bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title: Evaluation of methods incorporating biological function and GWAS summary statistics to accelerate discovery

Authors: Amy Moore[1], Jesse Marks[1], Bryan C. Quach[1], Yuelong Guo[2], Laura J. Bierut[3], Nathan C. Gaddis[1], Dana B. Hancock[1], Grier P. Page[1,4], Eric O. Johnson[1,4]

Affiliations: [1]GenOmics, Bioinformatics, and Translational Research Center, RTI International, Research Triangle Park, NC 27709, USA; [2]GeneCentric Therapeutics, Inc., Cary, NC, USA; [3]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA; [4]Fellow Program, RTI International, Research Triangle Park, NC 27709, USA.

Abstract

Where sufficiently large genome-wide association study (GWAS) samples are not currently available or feasible, methods that leverage increasing knowledge of the biological function of variants may illuminate discoveries without increasing sample size. We comprehensively evaluated 18 functional weighting methods for identifying novel associations. We assessed the performance of these methods using published results from multiple GWAS waves across each of five complex traits. Although no method achieved both high sensitivity and positive predictive value (PPV) for any trait, a subset of methods utilizing pleiotropy and expression quantitative trait loci nominated variants with high PPV (>75%) for multiple traits. Application of functionally weighting methods to enhance GWAS power for locus discovery is unlikely to circumvent the need for larger sample sizes in truly underpowered GWAS, but these results suggest that applying functional weighting to GWAS can accurately nominate additional novel loci from available samples for follow-up studies.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The genome-wide association study (GWAS) has been widely successful for discovering genetic loci contributing to complex traits[1]. Yet, a survey of the GWAS catalog identified 88 traits without genome-wide significant findings despite theoretically adequate sample size[2]. Traits with worse than expected performance even when thousands of cases are available include autism spectrum disorder[3], heart failure[4,5], major depressive disorder (MDD)[6,7], and some addictions[8–11]. Increasing sample size to increase statistical power for discovery is not always practical, as encountered for rare diseases[12], expensive phenotyping[13], phenotypic heterogeneity[14], hard-to-reach or socially disadvantaged populations[15], and population isolates[16]. Our ability to discover trait-associated loci that are ancestry-specific or subject to gene-environment interaction lags in a field where the overwhelming majority of GWAS samples are of European ancestry[17]. Further, increasing sample size sometimes fails to achieve the expected gain in significant loci[18].

Attempts to improve the discovery power of GWAS without increasing sample size by incorporating functional information, defined here as regulatory annotation of variants or evidence of pleiotropy, is not new[19]. An evaluation of gene- and pathway-based GWAS methods found low sensitivity overall for discovery, and that high sensitivity was achieved at the expense of more false positives[20]. Methods to combine GWAS summary statistics with additional information to perform *in silico* functional follow-up are plentiful[21–25] and range from fine-mapping to determining the biological underpinnings of the variant-trait association. Some authors suggest that a secondary usage of these methods is to increase the statistical power of GWAS to identify novel loci. Evaluation of the performance of such methods for locus discovery has been done *ad hoc* for select methods,[21–23,26] but to our knowledge, a comprehensive evaluation of many methods and multiple GWAS traits against objective criteria has not been published.

To identify suitable method(s) for improving GWAS statistical power to uncover novel loci, we performed the largest, most comprehensive evaluation of published functional weighting methods to date: 18 methods applied to multiple waves of GWAS for five diseases and traits. We applied these methods to publicly available GWAS summary statistics and evaluated their ability to nominate novel trait-associated loci that were confirmed by a subsequently larger, more powerful GWAS, henceforth referred to as GWAS1/GWAS2/GWAS3, for the same trait. To represent varying genetic architectures, phenotypic heterogeneity, and gene regulation by tissue type, we selected three psychiatric traits: schizophrenia, bipolar disorder, and MDD available from the Psychiatric Genomics Consortium (PGC); and two blood cell traits: mean platelet volume (MPV) and white blood cell (WBC) counts, available from the UK Biobank.

