Using genome-wide association results to identify drug repurposing candidates

Nathaniel Bell[1], Emil Uffelmann[1], Eva van Walree[1,3], Christiaan de Leeuw[1], & Danielle Posthuma[1,2]

[1] Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2] Department of Child and Adolescent Psychiatry and Pediatric Psychology, Section Complex Trait Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam, The Netherlands.

[3] Department of Clinical Genetics, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Drug repurposing may provide a solution to the substantial challenges facing *de novo* drug development. Given that 66% of FDA-approved drugs in 2021 were supported by human genetic evidence, drug repurposing methods based on genome wide association studies (GWAS), such as drug gene-set analysis, may prove an efficient way to identify new treatments. However, to our knowledge, drug gene-set analysis has not been tested in non-psychiatric phenotypes, and previous implementations may have contained statistical biases when testing groups of drugs. Here, 1201 drugs were tested for association with hypercholesterolemia, type 2 diabetes, coronary artery disease, asthma, schizophrenia, bipolar disorder, Alzheimer's disease, and Parkinson's disease. We show that drug gene-set analysis can identify clinically relevant drugs (e.g., simvastatin for hypercholesterolemia [p = 2.82E-06]; mitiglinide for type 2 diabetes [p = 2.66E-07]) and drug groups (e.g., C10A for coronary artery disease [p = 2.31E-05]; insulin secretagogues for type 2 diabetes [p = 1.09E-11]) for non-psychiatric phenotypes. Additionally, we demonstrate that when the overlap of genes between drug-gene sets is considered we find no groups containing approved drugs for the psychiatric phenotypes tested. However, several drug groups were identified for psychiatric phenotypes that may contain possible repurposing candidates, such as ATC codes J02A (p = 2.99E-09) and N07B (p = 0.0001) for schizophrenia. Our results demonstrate that clinically relevant drugs and groups of drugs can be identified using drug gene-set analysis for a number of phenotypes. These findings have implications for quickly identifying novel treatments based on the genetic mechanisms underlying diseases.

Keywords: genome-wide association study, drug repurposing, gene-set analysis.

Over the last 15 years, the use and effectiveness of genome-wide association studies (GWAS) have developed rapidly, resulting in the discovery of over 270,000 phenotypically associated genetic variants[1,2]. While translating GWAS findings into use for pharmacological and clinical purposes has been challenging, seminal work by Nelson et al. and King et al. demonstrated that drugs with genetic support for efficacy were 2-5 times more likely to gain approval status compared to compounds with no such evidence[3,4]. In fact, 66% of FDA-approved drugs in 2021 are supported by human genetic evidence that link the compounds to their gene product targets[5]. GWAS results have been used as early as 2009 to prioritize novel targets for pharmaceuticals[6]. For example, Wang et al identified the IL12/IL23 pathway as a novel target for Crohn's disease, which supported the subsequent development of Ustekinumab, an antibody that binds with IL12 and IL23[6,7].

However, the development of novel pharmaceuticals is a time and financially intensive process, averaging 12.8 years and $2.6 billion dollars for a single new medication[8,9]. Despite wide advances across science and technology in the past 60 years, the cost for approved drugs per billion spent has seen an 80-fold increase after adjusting for inflation[10]. Additionally, the development of novel drugs for psychiatric and neurological conditions has been particularly affected, with several large pharmaceutical companies shutting down research and development for psychiatric disorders entirely[11]. Given the size of the challenges facing *de novo* drug development, interest has turned towards drug repurposing, the process of identifying new clinical indications (the phenotype a drug is approved to treat) for currently approved medications[12]. Moreover, the vast quantities of GWAS results provide excellent data regarding possible gene targets to be leveraged for drug repurposing. To date, many methods have been created to repurpose drugs using GWAS summary statistics, including candidate gene approaches, Mendelian randomization, pharmagenic enrichment scores, transcriptomic signature matching, and drug gene-set analysis[13–17].

69 Drug gene-set analysis, an approach first published by De Jong and Breen[18], can be used to
70 identify candidate drugs based on GWAS results. Gene-set, or pathway, analysis involves mapping
71 genetic variants (e.g., single nucleotide polymorphisms [SNPs], copy number variants [CNVs]) to
72 genes, and grouping genes into sets based on, for example, *a priori* biological functions or co-
73 expression networks, and testing each gene set for association with the phenotype[19,20]. For drug
74 gene-set analysis, gene sets are created for drugs based on known biological and/or chemical
75 properties. For example, a drug-gene set can be made of the genes that code the proteins targeted
76 by a specific drug, and this gene set can be tested for association with a phenotype. While De Jong
77 & Breen found no gene-sets to be significantly associated with schizophrenia, their nominally
78 significant results included the known schizophrenia drug trifluoperazine, and metoclopramide
79 and neratinib, which target dopamine receptors and tyrosine kinases (both implicated in
80 schizophrenia)[18]. This method has been continued by Gaspar and colleagues, who developed the
81 drug repurposing tool Drug Targetor, which is largely based on drug gene-set analysis[17,21].
82 However, while drug gene-set analysis has been tested in schizophrenia and bipolar
83 disorder, it has never been verified in non-psychiatric phenotypes and, to our knowledge, no drugs
84 have ever been found to be significantly associated with schizophrenia. Additionally, previous
85 research testing groups of drugs for associations have used methods that violated statistical
86 assumptions of independence[17]. Here, we apply drug gene-set analysis to eight phenotypes –
87 hypercholesterolemia (i.e., high cholesterol), type 2 diabetes, coronary artery disease, asthma,
88 schizophrenia, bipolar disorder, Alzheimer's disease, and Parkinson's disease – and use a novel
89 statistical approach to test if groups of drugs are enriched for genetic signal associated with a
90 phenotype while accounting for the covariance between drugs. Our objectives were first, to assess
91 the ability of drug gene-set analysis at prioritizing clinically relevant drugs for each phenotype,
92 and second, to identify candidate drugs for repurposing for the phenotypes investigated.

Results

We extracted data for drug-gene targets and interactions from the Clue Repurposing Hub and the Drug Gene Interaction Database and created a gene set for each drug (n = 1201) composed of the genes whose protein products are known to be targeted by or interact with the drug in question (Figure 1A)[22,23]. To test drug-gene set associations, competitive gene-set analysis was performed using the MAGMA software tool for each of the eight phenotypes[19] while conditioning on a gene set of all drug target genes in our data (n = 2281; Methods). This was done to ensure that significant drug-gene set associations were due to effects unique to drug pathways, and not due to common properties shared by all drug target genes (e.g., many gene products targeted by drugs are receptors). For the individual drug-gene sets see Supplementary Table S1.

Individual drug-gene set results

We identified 22 significant drug-trait associations across the eight phenotypes we investigated using drug gene-set analysis (Table 1 & 2; Fig 1A). Drugs were significantly associated with each of the phenotypes except for asthma and Parkinson's disease. For the non-psychiatric/neurological phenotypes, many of our results included drugs approved to treat the diseases in question. For instance, all 4 of the drugs associated with type 2 diabetes – glipizide, glyburide, mitiglinide, and repaglinide – are approved for type 2 diabetes[24–26] (Tables 1 & 2). Similarly, statin drugs (e.g., pravastatin and atorvastatin) were associated with both coronary artery disease and hypercholesterolemia, which are used to lower LDL cholesterol and prevent myocardial infarctions[27]. Likewise, lomitapide and fenofibrate, which were also associated with coronary artery disease, are drugs used to treat hypercholesterolemia, a known risk factor for coronary artery disease[28–30]. We found no significant associations between approved psychiatric drugs and the psychiatric phenotypes that were tested. However, the antiepileptic drug gabapentin was

117 significantly associated with schizophrenia, which is known to have some efficacy in treating
118 psychiatric disorders[31]. For all individual drug gene-set analysis results see Supplementary Tables
119 S2 – S9.
120
121 **Table 1. Summary of results**

Phenotype	Total tested	Total Approved	% Approved	N Signif. drugs	N Signif. and approved	% Signif. of approved drugs
Hypercholesterolemia	1162	8	0.68%	5	3	37.5%
Type 2 diabetes	1056	30	2.84%	4	4	13.33%
Coronary artery disease	1172	14	1.19%	6	1	7.14%
Asthma	1172	29	2.47%	0	0	0%
Schizophrenia	1172	45	3.84%	4	0	0%
Bipolar disorder	1172	13	1.11%	2	0	0%
Alzheimer's disease	1172	5	0.42%	1	0	0%
Parkinson's disease	1172	29	2.47%	0	0	0%

