspondence between treatment targets and genes implicated in GWAS data for each disorder. For a given disorder and mapping method, circles represent the significance of the association between the gene scores obtained from a set of drugs, s^{drug}, and GWAS data, s^{GWAS}, computed as a permutation test relative to random drugs and plotted as $-\log_{10}(p)$ [see Online methods]. Subplots correspond to the four GWAS mapping methods: **(a)** SNP position, **(b)** PPI network, **(c)** Brain eQTL, **(d)** Gene-expression similarity (cf. Figs 1b–e). Higher $-\log_{10}(p)$ values indicate a stronger correspondence between GWAS and drug targets for a given disorder than expected from random treatments. Circles are colored by the disorder they represent (as labeled). Horizontal lines represent p-value thresholds: dark gray line – $p = 0.01$ (significant at $\alpha = 0.05$ after Bonferroni correction for five disorders); light gray line – $p = 0.05$.

GWAS-implicated genes based on the PPI mapping was very low for all disorders with the exception of bipolar disorder, where 7 of the top 20 categories involve G protein-coupled receptor signaling pathways and other signaling-related processes. These results suggest that the correspondence between treatment targets and GWAS-implicated genes for bipolar disorder is likely to occur due to the functional role of GWAS-implicated genes rather than the unique targets of bipolar disorder drugs.

For regional gene-expression-based mapping, the correspondence between drug targets and GWAS genes also significantly exceeded null expectation for bipolar disorder (Fig. 2d, $p^{\mathrm{perm}} = 0.006$, statistically significant following Bonferroni correction for $n = 5$ tests), but no associations were observed for any other disorder. Positional mapping (Fig. 2a) indicated a relatively strong ($p^{\mathrm{perm}} < 0.05$), but non-significant correspondence ($p^{\mathrm{perm}} > 0.01$) for both diabetes and bipolar disorder. Associations for brain eQTL-based mapping did not exceed null expectation for any disorder (Fig. 2c), indicating that the direct effect of SNPs on the cortical gene expression was not informative for matching GWAS variants to current treatment targets.

Our proposed method, which scores individual genes independently and then assesses the similarity in these scores, allows us to quantify the contribution of individual genes to the correspondence between GWAS and treatment-based scores, and thus delineate which genes drive the associations between the two sets of scores. The contribution of each gene to the similarity score, ρ, was quantified by comparing the inner product of s_i^{drug} and s_i^{GWAS} in real data to the distribution of inner product values derived from a null model [see Online methods]. As a result, genes are assigned p-values by the extent to which they contribute to the similarity score, ρ relative to the null expectation. To understand the functional roles of the most strongly implicated genes, we next investigated the associations identified for bipolar disorder and diabetes via PPI. We found that genes with the strongest involvement for bipolar disorder play roles in neurotransmission (*ADRA2C*, *CKS1B*, *GRIA3*, *GSK3B*, *HTR5A*, *HTR6*, *HTR7*, $p^{\mathrm{corr}} < 2.3 \times 10^{-5}$, Bonferroni correction for $n = 2155$ tests), whereas genes implicated in diabetes were predominantly involved in glucose metabolism and insulin secretion (e.g., *DPP4*, *GLP1R*, $p^{\mathrm{corr}} < 2.3 \times 10^{-5}$, Bonferroni correction for $n = 2155$ tests) [see Online methods, Supplementary Table S1]. The fact that we identified genes with distinct and specific functional roles for both disorders was partially predetermined by their treatment targets (as genes are scored based on the combined weight across both measures). However, these genes contributed to the similarity between GWAS and treatment targets over and above the expectation based on the null model indicating their specific contribution.

We next performed an exploratory analysis aiming to assess whether treatments for some psychiatric disorders overlap with the GWAS-identified genetic architecture of phenotypically and genetically similar disorders. To achieve this, we expanded the comparison to all pairwise combinations of GWAS and treatment scores. We found that, except for the previously identified associations for bipolar disorder and diabetes (Figs. S1b (IV,V), d (IV)), most other disorder pairs and mapping methods did not show a significant correspondence after correcting for

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

multiple comparisons. The strongest associations were identified for bipolar disorder, major depression, and schizophrenia. Treatment targets for bipolar disorder were linked to genes implicated in both major depression ($p^{\text{perm}} < 0.01$) and schizophrenia ($0.01 < p^{\text{perm}} < 0.05$) (Figs S1a (II, III), b (II, III)), whereas treatment targets for schizophrenia were related to major depression GWAS genes ($0.01 < p^{\text{perm}} < 0.05$) (Fig. S1a). Although not statistically significant, these cross-disorder associations are consistent with the well-described similarities in phenotypic manifestations and genetic architectures of schizophrenia, bipolar disorder, and major depression (52).

The impact of data-processing choices. The analyses presented above were based on a selected set of GWAS-to-gene scoring methods. To better understand the specificity of these results, and identify whether different data-processing choices could improve the correspondence between GWAS and current treatment targets, we explored other approaches for GWAS-to-drug target mapping. Specifically, we investigated five types of changes. First, we examined the effect of representing the PPI network at varying densities, by differently thresholding the pairwise confidence scores of protein–protein interactions: from 0 (the network includes all possible interactions), to 400 (where weakest interactions are not included), to 600 (used above), to 900 (only the interactions with the strongest evidence are kept) [see Online methods]. Second, we investigated the effect of scoring genes based on the total number of interacting neighbors that are implicated in the GWAS (rather than normalizing by the number of neighbors and computing a proportion of neighbors, as above) [see Online methods]. Third, in addition to selecting the initial set of genes for regional gene expression and PPI-based analyses using SNP position-based mapping ('PPI position', 'AHBA position'), we also performed these analyses using the initial set of genes based on brain eQTL mapping for these methods ('PPI eQTL brain', 'AHBA eQTL'). Fourth, we expanded brain-based eQTLs above to include other tissues such as blood, liver, and pancreas (hypothesizing that these might boost associations with T2D). Fifth, we introduced a novel mapping method based on long-range chromatin interactions (Hi-c) (53) in human brain tissues across two developmental time points, allowing us to identify developmentally specific genes [see Online methods]. These modifications resulted in a total of 27 methods of mapping from GWAS SNPs to gene scores: via SNP position (1 measure), PPI network (16 measures), eQTL (4 measures), chromatin interactions (4 measures), and spatial transcriptomics (2 measures).