Results

We selected 17 published functional weighting methods; we also evaluated a suggestive p-value threshold of 1×10^{-5} as an 18th method (Table 1). Nine methods provided results for individual variants, and nine provided gene-based results aggregated across variants. When evaluated on a per-variant basis, the number of potentially novel nominated variants, after excluding statistically significant variants from GWAS1, ranged from zero to 177,698 in the blood cell traits and zero to 4,147 in the psychiatric traits (Supplemental Table 4).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Briefly, we applied each published functional weighting method (Table 1) to genome-wide summary statistics from each GWAS1 study. Details of additional annotation datasets and statistical significance thresholds used for each method are described in the Methods and Supplemental Table 1. To facilitate cross-method comparisons, our primary way to evaluate both variant-based and gene-based method performance used a +/- 500kb window to define a locus, unless specified otherwise. Overlapping loci were merged. To exclude the possibility of methods re-discovering loci already identified as trait-associated in GWAS1, we did not consider loci if they overlapped with a +/- 500kb window surrounding the top variant of a locus that was genome-wide significant in GWAS1. Each functionally weighted GWAS1 was then compared to the corresponding GWAS2 of that trait to identify nominated loci from GWAS1 that overlapped with genome-wide significant loci first identified in GWAS2. A minimum overlap of 250kb was required. Our scheme for defining classification metrics (True Positive [TP], etc.), is illustrated in Supplemental Table 3. Our primary evaluation metrics, Positive Predictive Value (PPV) and Sensitivity (SN), were derived from these classification metrics.

Global Evaluation

No method had both high SN and PPV (>0.50, Figure 1, Quadrant I). In general, there was an inverse relationship between SN and PPV (Figure 1). Quadrant IV, with high SN and low PPV, was dominated by methods providing variant-level results and by the blood cell traits MPV and WBC. Quadrant II, with low SN and high PPV, was dominated by eQTL-based methods, which tended to nominate fewer loci than the variant-level methods (Supplemental Table 5). Exceptions to the pattern of finding eQTL-level methods in Quadrant II were MTAG and the weighted eQTL methods. These methods nominated fewer loci for their respective traits than was typical for other variant-based methods (Supplemental Table 4).

Quadrant III of Figure 1, representing low SN and low PPV, included results from all five traits and a preponderance of MDD, specifically around SN=0 and PPV=0. Only five out of nine methods nominated any variants for MDD (Supplemental Table 4), which had no significant hits in GWAS1. Like the variant-based methods, only four out of nine gene-based methods yielded any nominations for MDD, and none of those overlapped with the GWAS2 hits for MDD, regardless of the evaluation method used (Supplemental Table 5).

We provide representative Manhattan plots[36] to illustrate the performance of two functional weighting methods for the high PPV (MTAG, Figure 2a-b) and high SN (LSMM, Figure 2c-d) scenarios, respectively. When comparing the variants nominated by MTAG for SCZ1, using BPD1 as the pleiotropic trait, relative to both waves of the SCZ GWAS, the nominated variants of MTAG clustered around established “peaks”, including some that are just below the genome-wide significance threshold in GWAS1 (Figure 2a). Some of these variants (e.g., see Chromosomes 3 and 12) are in loci that become significant in GWAS2 (Figure 2b), contributing to the high PPV of this method-trait combination, while others fall below even the suggestive threshold in GWAS2 (e.g., see Chromosome 7). However, these particular non-significant nominated variants are within 500 kb of the novel GWAS2 top hit (Supplemental Table 4).

LSMM with global FDR nominated 3,395 more variants for SCZ than MTAG, resulting in high SN (Supplemental Table 4). In contrast to MTAG, a striking proportion of these nominated variants exhibited a sharp decrease in significance from GWAS1 (Figure 2c) to GWAS2 (Figure 2d), contributing to

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the low variant-based PPV under both FDR options for LSMM; the PPV also remained low in the locus-based evaluation (Supplemental Table 4).

Figure 3 illustrates the performance of gene-based methods. To provide parity in evaluating nominated genes, we calculated gene-based p-values using a modification of MAGMA (see Methods). The gene-based methods nominated fewer loci than the variant-based methods. For both EUGENE and SMR, which were applied using Brain eMETA cohort annotations, nominated genes tended to have higher MAGMA p-values (Figure 3a and 3c) but lower p-values in GWAS2 (Figure 3b and 3d).

Top Method for Positive Predictive Value

Focusing on the ability of methods to accurately nominate loci that were truly trait-associated but inadequately powered for detection in GWAS1, we compared PPV across all traits (Table 2). When multiple databases were applied to a functional weighting method, we chose its highest PPV to carry forward for overall evaluation. Any method ties were all assigned the lowest rank, and methods that failed to nominate any variants/eQTL/genes were ranked lower (NA) than methods with a PPV of 0%. Overall, the best-performing method was MTAG[27], even after a sensitivity analysis excluding the MDD rankings. This ranking was made despite MTAG failing to nominate any variants for MPV (Supplemental Table 4). The best-performing method for MPV alone was TWAS/FUSION, which failed to nominate any loci for MDD.