122 Total tested = total number of drug-gene sets tested for each phenotype; Total approved = number of
123 approved drug-gene sets tested for each phenotype; % approved = % of total drugs tested that are
124 approved for each phenotype; N signif. drugs = total number of significant drugs for each phenotype (see
125 Table 2 for all significant drugs). Bonferroni-corrected p-value thresholds for hypercholesterolemia = 0.05 /
126 1162 = 4.3E-05, for type 2 diabetes = 0.05 / 1056 = 4.73E-05, and for coronary artery disease, asthma,
127 schizophrenia, bipolar disorder, Alzheimer's disease, and Parkinson's disease = 0.05 / 1172 = 4.27E-05; N
128 signif. and approved = number of significant drugs that are approved for the given phenotype; % signif. of
129 approved drugs = % of the total number of approved drugs for a tested phenotype that were significant.
130

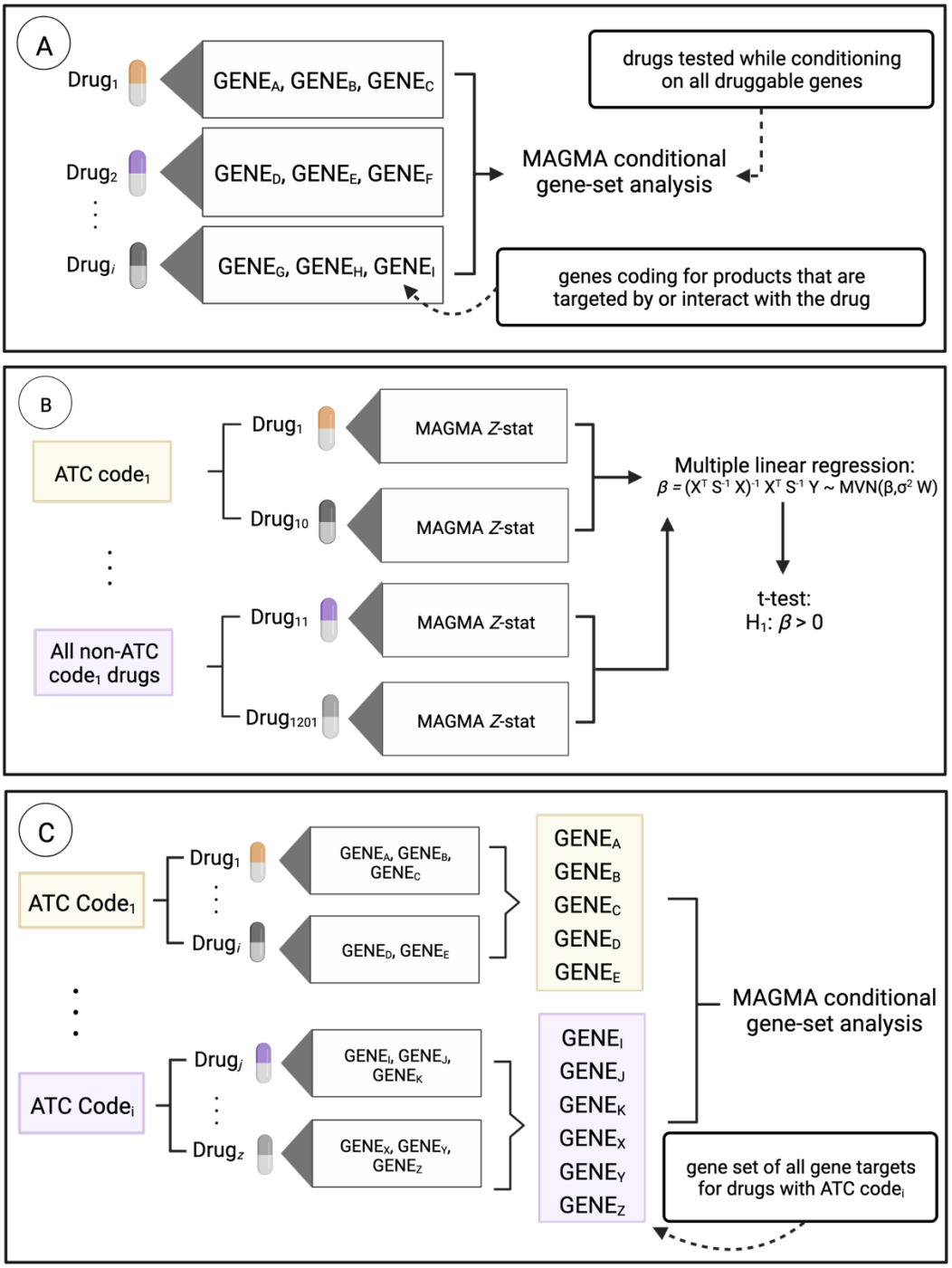

Figure 1 Overview of drug gene-set analysis. **(A)** Gene sets were created for every drug (n = 1201) consisting of the genes whose protein products are targeted by or interact with that drug. Since similar drugs share gene targets it is possible for overlap of genes to occur between drugs. MAGMA was used for gene-set analysis to test for significant relationships between drug-gene sets and each phenotype. Drug-gene sets were tested using a conditional competitive gene-set analysis (Methods). **(B)** Drug groups were tested for enrichment of genetic signal using a modified multiple linear regression model. This was done for ATC III groups (n = 85), mechanism of action drug groups (n = 79), and clinical indication drug groups (n = 118) with $n > 5$ (Methods). **(C)** Lastly, gene sets were created for each subcategory of each grouping method (i.e., ATC III codes, clinical indications, and mechanisms of action) and tested for association using the same MAGMA protocol as drugs in panel A (Methods).

142 **Table 2. Individual drug-gene set results for all phenotypes**

Drug gene set	Phenotype	Approved	N genes	MAGMA *p*
atorvastatin	TCL	No	60	1.77E-11
lomitapide	TCL	Yes	4	7.33E-08
pravastatin	TCL	Yes	34	6.15E-11
saroglitazar	TCL	No	2	3.22E-05
simvastatin	TCL	Yes	52	2.82E-06
mitiglinide	T2D	Yes	4	2.66E-07
glyburide	T2D	Yes	19	2.29E-05
repaglinide	T2D	Yes	15	9.01E-16
glipizide	T2D	Yes	6	1.08E-05
aminocaproic acid	CAD	No	3	1.59E-11
atorvastatin	CAD	No	61	8.40E-08
fenofibrate	CAD	No	28	7.63E-06
lomitapide	CAD	No	4	3.96E-11
lovastatin	CAD	Yes	43	3.38E-06
pravastatin	CAD	No	35	1.59E-07
amlodipine	SCZ	No	12	1.51E-05
benidipine	SCZ	No	2	1.32E-05
felodipine	SCZ	No	13	1.53E-05
gabapentin	SCZ	No	38	3.80E-08
manidipine	BIP	No	2	2.64E-06
nilvadipine	BIP	No	2	2.64E-06
lubiprostone	ALZ	No	2	3.76E-06

143 Drug-gene set = drugs whose gene set was significant after Bonferroni correction for the number of drug
144 gene sets tested (see Table 1 for number of drugs tested for each phenotype); Phenotype = the phenotype
145 the drug was associated with. TCL = hypercholesterolemia, T2D = type 2 diabetes, CAD = coronary artery
146 disease, BIP = bipolar disorder, SCZ = schizophrenia, ALZ = Alzheimer's disease; Approved = whether the
147 drug is approved to be used for the phenotype it was associated with; N genes = number of genes in the
148 drug gene set; MAGMA *p* = p-value for individual drug gene-sets from a competitive gene-set analysis while
149 conditioning on a gene set of all drug target genes in the data; No drugs were significantly associated with
150 Asthma or Parkinson's disease.
151

152 ***Drug group enrichment***

153 Drug gene sets were grouped in three ways: by Anatomical Therapeutic Classification (ATC) III
154 code, by clinical indication, and by the mechanism of action. Previous research that tested for
155 enrichment of genetic signal within groups of drug gene sets used drug enrichment curves derived
156 from a Wilcoxon Mann Whitney U test[17]. However, due to the overlap of genes between drug gene
157 sets, the assumption of data independence needed for a Wilcoxon Mann Whitney U test is not
158 valid. Thus, here we instead employed a multiple linear regression approach, modeling the drug
159 group membership as a predictor of the MAGMA drug gene-set t-statistic, to test whether drugs in
160 that drug group exhibited (on average) more strongly associated drug gene sets than other drugs.
161 Overlap between gene sets is accounted for through the residual covariance matrix of the
162 regression model.