As shown in Fig. S2, gene scores derived from different mapping methods showed varying levels of similarity. For example, PPI-based measures were generally more similar to each other than to Hi-c or eQTL-based scores, whereas regional gene expression was not similar to other mapping methods (Fig. S2). Similarly to the previous analyses, the correspondence between gene scores obtained from current treatments, s^{drug}, and GWAS, s^{GWAS} for each mapping method and disorder was quantified with a p-value (relative to an ensemble of random-drug nulls as above, while adjusting for 27 measures using Bonferroni correction), and shown in Fig. 3. The degree of correspondence between GWAS-implicated genes and treatment targets was markedly variable across disorders and mapping methods. For instance, none of the individual mapping methods yielded significant matches for ADHD, major depressive disorder, or schizophrenia (Fig. 3a-c), whereas bipolar disorder and diabetes exhibited a significant correspondence for several of PPI-based measures (Fig. 3d,e). Moreover, PPI-based measures using different confidence thresholds ('PPI position th900' and 'PPI position th0') outperformed the original measure ('PPI position th600', Fig. 3d,e). Consistent with the initial findings, gene scores assigned from the PPI network were generally the most informative about treatment targets for both bipolar disorder and diabetes, suggesting that indirect interactions through gene products might be involved in the mechanism of pharmacological treatments.

In addition to investigating each method in isolation, we evaluated whether different mapping methods can provide complementary information, resulting in an increased correspondence to treatment targets. To this end, we computed a single score using all 27 available measures as their linear combination by fitting a linear model using all available scores. The null model was also adjusted to incorporate the linear regression across all measures in each iteration to account for this optimization step. The results are shown in Fig. 3 using red circles. For ADHD, major depressive disorder, and diabetes, the combined score did not outperform individual measures, but provided a statistically significant correspondence for bipolar disorder ($p^{\text{perm}} = 4 \times 10^{-4}$, Bonferroni correction for 27 measures), and an improvement relative to all other measures for schizophrenia ($p^{\text{perm}} = 0.003$, slightly above the Bonferroni-corrected threshold, $p = 0.002$). Whereas our results are mixed, the example of schizophrenia demonstrates that complementary information from multiple bioinformatic sources can sometimes be combined to substantially improve the correspondence between GWAS and treatment targets.

To test the robustness of our general findings, we repeated the analysis using a different set of GWAS summary statistics data for all psychiatric disorders (5, 7, 9, 11). In this replication, we obtained results that were consistent with the initial findings (Fig. S4, Fig. S5, Fig. S6). To assess if our findings are specific to psychiatric disorders, we also evaluated the correspondence between treatment targets and genes implicated by GWAS for several non-psychiatric diseases such as heart failure, rheumatoid arthritis, and inflammatory bowl disease. We found that, in most cases, individ-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