Consistency of True Associations Nominated Across Methods

We evaluated whether loci nominated by multiple methods are more likely to be TP, as running the same summary statistics through multiple methods is cheaper than conducting a larger GWAS. In general, this was an effective strategy. For example, eight methods was the minimum number necessary to achieve PPV $\geq$ 50% (Supplemental Table 8) for three out of the five traits. For MPV, we did not see a monotonic increase in PPV with larger numbers of nominating methods, and for MDD, only two methods successfully nominated any TP loci. We examined combinations of functional weighting methods to determine if there existed an ensemble set that consistently achieved PPV $\geq$ 50% across traits (Supplemental Figure 4). Across SCZ, BPD, MPV, and WBC, the methods GenoCanyon and LSMM were common to all method ensembles with a minimum of six methods; however, the inclusion of one or both of these methods does not preclude a false positive (FP). None of the ensemble sets could be used to reliably nominate TPs across traits.

Evaluating False Positives

Some loci nominated by the functional weighting methods and labeled as FP by our definition may be truly associated with the trait but remain undiscovered in GWAS2. As a sensitivity analysis, we used GWAS3 waves and calculated the PPV of the nominated loci after removing findings from GWAS1, similar to our primary analysis approach. Figure 4 shows the SN and PPV of the functional weighting methods for the three psychiatric traits based on their GWAS3 waves. GWAS3 were not available for the blood cell traits MPV and WBC. Like Figure 1, no methods appeared in Quadrant I. In general, PPV was higher and SN was lower when using GWAS3, compared to using GWAS2, as the gold standard. A substantial number of the method-trait combinations remained in Quadrant III with low SN and PPV. Supplemental Table 7 shows that no methods had a worse PPV when GWAS3 was used as the gold standard rather than GWAS2. An improved PPV when compared to the larger GWAS3 is expected when

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

additional nominated loci are trait-associated. For BPD and MDD, most methods with any successful nominations still had PPV <50% when compared to GWAS3. For the variant-based methods, only MTAG outperformed the approach of simply using a suggestive p-value threshold in the original GWAS1 when using either GWAS2 or GWAS3 as the gold standard for both SCZ and BPD (Supplemental Table 7).

Evaluating the Stringency of Genome-wide Significance

The evaluated methods do not employ a consistent strategy for multiple testing correction or determination of statistical significance. We used a Bonferroni correction based on the number of valid test statistics for methods that calculated a p-value but did not provide a prespecified significance threshold. To evaluate whether this conservative approach hampered our ability to detect trait-associated loci, we performed a sensitivity analysis by calculating a local FDR and using a q-value of 0.05 as the threshold for statistical significance for those methods previously subjected to a Bonferroni correction. Results were largely unchanged (Supplemental Figure 2), except for a substantial drop for MTAG and JEPEG, which had achieved perfect PPV with some traits when using the Bonferroni correction.

Evaluating the Amount of Overlap

We evaluated the impact of our primary choice for defining a minimum overlap (250kb) between nominated loci and gold standard loci. We performed a sensitivity analysis utilizing different minimum overlaps of one base, 500kb, and 750kb. In general, we found a slight reduction in SN and PPV with increasing size of the required overlap for all five traits (Supplemental Figure 3a-3e). However, we did not find that our results, particularly our high PPV method-trait combinations, were dependent on overlap size.

Discussion

Our comprehensive, multi-method evaluation presents scenarios where functional weighting methods might prove helpful in expanding the number of novel loci uncovered by GWAS in lieu of increased sample size. None of the eighteen methods achieved both high PPV and high SN, which would have been the ideal result: nominating a substantial proportion of TP loci that would be found in the next GWAS wave without nominating excessive FP loci. Instead, our evaluation demonstrated that the use of functional weighting methods presents a tradeoff between high SN and high PPV. MTAG[27] had the best performance overall with respect to PPV, and LSMM with respect to SN.

When comparing functional weighting GWAS results to standard GWAS results from larger sample sizes as the gold standard, the PPV for many method-trait combinations exceeded 50%, indicating that most nominations were trait-associated by the standard defined here. For BPD and SCZ, where GWAS1 were adequately powered to detect genome-wide significant associations, most eQTL-based methods were able to consistently nominate TP loci when compared to GWAS3 as the gold standard; however, SN decreased across method-trait combinations, indicating that functional weighting GWAS methods combined with contemporaneous annotation databases were unable to identify a correspondingly large fraction of the trait-associated variants that can be captured with a larger GWAS sample size incorporating tens of thousands of additional cases. For SCZ, functionally

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

weighted GWAS applied to SCZ1 uncovered 74.7% of the novel loci found in SCZ2, which dropped to 58.1% of the novel loci found in SCZ3.