163 Each drug group with 5 or more drugs within the ATC (n = 85), mechanism of action (n =
164 79), and clinical indication (n = 118) categories was tested for enrichment of genetic signal using
165 the method above. Bonferroni correction was applied by correcting for the number of drug groups
166 tested within each of the three grouping methods, and then using a family-wise correction for the
167 number of grouping methods used. Therefore, the Bonferroni-corrected p-value threshold was
168 (0.05/85) / 3 = 0.00019 for ATC drug groups, (0.05/79) / 3 = 0.00021 for mechanism of action
169 drug groups, and (0.05/118) / 3 = 0.00014 for clinical indication drug groups. For all drug groups
170 see Supplementary Tables S10 – S12, and for the drug group enrichment results see Supplemental
171 Table S13 – S21.

172 As seen in Figure 2 (top), we identified 14 drug groups that were significantly enriched for
173 genetic signal across the 8 phenotypes we examined. For the non-psychiatric/neurological
174 phenotypes, we found at least one drug group containing one or more drugs approved for the
175 phenotype tested. For hypercholesterolemia, we identified two drug groups where more than 50%

176 of the drugs were approved for hypercholesterolemia (C10A and HMGCR inhibitors), and we
177 identified the mechanism of action group insulin secretagogues for type 2 diabetes, which contains
178 only drugs approved for type 2 diabetes. None of the drug groups that were associated with our
179 psychiatric and neurological phenotypes contained approved drugs. However, the ATC code N07B
180 – drugs used for addictive disorders – that is associated with schizophrenia contains several drugs
181 (e.g., naltrexone, buprenorphine) that have shown significant improvements in treating both
182 positive and negative symptoms in the disorder[32]. As seen in the bottom panels of Figure 2, none
183 of the clinical indication drug groups for the phenotypes we examined were significantly associated
184 with our phenotype after correction for multiple tests. The clinical indication groups for Parkinson's
185 disease, type 2 diabetes, and hypercholesterolemia were nominally enriched for genetic signal for
186 each of their respective phenotypes.
187

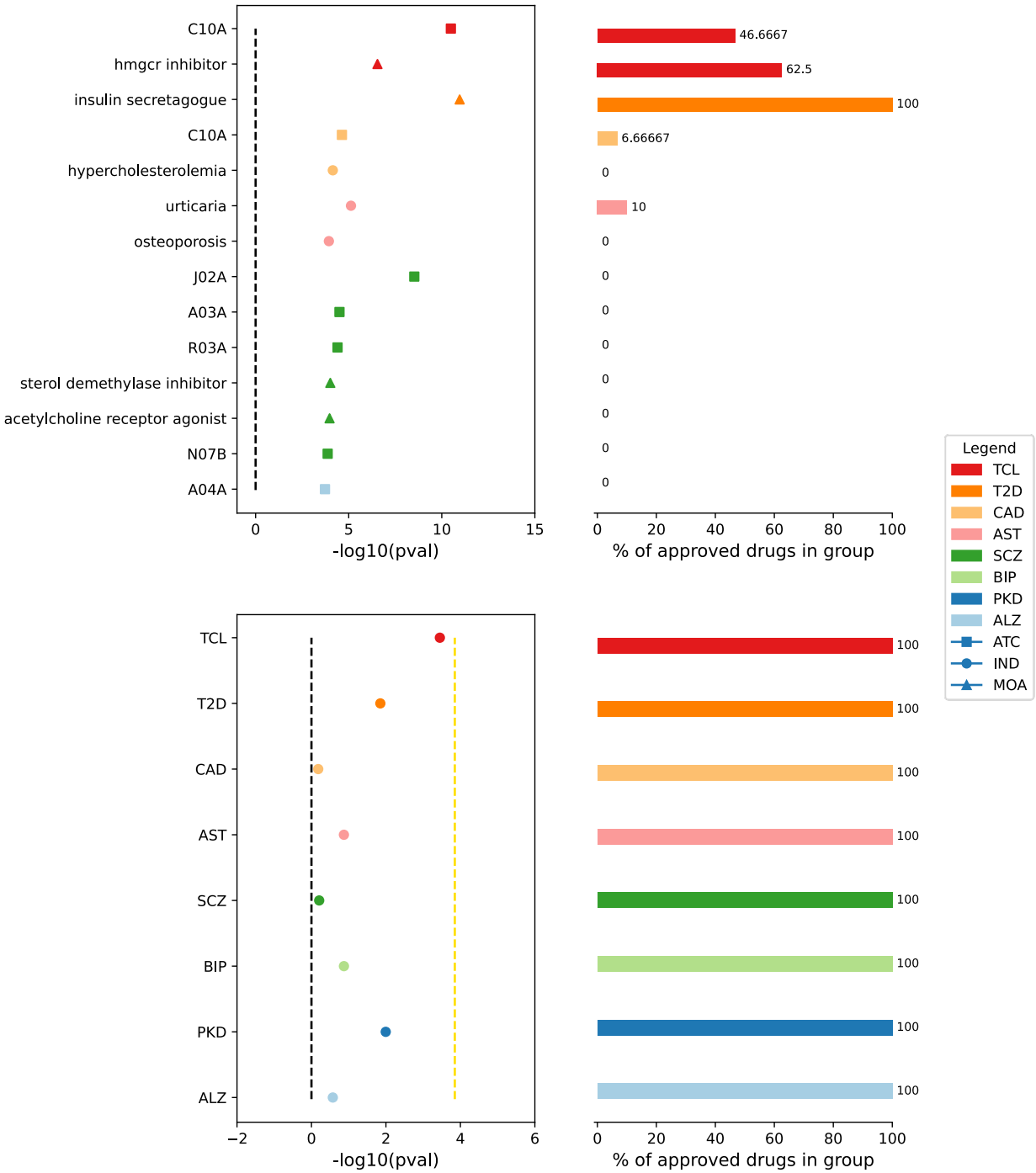

Figure 2 Drug group enrichment. Drug groups were tested for enrichment of genetic signal with each phenotype. The left panels show the -$\log_{10}$ p-values for each drug group, and the right panels show the percentage of approved drugs in the group. The top panels show all drug groups that are significant after multiple testing correction. The bottom panels portray the drug groups containing all approved drugs for each of the 8 phenotypes that were tested. TCL = hypercholesterolemia; T2D = type 2 diabetes; CAD = coronary artery disease; AST = asthma, SCZ = schizophrenia; BIP = bipolar disorder; PKD = Parkinson's disease; ALZ = Alzheimer's disease.

To examine the genes driving the significant results seen in the top panel of Figure 2, we generated heatmaps of the drug-gene sets for the most significant drug groups for each phenotype; insulin secretagogues for type 2 diabetes (Fig 3A), the clinical indication drug group of urticaria for asthma (Fig 3B), ATC code J02A for schizophrenia (Fig 3C), and ATC code A04A for Alzheimer's disease (Fig 3D). For the C10A heatmaps associated with hypercholesterolemia and coronary artery disease see Supplemental Figures S1 and S2.

As seen in Figure 3A, the genes with the highest MAGMA -$\log_{10}$ p-values targeted by insulin secretagogues are *KCNQ1*, *TCF7L2*, *PPARγ*, *IGF2BP2*, and *KCNJ11*. All these genes are known to be associated with type 2 diabetes and affect biological processes like insulin secretion and fat storage in adipose tissue[33,34]. For the clinical indication drug group urticaria (Fig 3B), there are several genes with evidence linking them to asthma. Namely, research has indicated that the *PPARδ* gene targeted by desloratadine may be a promising therapeutic target for asthma, with desloratadine itself – an allergy medication – having shown efficacy in treating seasonal allergic asthma[35,36]. Moreover, another study found that *EHMT2* was differentially expressed in patients with asthma compared to healthy controls[37]. The targets of drugs in the ATC J02A group also contain genes related to schizophrenia. Antigen levels for *DPYD* were found to be elevated in individuals with schizophrenia compared with healthy controls, and it is known to be associated with schizophrenia via previous genome-wide association studies[38,39]. Moreover, there is evidence indicating that splicing variation of *CYP2D6* may play a role in schizophrenia as well[40]. Lastly, the ATC group A04A associated with Alzheimer's disease targets a number of serotonergic receptors, which have shown promise as therapeutic targets in animal models by reducing levels of amyloid-beta[41]. Interestingly, all of the drug groups depicted appear to be driven by a variety of genes, as opposed to a small number of genes targeted by all drugs in a group (with the exceptions of *PPARγ*, *KCNJ11*, and *ABCC8* for the insulin secretagogues).