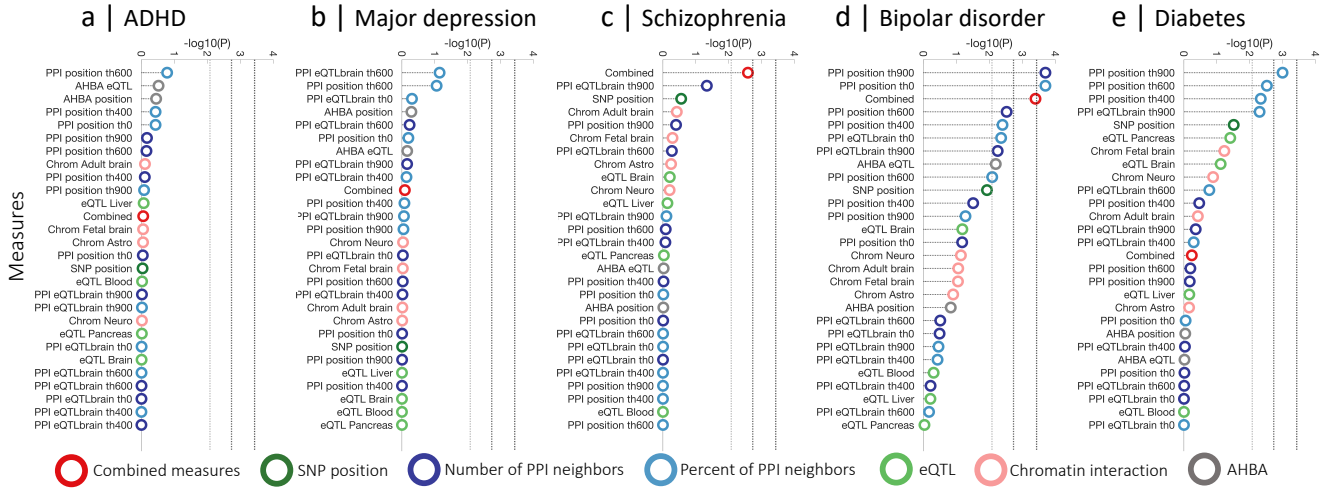

Fig. 3. PPI-based mapping methods show the strongest correspondence between treatment targets and genes implicated in a GWAS for bipolar disorder and diabetes. Circles represent the significance of the association between the treatment-based score, s^{drug}, and GWAS-based score, s^{GWAS}, for a selected disorder. The significance is quantified using a permutation-based p-value comparing the empirical GWAS-treatment matching score ρ to a set of 5000 matching scores generated using a selection of random drugs [see Online methods]. The position of the circle corresponds to $-\log_{10}(p)$ where higher values indicate more significant association. Subplots correspond to different disorders: **(a)** ADHD, **(b)** major depression, **(c)** schizophrenia, **(d)** bipolar disorder and **(e)** diabetes. Colors indicate different types of mapping methods: SNP position (dark green); spatial transcriptomic similarity (gray); chromatin interaction, Hi-c (pink); PPI network quantified as the proportion of neighbors (light blue); PPI network quantified as the total number of neighbors (dark blue); and eQTL (light green); linear combination of all measures (red). Horizontal lines represent the degree of statistical significance: light gray line – $p^{\mathrm{perm}} = 8.3 \times 10^{-3}$ (Bonferroni correction for six types of measures, this threshold provides guidance for the expected significance level considering that groups of measures derived from the same type of mapping method show a degree of similarity); dark gray line – $p^{\mathrm{perm}} = 1.9 \times 10^{-3}$ (Bonferroni correction for 27 measures); black line – $p^{\mathrm{perm}} = 3.7 \times 10^{-4}$ (Bonferroni correction for 27 measures and 5 disorders, this threshold provides guidance for a highly conservative correction assuming independence among all mapping methods and disorders). Permutation-based approach estimates a p-value with a minimum resolution of 2×10^{-4} (corresponding to $1/5000$). If the estimated p-value is lower, we conservatively place it at 2×10^{-4}.

ual mapping methods as well as the combined score did not provide significant correspondence between GWAS-implicated genes and treatment targets, suggesting poor overall correspondence for these non-psychiatric disorders (Fig. S3).

Discussion

Large-scale GWAS have led to significant developments in understanding the genetics of complex disorders, including a range of psychiatric conditions (5–12, 54). However, as current pharmacological treatments for psychiatric disorders were developed independently from these genomic findings, we aimed to investigate whether the mechanisms of action of these treatments relate to the disorder-associated genetic variation identified through GWAS. Under the general assumption that the genetic mechanisms driving the risk for a disorder are also implicated in its symptom expression and consequently can be used in pharmacological interventions, one of the main goals of GWAS research was to inform potential gene targets. However, the possibility that risk genes for psychiatric disorders may act by creating vulnerability to illness during development through pathways that are distinct from the pathophysiological mechanisms influencing symptom onset and severity is commonly overlooked. To test this, we introduced a new method for assessing the correspondence between genetic variation implicated in GWAS for complex disorders and their pharmacological treatment targets using a range of different types of biological information.

In line with previous findings (16, 17, 28, 34–37), we found that the overall degree of correspondence between GWAS-implicated genes for psychiatric disorders and their respective treatment targets across different mapping methods is low; only bipolar disorder exhibited a significant associations, that were detectable when incorporating functional information derived from the PPI network and, at a lower significance threshold, spatial gene expression. Importantly, these results were not strongly dependent on the size of the corresponding GWASs used in the analyses (Fig. S4, Fig. S5, Fig. S6), indicating that the lack of the associations does not simply result from insufficient power to identify relevant genetic variants. Considering the comprehensive nature of our analyses that investigated a range of different data modalities—including information derived from PPI network, eQTLs, chromatin interactions, and spatial transcriptomics—our findings suggest that genes giving rise to disease pathophysiology might differ from the systems that are currently targeted by their pharmacological treatments. This notion is also supported by the fact that a large fraction of genes identified through GWASs relate to neurodevelopmental processes such as presynaptic and neuron differentiation, neuronal morphogenesis and projection (9), neuronal development and synapse formation (7, 8), as well as synaptic plasticity (5, 10), whereas targets for pharmacological treatments mostly involve genes for synaptic signaling indicating a likely disconnect between the genetic risk for the disorders and the genetic mechanisms targeted by pharmacological treatments. Our re-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

sults demonstrate that, among different data-processing options, PPI network-based mapping methods consistently yielded the strongest GWAS–drug associations (Fig. 3d,e), supporting the idea that the concordance between GWAS and pharmacological treatments is likely to involve interactions between the corresponding gene products (34, 40, 41).

Besides investigating the concordance between GWAS and treatment targets for individual disorders, our approach also allowed us to directly test cross-disorder correspondence—namely, whether treatment targets for one disorder match genes implicated in the GWAS of others. The suggestive associations identified between treatments for bipolar disorder and genes implicated in both major depression and schizophrenia are consistent with the general profile of bipolar disorder treatments as they target some of the overlapping symptoms between these disorders, such as depressive moods and psychotic episodes. Similarity in the genetic architecture between bipolar disorder, major depression, and schizophrenia, quantified through the genetic correlation (52) as well as common biological pathways (55), is in line with the observed coupling between GWAS and treatment targets. Although not statistically significant (after accounting for multiple testing), these findings are in line with the results from a drug repurposing study by So et al. (42) indicating that repurposing candidates for bipolar disorder included a number of antipsychotics and antidepressants, whereas antipsychotics were commonly found among top hits for major depression.