If the goal is high-confidence nominations, then there is little additional cost to applying multiple functional weighting methods using publicly available annotation data. Our findings did not show an ideal ensemble approach, whereby nominations that intersected a subset of specific methods subsequently became genome-wide significant in later GWAS waves. Instead, for traits that had relatively few genome-wide significant loci identified in GWAS1, we found that increasing the number of functional weighting methods increased the PPV of those nominations. An ensemble approach may be achievable in the future as functional annotation data in disease-relevant tissues and cell types expands, enabling a comparison of methods with more complete annotations across the genome.

Applying functional weighting methods for the discovery of novel loci and variants carries important considerations. First, applying these methods to a GWAS in the "dead zone" of statistical power[18], where no genome-wide significant loci have been identified using standard GWAS, may not provide a reliable approach to find trait-associated variants. Using MDD as an example, only a minority of the tested methods were able to nominate any novel loci for MDD, and few nominated variants were significant in MDD2 (Supplemental Tables 4,5) or MDD3 (Supplemental Table 7). This difficultly in nominating TP loci for MDD suggests that functional annotation is unlikely to overcome insufficient statistical power for GWAS with sample sizes that are far below what is needed to identify robust genome-wide significant loci. For these situations, increasing the GWAS sample size is ideal[6,28,29]. However, if a second, contemporaneous GWAS of a highly pleiotropic trait is available, applying pleiotropy-based methods such as MTAG or GPA may provide an alternative approach. Although identifying the minimum required SNP-based genetic correlation is beyond the scope of this analysis, we note that SCZ and MDD have a SNP genetic correlation ranging from 0.34–0.51, depending on the study[6,30,31,32]. It is also worth considering that while improvements in trans-ethnic GWAS methods boost discovery power[33], uncovering ancestry-specific loci will require investments to increase the sample size of either the ancestry-specific GWAS or the ancestry-specific functional database[34,35].

Second, by using a distance-based locus definition, we could not evaluate whether the nomination captured the putative causal variant or gene identified in GWAS2. For example, MTAG and fgwas successfully nominated the HLA region as associated with MDD. As is typical with findings located in this region, more work is necessary to identify the causal mechanism for the association between HLA and MDD; initial work by the PGC noted that the C4A and C4B genes were unlikely to be causal for MDD[6], although these genes were functionally characterized as potentially causal for SCZ[36]. Subsequent fine-mapping of the classical MHC region by the PGC also did not support variation in C4 genes to be the source of the MDD association[37], though an eQTL-based analysis identified C4A as a candidate risk gene for MDD[38]. The extended HLA region was confirmed in a subsequent GWAS of MDD, though with a different lead SNP[28]. In our study, MTAG used pleiotropy to find what is likely a true association between the extended HLA region and MDD earlier than it could be discovered with the MDD GWAS1 sample size, but this is likely driven by linkage disequilibrium in the region, rather than genuinely shared pleiotropy between causal genes[37].

Third, our study focused on comparing three categories of methods: annotation, pleiotropy, and eQTL. These three categories represent some of the most popular and long-standing methods, collectively with >3,000 citations as of September 2021. Other categories of functional genomic

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

annotation exist, such as methylation[22] and protein[23] QTLs, and were beyond the scope of the present analysis; we expect that their performance would not substantially differ from the methods evaluated here, but we cannot account for significant GWAS2 loci acting through other mechanisms whose functional annotations were not evaluated here. Other mechanisms may explain some of the low SN, but PPV would be unaffected.

Fourth, our definition of "gold standard" using GWAS2 hits assumes that all novel discoveries in GWAS2 are TPs. In the modern GWAS era with independent replication as a best practice, this assumption likely holds for most loci and their variants. By evaluating variants and their broad flanking regions, our approach minimized FNs and FPs caused by changes in lead variants for a given locus across GWAS waves and equalized the playing field for variant- and eQTL/gene-based methods, allowing for simultaneous comparison.