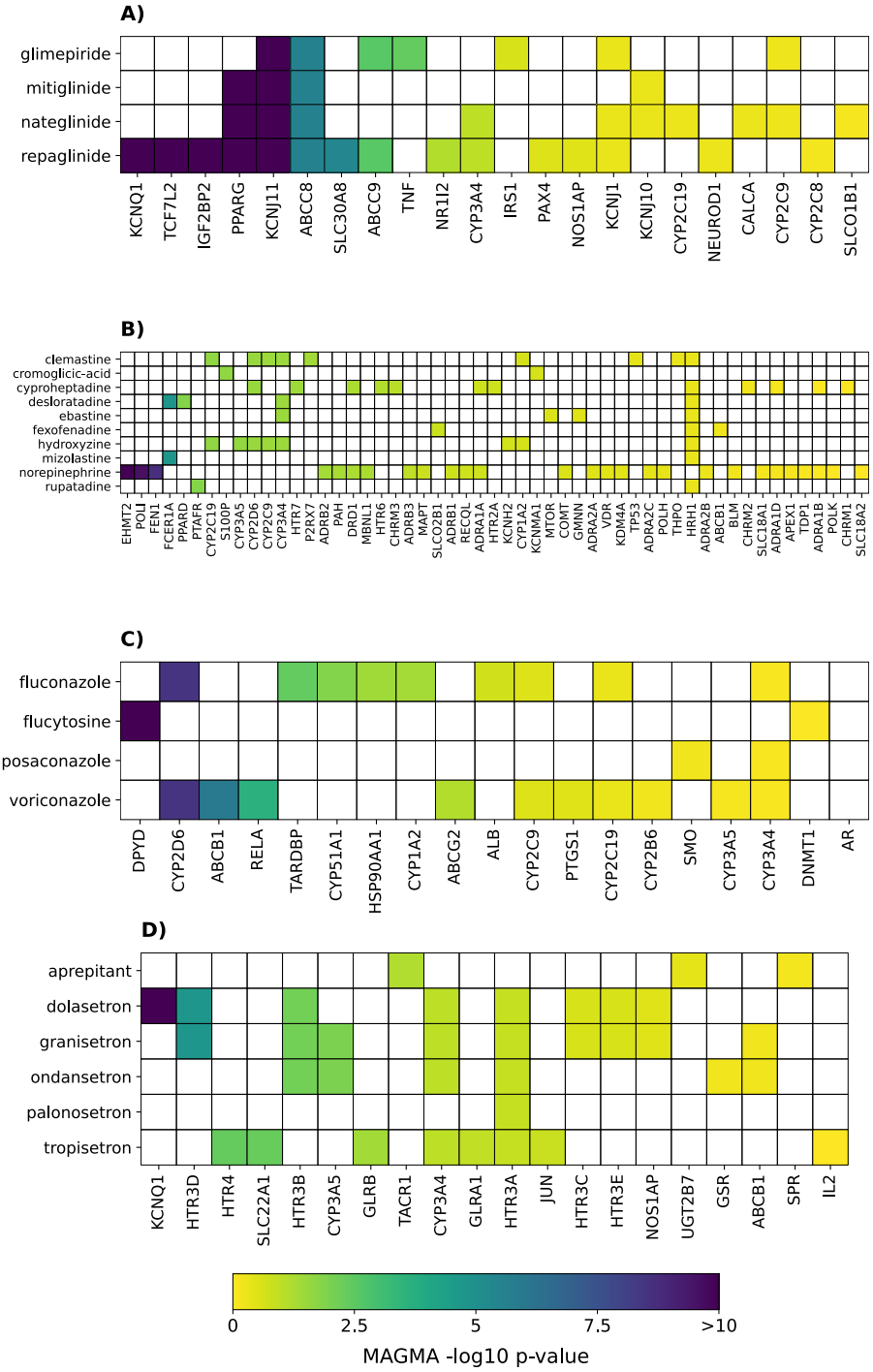

Figure 3 Heatmaps of drug group enrichment results. The heatmaps depict the drug-gene sets from drug groups that were significantly associated with a phenotype, with color signifying the gene's -$\log_{10}$ p-value from MAGMA. Panel **(A)** depicts drug-gene sets from the insulin secretagogues group associated with type 2 diabetes, panel **(B)** illustrates the drug-gene sets from the clinical indication group urticaria associated with asthma, panel **(C)** shows the drug-gene sets from ATC code J02A associated with schizophrenia, and panel **(D)** portrays the drug-gene sets from ATC code A04A associated with Alzheimer's disease. Drug gene set names are listed on the *y*-axis, and gene names are listed on the *x*-axis. Color = -$\log_{10}$ p-value from gene association tests in MAGMA.

Next, we compared the gene targets of approved drugs for each phenotype to further investigate why drug-gene set analysis was better at identifying approved drugs for the non-psychiatric/neurological phenotypes that were tested. As seen in Figure 4, there are 38 gene targets of approved drugs (8.44% of drug target genes) that are significant after Bonferroni correction for the non-psychiatric/neurological phenotypes investigated, compared to 17 significant gene targets (3.21% of drug target genes) for drugs approved for the psychiatric and neurological phenotypes. Moreover, each of the four non-psychiatric/neurological phenotypes has at least 4 significant genes targeted by approved drugs, while bipolar disorder and Parkinson's disease have only 1 significant gene targeted by an approved drug, and Alzheimer's disease has none. This may explain why none of the significant drug groups associated with the psychiatric and neurological phenotypes contained any drugs approved specifically for those disease and disorders.