In contrast to previous studies which mainly evaluated significance with respect to gene set and pathway enrichment (28, 31–33, 56), or compare the identified correspondences to what could be expected using all available genes (15, 17, 34), our approach introduces a more appropriate null model for evaluating associations based on other available treatments. In other words, instead of asking whether the correspondence between GWAS-identified genes and treatment targets exceeds the expectation compared to the rest of the genome, our approach addresses the specificity of those associations in the context of all available treatments by quantifying if the GWAS-implicated genes for a given disorder match its treatments targets to a higher extent than other drugs. Therefore, even if some SNPs relate to genes that serve as treatment targets for relevant drugs, that does not necessarily correspond to a significant association in the context of all available treatments. Based on this null model, GWAS-implicated genes for a particular disorder are expected to demonstrate a specific coupling with corresponding treatments that exceed the associations with treatments for unrelated disorders. Our proposed method also offers an advantage of tailoring the null model based on a research question by expanding or restricting the range of treatment categories used in generating the null distribution, ρ_{rand}. For instance, when focusing on psychiatric disorders, we can restrict the null model to psychiatric treatments or adjust their categories based on other criteria (Fig. S7). The key advantage of such a modification is the ability to refine the testing based on the relevance of selected drugs to evaluate and interpret the correspondence with respect to a plausible candidate set of treatments.

Our methods for identifying GWAS-implicated genes and evaluating the correspondence to treatment targets can be easily used to investigate other complex disorders and to uncover potential cross-disorder associations. Moreover, ranking individual genes based on their contribution to cross-disorder coupling could provide the means for prioritizing genes as potential treatment targets for selected disorders. With improvements in quantifying eQTLs, as well as the development of spatially comprehensive time-resolved and cell-specific gene-expression atlases, these methods could be used to investigate a range of complex disorders and uncover the links between the genetic variation influencing the liability for a disorder and pathophysiology.

Online methods

We first describe the selection of potential treatment targets, outline the scoring methods for both treatment and GWAS-based measures, and define the methods for computing similarity between those scores and evaluating the statistical significance. Code for reproducing all analyses described here is available at `https://github.com/AurinaBMH/GWAS_drugs`.

A. Target selection. The list of genes used in all analyses is based on the combined set of all approved treatment targets from the *DrugBank* database `https://go.drugbank.com/`, version 5.1.7, downloaded on the 3rd of August 2020. We selected all unique target genes annotated to the human species, resulting in a set, $\mathcal{G}$, containing 2155 genes. Each gene in $\mathcal{G}$ is then assigned a score, g based on different independent criteria in relation to pharmacological treatments or GWASs. Scores assigned to all genes in $\mathcal{G}$ can be represented as a score vector, s_i $(i = 1, 2, ..., |\mathcal{G}|)$. Below we describe procedures for generating an s_i according to treatment and GWAS-based scoring.

B. Treatment-based scoring. Our main analyses focus on four psychiatric disorders: ADHD, schizophrenia, major depression, and bipolar disorder. We use type 2 diabetes (referred as diabetes through the manuscript) as a benchmark for comparison, as a non-psychiatric disease that has a relatively well-known pathogenesis and effective treatments. Gene targets for each drug were assigned based on the *DrugBank* database (version 5.1.7). The curated list of pharmacological treatments for those disorders includes: ADHD (14 drugs with 63 unique gene targets); schizophrenia (29 drugs, 78 gene targets); major depression (48 drugs, 126 gene targets); bipolar disorder (22 drugs, 110 gene targets);

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

diabetes (45 drugs, 103 gene targets). Later, we extend the analysis to other non-psychiatric conditions such as heart failure (46 drugs, 114 gene targets), rheumatoid arthritis (58 drugs, 111 gene targets) and inflammatory bowl disease (26 drugs, 40 gene targets). Treatments for different conditions of interest were selected by searching `www.drugbank.ca` (accessed September 3, 2020). Specifically, drugs for each indication were searched in the DrugBank database using the following search terms: 'heart failure' (for heart failure); 'diabetes' (for type 2 diabetes) excluding 'type I diabetes' and 'diabetes insipidus'; 'schizophrenia' (for schizophrenia); 'bipolar' (for bipolar disorder) excluding 'bipolar depression'; 'major depression' (for major depression); 'attention deficit' (for ADHD); 'Crohn's' and 'ulcerative colitis' (for inflammatory bowel disease); and 'rheumatoid arthritis' (for rheumatoid arthritis).

Using this information, we constructed a score vector, s_i^{drug}, for each disorder across 2155 genes that quantifies the involvement of each gene as a treatment target for that disorder. To ensure that each drug was weighted equally, genes were scored with respect to the total number of targets for that drug: if a drug has L target genes, then the score for each gene is incremented by an amount $1/L$. As a result, if a drug is very specific and has only a few targets, each of those target genes will be assigned a relatively high score compared to a less specific drug targeting multiple genes. The total treatment-based score, s_i, for gene i, is then computed as the sum of weights across all drugs. Genes with higher scores, s_i, for a given disorder are those for which more drugs target them with high specificity; conversely, scores are minimal (0), for genes that are not targeted by any drugs used for treating a given disorder.

C. GWAS-based scoring. For each disorder GWAS [ADHD (8), schizophrenia (6), major depression (10), bipolar disorder (54), diabetes (57), heart failure (58), rheumatoid arthritis (41), and inflammatory bowl disease (59)], we constructed a score vector, s_i^{GWAS}, across 2155 genes that quantified the similarity of each gene to variants identified in each GWAS. The similarity was defined based on four main bioinformatic datasets including SNP position on the DNA, protein interaction network (PPI), brain eQTL, and regional gene expression across the brain as quantified using Allen Human Brain Atlas (AHBA) (47). Later, we expand GWAS-based mapping methods to include non-brain eQTLs, brain chromatin interaction profiles (Hi-c) and explored the effects of different data-processing options.