Fifth, our definition of a FP depended on the sample size of GWAS2. Variants or loci nominated by functional weighting methods could be classified as FP when compared against the gold standard of GWAS2, but it is possible that they represent TP associations that GWAS2 remained underpowered to detect. By performing a sensitivity analysis using recently published GWAS3 as our gold standard for the three psychiatric traits, we confirmed that a portion of the genome classified as FP in our primary analysis, with GWAS2 as the gold standard, were trait-associated. In a real-world application, the ability to arrive at this conclusion would require either substantial laboratory follow-up or an increase in GWAS sample size of 2–4 times to bridge the gap between GWAS2 and GWAS3, using the psychiatric traits as representative sample sizes.

Functional annotation databases continue to expand and contribute broadly to understanding human biology and uncovering causal underpinnings of variant-trait associations. Although functionally weighting GWAS is not a substitute for pursuing large samples for well-powered GWAS, these summary statistics-focused methods can be a cost-efficient approach to discovery. Our results show that no method applied systematically across five traits produced both high SN and high PPV. However, when focused on either high SN or high PPV, functional weighting GWAS methods boost statistical power where larger sample sizes are not feasible and the currently available GWAS has generated at least some genome-wide significant loci for the trait of interest. Greater tolerance for FPs can be endured by a research pipeline incorporating inexpensive, high-throughput, and/or *in silico* steps, while a pipeline intended to move GWAS nominations into model organisms may require more confidence that the nominated loci are truly trait-associated. Functional weighted GWAS results can generate leads for follow-up studies of the genetic drivers of complex traits with a reasonable likelihood of being true, particularly for associations that come through multiple methods.

Methods

1. Method Selection

We reviewed the published literature through February 2020 to identify methods that met the following criteria:

1. Categorized as a) annotation-based; b) pleiotropy-based; or c) eQTL-based
2. Utilized GWAS summary statistics, as opposed to individual-level genotype data
3. Implemented using freely-available software or packages.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Provided either method-specific annotation or eQTL files for use with the method, or were amenable to use with publicly available annotation datasets (e.g., GTEx[39])
5. Originally proposed primary or secondary usage included the discovery of novel trait-associated variants, genes, eQTL, or loci.

We found 17 functional weighting methods that met our inclusion criteria. We also evaluated the performance of a "suggestive" p-value threshold, defined as $5\times10^{-8} \leq p < 1\times10^{-5}$ to illustrate the tradeoffs of simply choosing a more liberal p-value cutoff, without the addition of any functional weighting information. The full list of 18 methods evaluated in the present analysis is presented in Table 1. These methods varied in their determination of significant trait associations. For methods that listed specific threshold values for test statistics, we used those thresholds. For methods whose test statistics were p-values and whose authors did not provide a significance threshold, we used a Bonferroni correction on the number of valid p-values output by the method. Details for significant trait association determinations for each method are detailed below and in Supplemental Table 1.

Suggestive

For all traits, we considered "suggestive" variants as those with p-values $< 1\times10^{-5}$ and $\geq 5\times10^{-8}$. To define suggestive loci, we defined a region +/- 500 kilobases surrounding the variant with the smallest suggestive p-value, and collapsed regions that overlapped by any amount into a single locus.

GenoCanyon

We downloaded the prediction scores for the human genome smoothed over 10-kilobase segments[40] (zhaocenter.org/GenoCanyon_Downloads.html) and applied them to each of the five GWAS1 using the signal prioritization software GenoWAP[41]. We used the recommended posterior probability of 0.50 to define statistical significance.

GenoSkyline

We downloaded tissue-specific functional predictions[42] (http://genocanyon.med.yale.edu/GenoSkyline) based on the Roadmap Epigenomics Project (Roadmap) for whole blood and brain tissue and applied them to the blood traits and psychiatric traits, respectively, using the signal prioritization software GenoWAP[41]. We used the recommended posterior probability of 0.50 to define statistical significance.

Weighted eQTL

Following the method of Li et al.[43], we calculated both binary and general eQTL-based weights for all five traits. In each case, we set $\alpha = 0.05$ and power = 0.6. For binary weights, the parameter M was the number of included variants in each GWAS1, respectively, and ε was calculated as the percentage of eSNPs, defined as those with significant eQTL associations in the relevant tissue. We used the significant GTEx v7 Brain Nucleus Accumbens for PGC traits and the significant GTEx v7 Whole Blood for blood cell traits. Weights were then normalized and applied to the downloaded p-values. Statistical significance was defined as $p_{weighted} < 5\times10^{-8}$.

The general eQTL weight was calculated as $\sqrt{(-\log_{10} p_{eQTL})}$ for eSNPs and 1 for all others, where eSNPs are defined as above. Weights were then normalized and applied to the downloaded p-values. Statistical significance was defined as $p_{weighted} < 5\times10^{-8}$. The parameters α, power, and M were also defined as above.