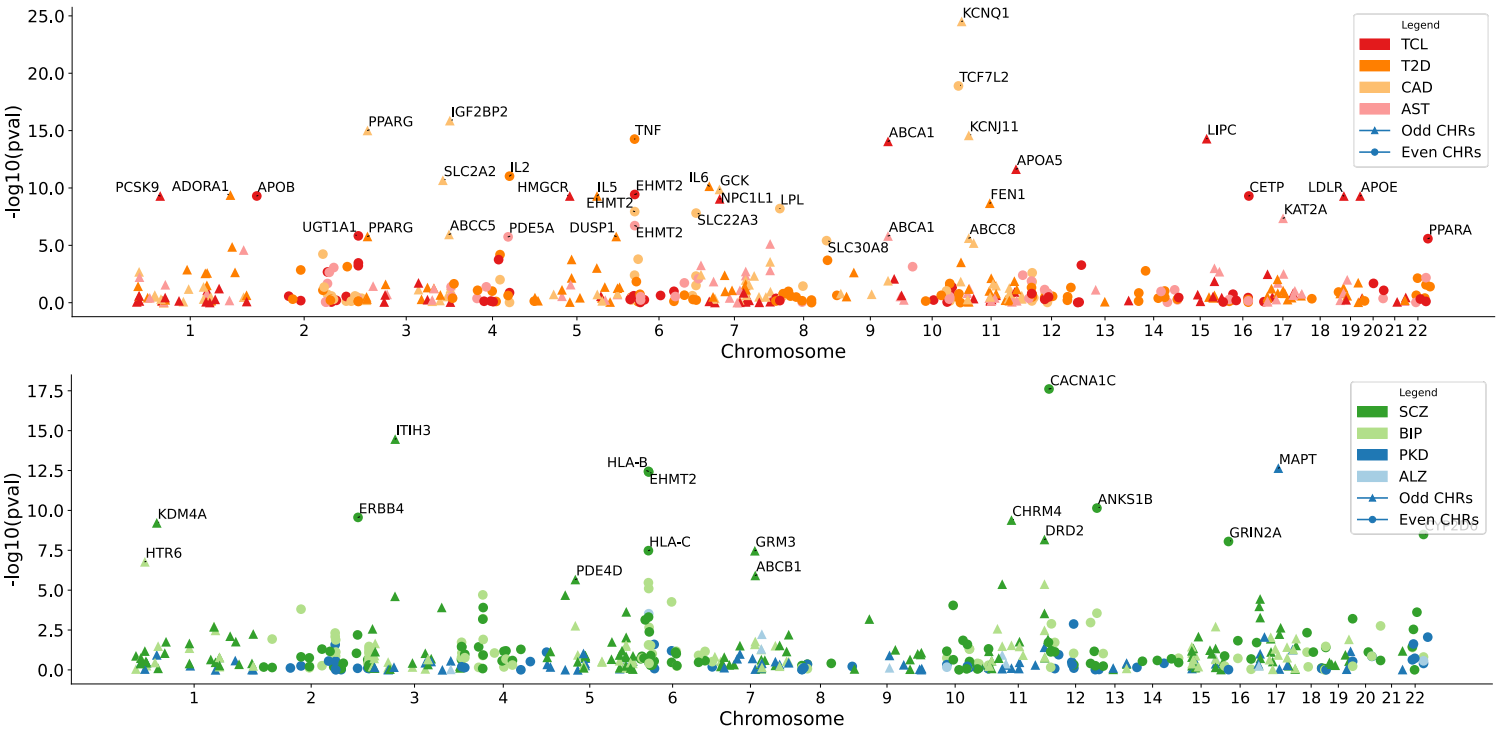

Figure 4 Manhattan plots of gene targets for approved drugs. The top panel depicts the gene-level Manhattan plot for the targets of approved drugs for the four non-psychiatric/neurological phenotypes (n = 450), and the bottom panel depicts the same for the psychiatric/neurological phenotypes (n = 530). Genes that are labeled are significant after Bonferroni correction for their respective phenotype (since the number of genes tested varied per phenotype, different Bonferroni thresholds were applied to each phenotype; see Methods). Color = phenotype; shape = even and odd chromosomes; TCL = hypercholesterolemia; T2D = type 2 diabetes; CAD = coronary artery disease; AST = asthma, SCZ = schizophrenia; BIP = bipolar disorder; PKD = Parkinson's disease; ALZ = Alzheimer's disease.

256 ***Merged drug gene-set analysis***

257 Some drug-gene sets have a very small number of gene targets, resulting in lower power
258 to detect associations with phenotypes[20]. Additionally, many drug categories in our data do not
259 contain enough drugs to test for enrichment. Therefore, to be able to include drug targets with
260 small gene set sizes and to test association for drug categories with low numbers of drugs, we also
261 created gene sets by grouping drugs again by ATC III code, clinical indication, and mechanism of
262 action. For each category within the grouping methods, a single gene set was created made up of
263 all the gene targets for drugs in the group (Figure 1C), resulting in 164 ATC III code gene sets, 515
264 clinical indication gene sets, and 354 mechanism of action gene sets (Tables S22 – S24).
265 Competitive gene-set analysis was performed with the same protocol used when testing individual
266 drug-gene sets for each of the three types of grouped drug-gene sets (Methods). We identified 32
267 significant categorical gene set-trait relationships, with every phenotype but asthma and
268 Parkinson's disease having at least one associated gene set. For all significant categorical drug gene
269 sets see Supplementary Table S25, and for all categorical drug-gene set results see Supplementary
270 Tables S26 – S49.

271

272 ***Drug similarity networks***

273 For each phenotype, we created network graphs to investigate how the top drugs per
274 phenotype cluster based on their Jaccard similarity coefficient[42] (Figure 5), which is calculated
275 using the proportion of gene targets shared by both drugs (see Methods). Central nodes in each
276 network graph were defined using all drugs with a MAGMA competitive gene-set analysis p-value
277 < 0.01 per phenotype. All drugs with a Jaccard similarity of 0.25 or greater with a central node
278 were included as peripheral nodes. Figure 5A displays the similarity network graph for drugs

associated with hypercholesterolemia and Figure 5B illustrates the similarity of drugs associated with schizophrenia.

As may be expected, statin drugs approved for hypercholesterolemia tend to cluster together, while the other two significant drugs, lomitapride and probucol, have no similar compounds (Figure 5A). There is also a cluster of PPAR agonists around the significant drug saroglitazar and nominally significant drug bezafibrate, which both target the PPAR α/γ subunits. Figure 5B demonstrates that almost none of the Bonferroni-significant drugs identified using drug gene-set analysis for schizophrenia cluster with approved schizophrenia drugs, with the exception being one approved drug sharing an edge with benidipine. Most nominally significant drugs for schizophrenia (Figure 5B) appear in the same large cluster, which ranges from antiepileptics in one subcluster (e.g., ethosuximide), calcium channel blockers in the middle (e.g., felodipine), and ACE inhibitors on the other side (e.g., quinapril). The exception is gabapentin, which appears by itself connected to only one other drug, spironolactone. Interestingly, spironolactone has been found to improve schizophrenia symptoms in mice by acting as an antagonist to the NRG1-ERBB4 signaling pathway, and is currently being tested in humans as an add-on treatment for SCZ[43]. The rest of the approved schizophrenia drugs (triangles) appear in the cluster centered around amisulpride, azaperone, domperidone, and perospirone. For network graphs of all phenotypes see the Supplemental materials (Figures S3 – S8).

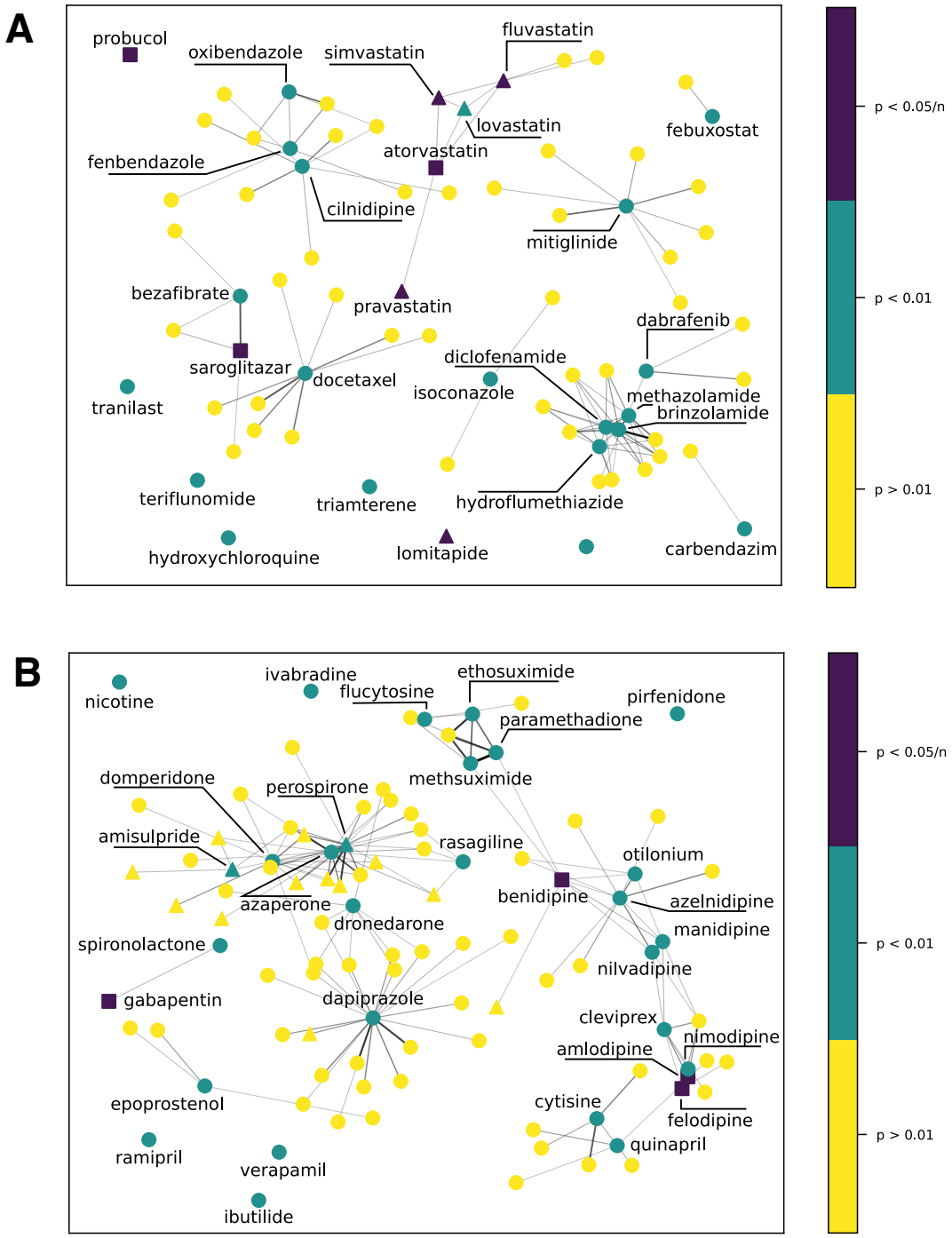

Figure 5 Drug similarity networks. Drug similarity was calculated using the Jaccard index to measure the proportion of shared gene targets between every pair of drugs. Drugs nominally associated with our phenotypes ($p < 0.01$) were used as central nodes, with peripheral nodes being any drug with a Jaccard similarity > 0.25 with any central node. Panel **(A)** displays the results for hypercholesterolemia, and **(B)** shows the network for drugs associated with schizophrenia. Triangle = approved drug for the phenotype, square = Bonferroni-significant drug that is not approved for the phenotype, edge width = Jaccard similarity coefficient.