C.1. SNP position-based GWAS scoring. For a given disorder's GWAS, each SNP was mapped to a corresponding gene, based on its position on the DNA using MAGMA software package (`https://ctg.cncr.nl/software/magma`) (60). Corresponding annotation file `MAGMAdefault.genes.annot` and reference panel (1000 Genomes European ancestry) for linkage disequilibrium (LD) calculation were downloaded from `https://github.com/thewonlab/H-MAGMA`. Gene analyses in MAGMA are based on a multiple linear principal components regression model where the gene p-values are computed using an F-test. A score vector $s^{\text{GWAS-position}}$ was constructed using $-\log_{10}$ transformed gene p-values for each of 2155 genes $s_i^{\text{GWAS-position}} = -\log_{10}(P_i)$. Therefore, genes with a more significant involvement in a GWAS were assigned higher values. Genes that were absent from the MAGMA output were assigned a 0 score.

C.2. PPI-based GWAS scoring. We developed a gene-scoring procedure to incorporate information derived from the protein–protein interaction (PPI) network. PPI data were downloaded from the STRING database (version 11.0) on the 24th of June 2020 `https://string-db.org/cgi/download.pl?sessionId=a1fHJhN5R9Md&species_text=Homo+sapiens`. Protein interactions were quantified using confidence scores ranging from 150 to 999, with higher values assigned to interactions with stronger evidence. Applying different thresholds to those confidence scores allows rendering binary PPI networks where interactions exceeding a selected threshold are considered present while others are treated as absent. We selected a range of confidence score thresholds—0, 400, 600 and 900—yielding four binary PPI networks with varying densities. These networks were transformed into PPI distance matrices that capture the shortest path between all pairs of nodes using `distance_bin` function from the Brain Connectivity Toolbox (61). We then matched proteins to their corresponding genes through `biomaRt` package in `Bioconductor`. The main analyses here are performed using PPI distance matrix generated from relatively high-confidence interactions (>600) providing a balance between sensitivity and specificity. Although note that we investigate the effect of this threshold in the impact of data-processing choices section.

The construction of PPI-based GWAS scores relied on the selection of genes implicated in each GWAS. To identify a list of genes for each GWAS, we controlled the family-wise error at 0.05 using Bonferroni correction (by setting a gene p-value threshold equal to 0.05 divided by the number of identified genes in MAGMA analysis). The number of identified genes for psychiatric disorder GWASs ranged from 30 for ADHD to 479 genes for schizophrenia.

Our PPI-based scoring method assigned each of 2155 target genes (g_i) a score according to its vicinity to G_{GWAS} genes on the PPI network along paths of length k. We started the analyses considering one-step paths, $k=1$ and for each of 2155 target genes identified all genes that directly interact with gene g_i ($k=1$) resulting in a gene set, G_{int}. A subset of these genes were then labeled as being implicated in GWAS, $G_{\text{GWAS}} \in G_{\text{int}}$. The absolute number of overlapping genes for each target g_i ($G_{\text{GWAS}} \in G_{\text{int}}$) is proportional to the total number of its neighbors. Therefore, we scored each gene as the

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

proportion of their interacting neighbors that are implicated in GWAS: $s_i^{\mathrm{GWAS-PPI}} = |G_{\mathrm{GWAS}}|/|G_{\mathrm{int}}|$. Repeating the process across all 2155 genes we constructed the full score vector, $s^{\mathrm{GWAS-PPI}}$. Following this general procedure, a score vector can be computed at any given path-length, k (all genes within k steps on the PPI network are included in G_{int}). Exploratory analyses indicated that increasing k above 1 significantly reduces the specificity of the results, therefore all PPI-based analyses presented in this manuscript consider only 1-step neighbors ($k = 1$).

C.3. eQTL-based GWAS scoring. Using eMAGMA we constructed eQTL-based gene scores ($s^{\mathrm{GWAS-eQTL}}$) quantifying each of 2155 genes based on their involvement in tissue-specific gene expression. eMAGMA is a validated pipeline based on the original MAGMA software that uses tissue-specific SNP-gene associations to assign SNPs to genes based on their association with gene expression. The SNP–gene associations are then aggregated in a gene-based test, while adjusting for linkage disequilibrium and correlated gene expression (62). Our initial analyses were based on brain eQTL information derived from the psychENCODE database (63) and then expanded to other tissues including liver, whole blood and pancreas with corresponding annotation files derived from GTEx database (64) (version 8). Similarly to position-based gene scoring, a score vector $s^{\mathrm{GWAS-eQTL}}$ was constructed using $-\log_{10}$ transformed gene p-values for each of 2155 genes $s_i^{\mathrm{GWAS-eQTL}} = -log10(P_i)$. Genes that were absent from the eMAGMA output were assigned a score of 0.

C.4. Regional expression-based GWAS scoring. We developed a gene-scoring procedure based on high spatial resolution gene-expression data, that assigns a score for each gene based on its spatial gene co-expression patterns. The AHBA (47) provides high resolution whole-brain gene expression data derived from six post-mortem donor brains. First, genes implicated in GWAS were selected using the same procedure as described for PPI-based scoring. Here we included 15 744 genes that passed our quality-control criteria (65) across 180 regions of the left cortical hemisphere. First, we calculated a gene–gene coexpression matrix between each pair of 15 744 genes that captures correlations in regional expression profiles. High scores in the coexpression matrix indicated that a pair of genes had similar spatial expression patterns across the cortex. Gene scores, $s_i^{\mathrm{GWAS-AHBA}}$, for each of 2155 genes, g_i, were then calculated by comparing each gene's coexpression with genes implicated in GWAS *vs* all other genes using z-scores from a Wilcoxon rank-sum test. Therefore, high scores indicate that a gene has stronger coexpression with GWAS-implicated genes than other genes.

C.5. Hi-c-based GWAS scoring. H-MAGMA assigns non-coding SNPs to their cognate genes based on long-range chromatin interactions in human brain tissue across two developmental epochs and two brain cell types measured by Hi-C (66). This technique identifies developmentally specific and cell-specific neurobiologically relevant genes. Using H-MAGMA software (66) (`https://github.com/thewonlab/H-MAGMA`) we constructed four chromatin interaction gene scores $s^{\mathrm{GWAS-Hi-C}}$ based on fetal brain, adult brain, neuronal, and astrocytic brain Hi-C. Similarly to eQTL and position-based gene scoring, a score vector $s^{\mathrm{GWAS-Hi-C}}$ was constructed using $-\log_{10}$ transformed gene p-values for each of 2155 genes, as $s_i^{\mathrm{GWAS-Hi-C}} = -\log_{10}(P_i)$.

D. Gene-score similarity. We aimed to evaluate the extent to which genes involved in pharmacological treatments for psychiatric disorders match genes implicated in their corresponding GWAS. To achieve this, we independently constructed two sets of scores quantifying the involvement of each gene in: (i) pharmacological treatments s_i^{drug}; and (ii) GWAS s_i^{GWAS} based on different bioinformatic datasets (e.g. $s^{\mathrm{GWAS-position}}$, $s^{\mathrm{GWAS-PPI}}$, $s^{\mathrm{GWAS-eQTL}}$, $s^{\mathrm{GWAS-AHBA}}$). The similarity between two different score vectors was defined using a similarity function $\rho(s^{\mathrm{drug}}, s^{\mathrm{GWAS}})$. First, each score vector was normalized as $\hat{s}_i^{(x)} = s_i^{(x)} / \sum_i s_i^{(x)}$, such that normalized score vectors define positive mass across genes that sums to unity; hence both drug and GWAS-based scores have equal contribution. We then defined the weighted similarity score, ρ, as a simple inner product between two unit vectors $\rho = \hat{\mathbf{s}}^{\mathrm{drug}} \cdot \hat{\mathbf{s}}^{\mathrm{GWAS}}$:

$$\rho = \sum_i \hat{s}_i^{\mathrm{drug}} \hat{s}_i^{\mathrm{GWAS}} . \quad (1)$$

Gene-score vectors in gene space that point in the same direction receive a maximal score, $\rho = 1$; those that are orthogonal receive a minimal score, $\rho = 0$. For example, if pharmacological treatments for a disorder target the same genes that are implicated in GWAS, then the resulting normalized score vectors, $\hat{\mathbf{s}}^{drug}$ and $\hat{\mathbf{s}}^{GWAS}$, will place similar weight on similar genes and thus obtain a high weighted similarity score, ρ.
The contribution of each gene to the similarity score ρ was quantified by comparing the inner product of s_i^{drug} and s_i^{GWAS} in real data to the distribution of inner product values derived using a null model [see Evaluating significance based on random treatments]. The gene ranking was based on the significance of p-values derived from such permutation testing.

E. Evaluating significance based on random treatments. We evaluated the statistical significance of our results by estimating a p-value using a permutation test comparing empirically derived weighted similarity scores, ρ_{emp}, to an ensemble of 5000 null similarity scores generated by selecting a set of random drugs, ρ_{rand}. Specifically, for each disorder we repeatedly selected N random drugs (where N is the number of drugs in the

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

curated list of pharmacological treatments for that disorder) from the total of 1900 drugs in the *DrugBank* database and generated an ensemble of gene-score vectors, $\hat{\mathbf{s}}^{drug}_{rand}$. Random gene-score vectors where then normalized and compared to the real $\hat{\mathbf{s}}^{GWAS}$ resulting in a distribution of ρ_{rand} values used to derive p-values.
Under the assumption that treatments for psychiatric disorders are likely to target different sets of genes compared to drugs for non-psychiatric conditions, we also tested the significance using a constrained set of treatments relevant only to psychiatric disorders. In this case, the random treatments were selected from a set of 82 drugs used to treat four psychiatric conditions investigated in our analysis, namely ADHD, schizophrenia, major depression, and bipolar disorder. Each disorder had a different number of available treatments, therefore, to avoid the over-sampling of drugs for disorders that have more treatments, we selected drugs from a probability distribution that gave an equal chance for all disorder drugs to be selected. Specifically, for each drug the probability of being selected was inversely proportional to the total number of drugs for that disorder ($1/N$). If a drug is used to treat more than one disorder, these scores were summed. Similarly to the general case, we then generated an assembly of 5000 gene score vectors based on random drug selection for each disorder and used them to compute the distributions of $\rho_{\text{randpsych}}$ values to derive p-values.

Gene-set enrichment analysis using over-representation analysis. Over-representation analysis (ORA) examines if there are gene sets [e.g., annotated using Gene Ontology (GO)] within a selected list of genes that are statistically over-represented in that list. Here we used ORA to investigate which GO categories are enriched in treatment-based scores (s^{drug}) and PPI-based GWAS scores (s^{gwas}) for psychiatric disorders. Considering that only the minority of all 2155 genes were assigned a non-zero scores in both s^{drug} and s^{gwas}, we applied a threshold to retain all genes with non-zero scores while using the remaining genes as a reference list. Functional gene group analyses were performed using version 3.2 of ErmineJ software (67). Gene ontology annotations were obtained from GEMMA (68) (https://gemma.msl.ubc.ca/arrays/showArrayDesign.html?id=735) as `Generic_human_ncbiIds_noParents.an.txt.gz` on April 29, 2021. Gene Ontology terms and definitions were automatically downloaded by ErmineJ on April 29, 2021 as `go.obo` (data version 2021-02-01) and can be downloaded from http://release.geneontology.org/2021-02-01/ontology/index.html. We performed ORA on the thresholded treatment-based scores (s^{drug}) and PPI-based GWAS scores (s^{drug}) for each psychiatric disorder testing the biological process GO categories with 5 to 100 genes available. The resulting p-values were corrected across 3807 and 3753 GO categories for s^{drug} and s^{gwas} respectively, controlling the false discovery rate (FDR) at 0.05 using the method of Benjamini and Hochberg (69).