Discussion

In the current study, we applied drug gene-set analysis to four non-psychiatric/neurological phenotypes and four psychiatric/neurological phenotypes, to investigate the performance of drug gene-set analysis at identifying clinically relevant drugs (i.e., drugs approved to treat the phenotypes we examined), and to identify potential repurposing candidates. We found the first drugs whose targets were significantly enriched for genes associated with hypercholesterolemia, type 2 diabetes, coronary artery disease, schizophrenia, and Alzheimer's disease using this method, and identified drugs with the same mechanism of action (calcium channel blockers) as those published in the latest genome-wide association study of bipolar disorder[44]. In all, 22 drug-trait associations were identified across the eight phenotypes we examined. Notably, 4 of the 4 drugs identified for type 2 diabetes are approved to treat type 2 diabetes, and 3 statin drugs were significantly associated with both coronary artery disease and hypercholesterolemia, respectively[24–26,45,46]. We also demonstrated that drug gene-set analysis does not have to be used to isolate single drugs for repurposing and can also pinpoint categories of drugs by using two different methods: by testing for the enrichment of individual drugs in a category and by combining gene targets of all drugs in a category into a single gene set, thereby also pointing into the direction of putative novel drug development that fall within an identified category. Specifically, we identified drug groups enriched for genetic signal for hypercholesterolemia, type 2 diabetes, coronary artery disease, and asthma that contained approved drugs for those diseases. Additionally, we found that the clinical indication drug groups for hypercholesterolemia, type 2 diabetes, and Parkinson's disease were nominally enriched.

Phenotypes that were not enriched for drugs that are already approved for treatment may not reflect a shortcoming of drug gene-set analysis, but more so the gap between the gene products targeted by current treatments and the genes that are most associated with it as measured by

genome-wide association studies. Thus, while no approved drugs for the psychiatric/neurological phenotypes examined here were identified, this may be because current treatments for these diseases only target disease symptoms and not the underlying disease mechanisms[47–50]. The results presented here establish that drug gene-set analysis can identify clinically relevant drugs – with accuracy varying across the phenotypes we examined – and is able to offer new drug candidates for repurposing.

Across our drug gene-set analysis results there is evidence for repurposable drugs for several phenotypes. For instance, we find repeated evidence for the potential treatment of schizophrenia with calcium channel blockers. The calcium channel blockers felodipine, amlodipine, and benidipine were significantly associated with schizophrenia after Bonferroni correction. The use of calcium channel blockers in psychiatry has shown mixed efficacy in the literature. Previous research has tested the use of calcium channel blockers for psychiatric disorders and found them ineffective, although many of these studies used first-generation drugs (e.g., verapamil)[51]. Recent research has found that calcium channel blocker use is associated with reduced illness severity for schizophrenia and lower rates of psychiatric admission and self-harm during psychosis[52,53]. Additionally, Lintunen et al. found that calcium channel blocker use was associated with lower risk for psychiatric hospitalization in individuals with schizophrenia, an effect which was specific to the calcium channel blocker subclass of dihydropyridines[54]. This is a significant distinction, as drugs in this subclass are also known to have higher selectivity and permeability across the blood brain barrier, perhaps making them better candidates for psychiatric use than first generation calcium channel blockers[51]. Notably, all calcium channel blockers that were identified for both schizophrenia and bipolar disorder, felodipine, amlodipine, benidipine, manidipine, and nilvadipine, are all dihydropyridines[55]. Likewise, one of the two drugs identified using drug gene-set analysis in the 2021 genome-wide association study on bipolar disorder was

355 nisoldipine, which is also a dihydropyridine calcium channel blocker[44]. Thus, further research into
356 the use of dihydropyridines in psychiatry may be warranted.

357 Despite the promising results found here, drug gene-set analysis still leaves several
358 questions unanswered regarding repurposable drugs prioritized with this method. First, drug gene-
359 set analysis does not specifically test for direct associations between a drug and a phenotype.
360 Rather, it tests whether the gene targets of drugs are enriched for genetic signal that is associated
361 with a phenotype. Second, and perhaps most importantly, is that drug gene-set analysis does not
362 indicate the direction of effect, meaning that drugs identified could either alleviate or exacerbate
363 disease progression and symptoms. Additionally, drug gene-set analysis provides no information
364 on the way a drug interacts with its targets (e.g., agonism versus antagonism). These two
365 limitations mean that drug gene-set analysis may best be used in the first steps of a drug
366 repurposing pipeline, or as one of several methods used to triangulate novel drugs for a
367 phenotype[56]. Drugs prioritized with drug gene-set analysis could be followed up using genomics
368 approaches that consider gene expression (e.g., transcriptomic signature matching) or more
369 statistical approaches like Mendelian randomization[15] or local genetic correlation analysis[57], using
370 protein levels of the drug-gene targets for each gene in a drug gene set. This may prove tedious
371 for drugs with large gene sets and may fail to capture interaction effects between different proteins.
372 However, transcriptomic signature matching (e.g., correlating gene expression induced by a drug
373 with the gene expression associated with a phenotype) could be an efficient follow-up for large
374 drug-gene sets. Whilst the limitations above are important to note, they do not outweigh the
375 benefits. Given the excessive financial and time challenges facing psychiatric drug development, a
376 cheap in-silico, and quick method for genetically informed drug repurposing, such as drug gene-
377 set analysis, may prove invaluable. The drug gene-set analysis pipeline used here is publicly
378 available (Methods) and can be used with any user provided GWAS summary statistics.

Materials & Methods

Data. *Clue Repurposing Hub*. The Clue Repurposing Hub is an annotated data set of FDA-approved, clinical trial, and pre-clinical drugs. We extracted gene targets for every FDA-approved drug from the Clue Repurposing Hub that had a known clinical indication, disease category, and mechanism of action (n = 1201).

Drug Gene Interaction Database. To augment the sets of gene targets downloaded from the Clue Repurposing Hub, we queried the Drug Gene Interaction Database API for additional genes targeted by or interacting with each compound in our data set.

GWAS. Summary statistics for hypercholesterolemia, coronary artery disease, and Parkinson's disease were downloaded from GWAS Atlas[58–60] (https://atlas.ctglab.nl/). Summary statistics for type 2 diabetes were downloaded from the DIAGRAM Consortium website (http://diagram-consortium.org/downloads.html)[33], summary statistics for asthma were retrieved from GWAS Catalog (https://www.ebi.ac.uk/gwas/)[61], and those for bipolar disorder and schizophrenia were accessed from the Psychiatric Genomics Consortium website[44,61,62] (https://www.med.unc.edu/pgc/download-results/). Lastly, GWAS summary statistics for Alzheimer's disease (excluding 23andme data) were downloaded from the CTG Lab website[63] (https://ctg.cncr.nl/software/summary_statistics).

Analyses. *Drug gene-set analysis.* Drug gene-sets were created using the gene targets and interactions retrieved from both the Clue Repurposing Hub and Drug Gene Interaction Database for every compound, resulting in 1201 drug gene sets with $n >= 2$. Additionally, we created 3 types of drug gene-set by grouping drugs by clinical indication (n = 515), by mechanism of action

($n = 354$), and by ATC III code ($n = 164$). Competitive gene-set analysis was performed separately with MAGMA v1.09b for each type of drug-gene set, using the 1K Genomes European reference panel[19,64]. Competitive gene set analysis was computed while conditioning on a gene set of every gene target in our data set ($n = 2281$) to assess whether drug gene-set associations were specific to the drug-gene set or driven by common properties of drug target genes.