ACKNOWLEDGEMENTS
We are thankful to iPSYCH (Ditte Demontis and Anders Børglum) and deCODE (Bragi Walters and Hreinn Stefansson) for giving us access to the recent GWAS meta-analysis summary statistics of ADHD (8). MAB was supported by a National Health and Medical Research Council of Australia Senior Research Fellowship. AF was supported by the National Health and Medical Research Council of Australia (ID:1050504) and Sylvia and Charles Viertel Charitable Foundation. BDF was supported by National Health and Medical Research Council Grant 1089718.

1. A. Buniello, J. MacArthur, o. M. Cerez, et al. The nhgri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* **47**, D1005 (2019).
2. M. C. Mills and C. Rahal. A scientometric review of genome-wide association studies. *Communications Biology* **2**, 1 (2019).
3. P. M. Visscher, N. R. Wray, Q. Zhang, et al. 10 years of GWAS discovery: Biology, function, and translation. *American Journal of Human Genetics* **101**, 5 (2017).
4. D. J. M. Crouch and W. F. Bodmer. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proceedings of the National Academy of Sciences* **117**, 18924 (2020).
5. S. Ripke, B. M. Neale, A. Corvin, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421 (2014).
6. The Schizophrenia Working Group of the Psychiatric Genomics Consortium, S. Ripke, J. T. Walters, and M. C. O'Donovan. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv* p. 2020.09.12.20192922 (2020).
7. D. Demontis, R. K. Walters, J. Martin, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63 (2018).
8. D. Demontis, G. B. Walters, G. Athanasiadis, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *medRxiv* p. 2022.02.14.22270780 (2022).
9. N. R. Wray, S. Ripke, M. Mattheisen, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668 (2018).
10. D. M. Howard, M. J. Adams, T.-K. Clarke, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343 (2019).
11. D. M. Ruderfer, S. Ripke, A. McQuillin, et al. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705 (2018).
12. J. Grove, S. Ripke, T. D. Als, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* **51**, 431 (2019).
13. W. S. Bush and J. H. Moore. Chapter 11: Genome-wide association studies. *PLoS Comput. Biol.* **8**, e1002822 (2012).
14. G. Breen, Q. Li, B. L. Roth, et al. Translating genome-wide association findings into new therapeutics for psychiatry. *Nature Neuroscience* **19**, 1392 (2016).
15. Y. Okada, D. Wu, G. Trynka, and et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
16. C. M. Schooling, J. V. Huang, J. V. Zhao, et al. Disconnect between genes associated with ischemic heart disease and targets of ischemic heart disease treatments. *EBioMedicine* **28**, 311 (2018).
17. C. Finan, A. Gaulton, F. A. Kruger, et al. The druggable genome and support for target identification and validation in drug development. *Science Translational Medicine* **9**, eaag1166 (2017).
18. M. J. Owen, M. C. O'Donovan, A. Thapar, and N. Craddock. Neurodevelopmental hypothesis of schizophrenia. *The British Journal of Psychiatry* **198**, 173 (2011).
19. S. E. Hyman. Psychiatric drug development: Diagnosing a crisis. *Cerebrum: the Dana Forum on Brain Science* **2013** (2013).
20. T. E. Wilens. Mechanism of action of agents used in attention-deficit/hyperactivity disorder. *The Journal of Clinical Psychiatry* **67 Suppl 8**, 32 (2006).
21. C. Taylor, A. D. Fricker, L. A. Devi, and I. Gomes. Mechanisms of action of antidepressants: from neurotransmitter systems to signaling pathways. *Cellular signalling* **17**, 549 (2005).
22. S. Miyamoto, G. E. Duncan, C. E. Marx, and J. A. Lieberman. Treatments for schizophrenia: a critical review of pharmacology and mechanisms of action of antipsychotic drugs. *Molecular Psychiatry* **10**, 79 (2005).
23. A. Cipriani, T. A. Furukawa, G. Salanti, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet* **391** (2018).
24. S. Cortese, N. Adamo, C. D. Giovane, et al. Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. *The Lancet Psychiatry* **5**, 727 (2018).
25. M. Huhn, A. Nikolakopoulou, J. Schneider-Thoma, et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. *The Lancet* **394**, 939 (2019).
26. A. N. de Araújo, E. P. de Sena, I. R. de Oliveira, and M. F. Juruena. Antipsychotic agents: efficacy and safety in schizophrenia. *Drug, Healthcare and Patient Safety* **4**, 173 (2012).
27. T. P. Blackburn. Depressive disorders: Treatment failures and poor prognosis over the last 50 years. *Pharmacology Research & Perspectives* **7**, e00472 (2019).
28. M. R. Nelson, H. Tipney, J. L. Painter, et al. The support of human genetic evidence for approved drug indications. *Nature Genetics* **47** (2015).
29. E. A. King, J. W. Davis, and J. F. Degner. Are drug targets with genetic support twice as likely to be approved? revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS genetics* **15**, e1008489 (2019).
30. P. Sanseau, P. Agarwal, M. R. Barnes, et al. Use of genome-wide association studies for drug repositioning. *Nature Biotechnology* **30**, 317 (2012).
31. H. A. Gaspar and G. Breen. Drug enrichment and discovery from schizophrenia genome-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.11.22274981; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

wide association results: an analysis and visualisation approach. *Scientific Reports* **7**, 12460 (2017).