Drug group associations. Groups of drugs were examined for enrichment of genetic signal using a multiple linear regression model to estimate the effect of drug group membership on MAGMA t-statistics. Given that the MAGMA t-statistic is a measure of the degree to which a drug gene set is enriched for genetic signal for a phenotype, this allows us to test whether drugs in that drug group exhibited (on average) greater enrichment than other drugs.

Denoting the vector of drug gene-set t-statistics as Y and encoding a binary indicator variable G_j for drug group j, scored 1 for drugs in that drug group and 0 otherwise, we fit the model $Y = \beta_0 + C\beta_C + G_j\beta_G + \varepsilon = X\beta + \varepsilon$. Here, C is a covariate matrix containing the drug gene-set size as well as the log of the drug gene-set size, and ε is the residual term with distribution $\varepsilon \sim \mathrm{MVN}(0, \sigma^2 S)$, where S is the sampling correlation matrix of the MAGMA t-statistics. Combining the intercept and predictor variables in X, the model can be estimated using Generalized Least Squares, yielding the estimate $\hat{\beta} = (X^T S^{-1} X)^{-1} X^T S^{-1} Y$, with sampling covariance matrix $W = \hat{\sigma}^2 (X^T S^{-1} X)^{-1}$. For the residual variance we have the estimate $\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T S^{-1} \hat{\varepsilon}}{N-K}$, with $\hat{\varepsilon} = Y - X\hat{\beta}$ and with N the number of drugs and $K = 4$ the number of parameters in the regression equation.

For each drug group j, we test the null hypothesis $H_0: \beta_G = 0$ against the one-sided alternative $H_A: \beta_G > 0$ using the test statistic $T = \frac{\hat{\beta}_G}{\sqrt{W_G}}$, where W_G is the diagonal element of W corresponding to $\hat{\beta}_G$ and T has a t-distribution with $N - K$ degrees of freedom. Correction for

multiple tests was done using a two-stage Bonferroni correction, correcting for the three grouping methods used and then for the number of groups tested within each grouping method. This resulted in a threshold of $0.05/N_g/3$ for each grouping method g with N_g the number of groups for g, where $N_g = 85$ for the ATC III code drug groups, $N_g = 79$ for the mechanism of action drug groups, and $N_g = 118$ for the clinical indication drug groups.

Manhattan Plots of Genes Targeted by Approved Drugs. The Bonferroni-significant thresholds for the gene-level Manhattan plots were derived by $0.05/N_{x,y}$, where N_x is the number of genes tested in MAGMA for a phenotype y. The Bonferroni-significant thresholds are as follows: $p_{hypercholeterolemia} = 0.05/17493 = 2.86\text{E-}06$, $p_{diabetes} = 0.05/11631 = 4.29\text{E-}06$, $p_{cad} = 0.05/18180 = 2.75\text{E-}06$, $p_{asthma} = 0.05/18315 = 2.73\text{E-}06$, $p_{schizophrenia} = 0.05/18265 = 2.74\text{E-}06$, $p_{bipolar} = 0.05/18262 = 2.74\text{E-}06$, $p_{parkinsons} = 0.05/18367 = 2.72\text{E-}06$, $p_{alzheimers} = 0.05/18409 = 2.72\text{E-}06$.

Drug Similarity Network Graphs. Network graphs were created to investigate how the top results cluster based on the similarity of their gene targets. The similarity of each pair of drugs was evaluated using the Jaccard similarity coefficient, which is calculated using the formula $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$ for drug gene sets A and B, where $|A \cap B|$ denotes the number of genes shared by those gene sets and $|A \cup B|$ the number of genes that occurs in at least one of the two gene sets. Central nodes for each phenotype were defined as drug-gene sets with a MAGMA gene-set analysis p-value < 0.01, and all drugs with a Jaccard index $>= 0.25$ with a central node were included as peripheral nodes. Network graphs were created in Python using the package *networkx 2.6.3*.

URLs. CTG Lab, https://ctg.cncr.nl/software/summary_statistics; GWAS Atlas, https://atlas.ctglab.nl/; GWAS Catalog, https://www.ebi.ac.uk/gwas/; Drug Gene Interaction Database, https://www.dgidb.org/; DIAGRAM Consortium, http://diagram-consortium.org/downloads.html); Psychiatric Genomics Consortium, https://www.med.unc.edu/pgc/download-results/; Clue Repurposing Hub, https://clue.io/repurposing; MAGMA, https://ctg.cncr.nl/software/magma; Drug gene-set analysis code, https://github.com/nybell/drugsets.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964874. DP is supported by the Netherlands Organization for Scientific Research - Gravitation project 'BRAINSCAPES: A Roadmap from Neurogenetics to Neurobiology' (024.004.012), and DP and EU are funded by the European Research Council advanced grant 'From GWAS to Function' (ERC-2018-ADG 834057). NYB, EU, EvW, and DP have no conflicts of interest to declare. CdL was funded by F. Hoffman-La Roche AG.

466 **References**

467 1. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.
468 *Nucl. Acids Res.* **42**, D1001–D1006 (2014).
469 2. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation.
470 *The American Journal of Human Genetics* **101**, 5–22 (2017).
471 3. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications.
472 *Nat Genet* **47**, 856–860 (2015).
473 4. King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as
474 likely to be approved? Revised estimates of the impact of genetic support for drug
475 mechanisms on the probability of drug approval. *PLoS Genet* **15**, e1008489 (2019).
476 5. Ochoa, D. *et al.* Human genetics evidence supports two-thirds of the 2021 FDA-approved
477 drugs. *Nature Reviews Drug Discovery* (2022) doi:10.1038/d41573-022-00120-3.
478 6. Wang, K. *et al.* Diverse Genome-wide Association Studies Associate the IL12/IL23 Pathway
479 with Crohn Disease. *The American Journal of Human Genetics* **84**, 399–405 (2009).
480 7. Moschen, A. R., Tilg, H. & Raine, T. IL-12, IL-23 and IL-17 in IBD: immunobiology and
481 therapeutic targeting. *Nat Rev Gastroenterol Hepatol* **16**, 185–196 (2019).
482 8. Dickson, M. & Gagnon, J. P. Key factors in the rising cost of new drug discovery and
483 development. *Nat Rev Drug Discov* **3**, 417–429 (2004).
484 9. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry:
485 New estimates of R&D costs. *Journal of Health Economics* **47**, 20–33 (2016).
486 10. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in
487 pharmaceutical R&D efficiency. *Nat Rev Drug Discov* **11**, 191–200 (2012).
488 11. Hyman, S. E. Psychiatric Drug Development: Diagnosing a Crisis. 11 (2013).

489 12. Lau, A. & So, H.-C. Turning genome-wide association study findings into opportunities for
490 drug repositioning. *Computational and Structural Biotechnology Journal* **18**, 1639–1650
491 (2020).
492 13. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel
493 disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979–986 (2015).
494 14. Schmidt, A. F. *et al.* Genetic drug target validation using Mendelian randomisation. *Nat*
495 *Commun* **11**, 3255 (2020).
496 15. Reay, W. R. & Cairns, M. J. Advancing the use of genome-wide association studies for drug
497 repurposing. *Nat Rev Genet* (2021) doi:10.1038/s41576-021-00387-z.
498 16. So, H.-C. *et al.* Analysis of genome-wide association data highlights candidates for drug
499 repositioning in psychiatry. *Nat Neurosci* **20**, 1342–1349 (2017).
500 17. Gaspar, H. A. & Breen, G. Drug enrichment and discovery from schizophrenia genome-wide
501 association results: an analysis and visualisation approach. *Sci Rep* **7**, 12460 (2017).
502 18. de Jong, S., Vidler, L. R., Mokrab, Y., Collier, D. A. & Breen, G. Gene-set analysis based on
503 the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia.
504 *J Psychopharmacol* **30**, 826–830 (2016).
505 19. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set
506 Analysis of GWAS Data. *PLoS Comput Biol* **11**, e1004219 (2015).
507 20. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-
508 set analysis. *Nat Rev Genet* **17**, 353–364 (2016).
509 21. Gaspar, H. A., Hübel, C. & Breen, G. Drug Targetor: a web interface to investigate the
510 human druggome for over 500 phenotypes. *Bioinformatics* **35**, 2515–2517 (2019).