32. D. M. Ruderfer, A. W. Charney, B. Readhead, et al. Polygenic overlap between schizophrenia risk and antipsychotic response: a genomic medicine approach. *The Lancet Psychiatry* **3**, 350 (2016).
33. H.-C. So, C. K.-L. Chau, A. Lau, S.-Y. Wong, and K. Zhao. Translating GWAS findings into therapies for depression and anxiety disorders: gene-set analyses reveal enrichment of psychiatric drug classes and implications for drug repositioning. *Psychological Medicine* **49** (2019).
34. C. Cao and J. Moult. GWAS and drug targets. *BMC Genomics* **15**, S5 (2014).
35. M. J. Gandal, V. Leppa, H. Won, N. N. Parikshak, and D. H. Geschwind. The road to precision psychiatry: translating genetics into disease mechanisms. *Nature Neuroscience* **19**, 1397 (2016).
36. T. Lencz and A. K. Malhotra. Targeting the schizophrenia genome: a fast track strategy from GWAS to clinic. *Molecular Psychiatry* **20**, 820 (2015).
37. D. Jhamb, M. Magid-Slav, M. R. Hurle, and P. Agarwal. Pathway analysis of GWAS loci identifies novel drug targets and repurposing opportunities. *Drug Discovery Today* **24**, 1232 (2019).
38. H. Fang, H. D. Wolf, B. Knezevic, et al. A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nature Genetics* **51**, 1082 (2019).
39. N. Safari-Alighiarloo, M. Taghizadeh, M. Rezaei-Tavirani, B. Goliaei, and A. A. Peyvandi. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterology and Hepatology From Bed to Bench* **7**, 17 (2014).
40. M. Imamura, A. Takahashi, T. Yamauchi, et al. Genome-wide association studies in the japanese population identify seven novel loci for type 2 diabetes. *Nature Communications* **7**, 10531 (2016).
41. Y. Okada, D. Wu, G. Trynka, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376 (2014).
42. H.-C. So, C. K.-L. Chau, W.-T. Chiu, et al. Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nature Neuroscience* **20**, 1342 (2017).
43. A. Lau and H.-C. So. Turning genome-wide association study findings into opportunities for drug repositioning. *Computational and Structural Biotechnology Journal* **18**, 1639 (2020).
44. D. A. Baird, J. Z. Liu, J. Zheng, et al. Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome. *PLOS Genetics* **17**, e1009224 (2021).
45. S. E. Kahn, M. E. Cooper, and S. Del Prato. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present and future. *Lancet* **383**, 1068 (2014).
46. D. S. Wishart, C. Knox, A. C. Guo, et al. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research* **34**, D668 (2006).
47. M. J. Hawrylycz, E. S. Lein, A. L. Guillozet-Bongaarts, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391 (2012).
48. Y. G. Tak and P. J. Farnham. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics & Chromatin* **8**, 57 (2015).
49. C. A. d. Leeuw, J. M. Mooij, T. Heskes, and D. Posthuma. MAGMA: Generalized gene-set analysis of GWAS data. *PLOS Computational Biology* **11**, e1004219 (2015).
50. R. C. Ramaker, K. M. Bowling, B. N. Lasseigne, et al. Post-mortem molecular profiling of three psychiatric disorders. *Genome Medicine* **9**, 72 (2017).
51. M. Korostynski, M. Piechota, J. Dzbek, et al. Novel drug-regulated transcriptional networks in brain reveal pharmacological properties of psychotropic drugs. *BMC Genomics* **14**, 606 (2013).
52. Psychiatric Genomics Consortium, S. Lee, and S. e. a. Ripke. Genetic relationship between five psychiatric disorders estimated from genome-wide snps. *Nat Gene* **45**, 984–994 (2013).
53. J. Dekker and T. Misteli. Long-range chromatin interactions. *Cold Spring Harbor Perspectives in Biology* **7**, a019356 (2015).
54. N. Mullins, A. J. Forstner, K. S. O'Connell, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics* **53**, 817 (2021).
55. C. O'Dushlaine, L. Rossin, P. H. Lee, et al. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature Neuroscience* **18**, 199 (2015).
56. S. de Jong, L. R. Vidler, Y. Mokrab, D. A. Collier, and G. Breen. Gene-set analysis based on the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia. *Journal of Psychopharmacology* **30**, 826 (2016).
57. A. Xue, Y. Wu, Z. Zhu, and et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9** (2018).
58. S. Shah, A. Henry, C. Roselli, and et al. Genome-wide association and mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11** (2020).
59. K. M. de Lange, L. Moutsianas, J. C. Lee, and et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
60. C. A. Leeuw, J. M. Mooij, T. Heskes, and D. Posthuma. Magma: Generalized gene-set analysis of gwas data. *PLoS Comput Biol* **11**, e1004219 (2015).
61. M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059 (2010).
62. Z. F. Gerring, E. R. Gamazon, E. M. Derks, and for the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. A gene co-expression network-based analysis of multiple brain tissues reveals novel genes and molecular pathways underlying major depression. *PLoS Genetics* **15**, 1 (2019).
63. D. Wang, S. Liu, J. Warrell, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362** (2018).
64. G. Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204 (2017).
65. A. Arnatkevičiūtė, B. D. Fulcher, and A. Fornito. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* (2019).
66. N. Y. A. Sey, B. Hu, W. Mah, et al. A computational tool (h-magma) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci* **23**, 583–593 (2020).
67. J. Gillis, M. Mistry, and P. Pavlidis. Gene function analysis in complex data sets using ErmineJ. *Nat. Protoc.* **5**, 1148 (2010).
68. A. Zoubarev, K. M. Hamer, K. D. Keshav, et al. Gemma: a resource for the reuse, sharing and meta-analysis of expression profiling data. *Bioinformatics* **28**, 2272 (2012).
69. Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289 (1995).