511 22. Corsello, S. M. *et al.* The Drug Repurposing Hub: a next-generation drug library and
512 information resource. *Nat Med* **23**, 405–408 (2017).
513 23. Freshour, S. L. *et al.* Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with
514 open crowdsource efforts. *Nucleic Acids Research* **49**, D1144–D1151 (2021).
515 24. Plosker, G. L. & Figgitt, D. P. Repaglinide: A Pharmacoeconomic Review of its Use in Type
516 2 Diabetes Mellitus. *PharmacoEconomics* **22**, 389–411 (2004).
517 25. Phillippe, H. M. & Wargo, K. A. Mitiglinide for type 2 diabetes treatment. *Expert Opinion*
518 *on Pharmacotherapy* **14**, 2133–2144 (2013).
519 26. Dahlén, A. D. *et al.* Trends in Antidiabetic Drug Discovery: FDA Approved Drugs, New
520 Drugs in Clinical Trials and Global Sales. *Front. Pharmacol.* **12**, 807548 (2022).
521 27. Maron, D. J., Fazio, S. & Linton, M. F. Current Perspectives on Statins. *Circulation* **101**,
522 207–213 (2000).
523 28. Alonso, R., Cuevas, A. & Mata, P. Lomitapide: a review of its clinical use, efficacy, and
524 tolerability. *Core Evid* **14**, 19–30 (2019).
525 29. Keating, G. M. & Croom, K. F. Fenofibrate: A Review of its Use in Primary Dyslipidaemia,
526 the Metabolic Syndrome and Type 2 Diabetes Mellitus. *Drugs* **67**, 121–153 (2007).
527 30. Libby, P. & Theroux, P. Pathophysiology of Coronary Artery Disease. *Circulation* **111**,
528 3481–3488 (2005).
529 31. Ahmed, S. *et al.* Use of Gabapentin in the Treatment of Substance Use and Psychiatric
530 Disorders: A Systematic Review. *Frontiers in Psychiatry* **10**, (2019).
531 32. Clark, S. D., Van Snellenberg, J. X., Lawson, J. M. & Abi-Dargham, A. Opioid antagonists
532 are associated with a reduction in the symptoms of schizophrenia: a meta-analysis of
533 controlled trials. *Neuropsychopharmacol.* **45**, 1860–1869 (2020).

534 33. Mahajan, A. *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of
535 diverse populations for discovery and translation. *Nat Genet* **54**, 560–572 (2022).
536 34. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-
537 density imputation and islet-specific epigenome maps. *Nat Genet* **50**, 1505–1513 (2018).
538 35. Banno, A., Reddy, A. T., Lakshmi, S. P. & Reddy, R. C. PPARs: Key Regulators of Airway
539 Inflammation and Potential Therapeutic Targets in Asthma. *Nucl Receptor Res* **5**, 101306
540 (2018).
541 36. Berger, W. E., Schenkel, E. J., Mansfield, L. E., & Desloratadine Study Group. Safety and
542 efficacy of desloratadine 5 mg in asthma patients with seasonal allergic rhinitis and nasal
543 congestion. *Ann Allergy Asthma Immunol* **89**, 485–491 (2002).
544 37. Stefanowicz, D. *et al.* Epigenetic modifying enzyme expression in asthmatic airway
545 epithelial cells and fibroblasts. *BMC Pulm Med* **17**, 24 (2017).
546 38. Whelan, R., St Clair, D., Mustard, C. J., Hallford, P. & Wei, J. Study of Novel
547 Autoantibodies in Schizophrenia. *Schizophrenia Bulletin* **44**, 1341–1349 (2018).
548 39. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature*
549 **511**, 421–427 (2014).
550 40. Ma, L., Shcherbina, A. & Chetty, S. Variations and expression features of CYP2D6
551 contribute to schizophrenia risk. *Mol Psychiatry* **26**, 2605–2615 (2021).
552 41. Claeysen, S., Bockaert, J. & Giannoni, P. Serotonin: A New Hope in Alzheimer's Disease?
553 *ACS Chem. Neurosci.* **6**, 940–943 (2015).
554 42. Jaccard, P. The Distribution of the Flora in the Alpine Zone.1. *New Phytologist* **11**, 37–50
555 (1912).

556 43. Hasan, A. *et al.* Add-on spironolactone as antagonist of the NRG1-ERBB4 signaling
557 pathway for the treatment of schizophrenia: Study design and methodology of a multicenter
558 randomized, placebo-controlled trial. *Contemp Clin Trials Commun* **17**, 100537 (2020).
559 44. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases
560 provides new insights into the underlying biology. *Nat Genet* **53**, 817–829 (2021).
561 45. Agrawal, R. The first approved agent in the Glitazar's Class: Saroglitazar. *Curr Drug*
562 *Targets* **15**, 151–155 (2014).
563 46. Aravind Sosale, S., Saboo, B. & Sosale, B. Saroglitazar for the treatment of
564 hypertriglyceridemia in patients with type 2 diabetes: current evidence. *DMSO* 189 (2015)
565 doi:10.2147/DMSO.S49592.
566 47. Rizek, P., Kumar, N. & Jog, M. S. An update on the diagnosis and treatment of Parkinson
567 disease. *CMAJ* **188**, 1157–1165 (2016).
568 48. Cummings, J. New approaches to symptomatic treatments for Alzheimer's disease.
569 *Molecular Neurodegeneration* **16**, 2 (2021).
570 49. Stępnicki, P., Kondej, M. & Kaczor, A. A. Current Concepts and Treatments of
571 Schizophrenia. *Molecules* **23**, 2087 (2018).
572 50. McIntyre, R. S. *et al.* Bipolar disorders. *The Lancet* **396**, 1841–1856 (2020).
573 51. Harrison, P. J., Tunbridge, E. M., Dolphin, A. C. & Hall, J. Voltage-gated calcium channel
574 blockers for psychiatric disorders: genomic reappraisal. *Br J Psychiatry* **216**, 250–253
575 (2020).
576 52. Hayes, J. F. *et al.* Association of Hydroxylmethyl Glutaryl Coenzyme A Reductase
577 Inhibitors, L-Type Calcium Channel Antagonists, and Biguanides With Rates of Psychiatric

578 Hospitalization and Self-Harm in Individuals With Serious Mental Illness. *JAMA Psychiatry*
579 **76**, 382 (2019).
580 53. Van Dyke, P. & Thomas, K. L. Concomitant calcium channel blocker and antipsychotic
581 therapy in patients with schizophrenia: Efficacy analysis of the CATIE-Sz phase 1 data. *Ann*
582 *Clin Psychiatry* **30**, 6–16 (2018).
583 54. Lintunen, J., Lähteenvuo, M., Tiihonen, J., Tanskanen, A. & Taipale, H. Adenosine
584 modulators and calcium channel blockers as add-on treatment for schizophrenia. *npj*
585 *Schizophr* **7**, 1–7 (2021).
586 55. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018.
587 *Nucleic Acids Research* **46**, D1074–D1082 (2018).
588 56. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology.
589 *International Journal of Epidemiology* **45**, 1866–1886 (2016).
590 57. Werme, J., Sluis, S. van der, Posthuma, D. & Leeuw, C. A. de. *LAVA: An integrated*
591 *framework for local genetic correlation analysis*. 2020.12.31.424652
592 https://www.biorxiv.org/content/10.1101/2020.12.31.424652v2 (2021)
593 doi:10.1101/2020.12.31.424652.
594 58. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet*
595 **45**, 1274–1283 (2013).
596 59. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for
597 Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet*
598 *Neurology* **18**, 1091–1102 (2019).
599 60. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex
600 traits. *Nat Genet* **51**, 1339–1348 (2019).

601 61. Han, Y. *et al.* Genome-wide analysis highlights contribution of immune system pathways to
602 the genetic architecture of asthma. *Nat Commun* **11**, 1776 (2020).
603 62. Consortium, T. S. W. G. of the P. G., Ripke, S., Walters, J. T. & O'Donovan, M. C. *Mapping*
604 *genomic loci prioritises genes and implicates synaptic biology in schizophrenia*.
605 2020.09.12.20192922 https://www.medrxiv.org/content/10.1101/2020.09.12.20192922v1
606 (2020) doi:10.1101/2020.09.12.20192922.
607 63. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals
608 identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**, 1276–1282 (2021).
609 64. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
610