GPA

Genetic analysis incorporating Pleiotropy and Annotation[44] (GPA) was performed using pairwise comparisons between two traits of interest. For each pair of traits, we matched variants on hg19

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chromosome and position. In the case of duplicate variants, the variant with the smaller p-value was retained. We performed GPA with both global and local FDR strategies using a cutoff of 0.05 in both cases to determine statistical significance.

For each of the three PGC traits, we used as pleiotropic traits the remaining two PGC traits. For blood cell traits, we used the second blood cell trait as a pleiotropic trait for the first. Additionally, we used SCZ1[45], BMI[46], height[47], and two GWAS for HDL[48,49]. FDR cutoffs were defined as above in all cases.

MTAG

Multi-Trait Analysis of Genome-wide association summary statistics (MTAG[27]) was performed using pairwise comparisons between two traits of interest. For all traits, we used the subset of downloaded variants with valid rsIDs and allele frequencies in the downloaded GWAS1 summary statistics. Statistical significance was defined as $p_{MTAG} < 5\times10^{-8}$. For each of the three PGC traits, we used as pleiotropic traits the remaining two PGC traits. For blood cell traits, we used the second blood cell trait as a pleiotropic trait for the first. Additionally, we used SCZ1[45], BMI[46], height[47], and two GWAS for HDL[48,49].

fgwas

We combined fgwas[26] with eQTL results from GTEx[39] as the annotation database. Our eQTL dataset of choice was the significant eQTL dataset for expression in the Nucleus Accumbens from GTEx, v7 for PGC traits and the significant GTEx v7 Whole Blood eQTL dataset for blood cell traits. Each significant eQTL was defined as a "segment". All GWAS1 variants whose position fell within the start and end positions[50] of a significant eQTL were assigned to that segment. Variants that remained unassigned to any segment were excluded, along with variants having missing allele frequencies or odds ratios of zero in the downloaded summary statistics. If more than one variant was localized to the same position in GWAS1, the variant with the smallest p-value was retained. We used the default likelihood penalty of 0.2 to run fgwas. Statistical significance was defined as a PPA > 0.9.

Naïve

We applied the functional-weighted GWAS method described by Sveinbjornsson, et al.[51], dubbed here the "naïve" method. This method relies on an annotation classification for each variant into one of four categories, where each category has a Bonferroni-adjusted family-wise error weight reflecting the likelihood of protein function alterations caused by that variant. The categories and p-value thresholds are loss-of-function ($p < 5.5\times10^{-7}$), moderate impact ($p < 1.1\times10^{-7}$), low impact ($p < 1.0\times10^{-8}$), and other ($p < 1.7\times10^{-9}$). We annotated GWAS1 summary statistics using SnpEff software[52] and applied the aforementioned p-value cutoffs according to the annotation category to determine statistical significance.

LSMM

We performed latent sparse mixed model (LSMM[53]) following the example annotations of the method authors, requiring three sets of input: variants and p-values from GWAS summary statistics, ANNOVAR[54], and GenoSkylinePlus[55] using annotations from the original source. We downloaded the hg19 annotations from the ANNOVAR website and used the dbSNP147 database to annotate GWAS1 variants. Annotations were then collapsed into nine categories: downstream, exonic, intergenic, intronic, ncRNA/exonic, ncRNA/intronic, upstream, 3'UTR, and 5'UTR, with each variant assigned a value of 0 or 1 to denote category membership.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

COLOC

For PGC traits, our colocalization[25] dataset of choice was the significant eQTL dataset for expression in the Nucleus Accumbensfrom GTEx, v7[39]. We defined a region to test for colocalized signal as +/- 200 kilobases upstream and downstream from start and stop positions of a single eQTL probe, and included all GWAS1 SNPs contained within that region. This was repeated for all eQTL probes available in the downloaded dataset.

For blood cell traits, we repeated the same procedure using the significant Whole Blood GTEx v7 dataset[39]. For all GWAS1, evidence of statistically significant colocalization was defined as an Approximate Bayes Factor greater than 0.75.

ENLOC

We performed the fastENLOC implementation of the ENLOC method[56]. We downloaded the multi-tissue eQTL annotation derived from GTEx v8[57] hg38 position and provided European LD definition file (https://github.com/xqwen/fastenloc/). We then used LiftOver[58] to convert all five GWAS1 from hg19 to hg38 genomic coordinates. We applied the Nucleus Accumbens eQTL dataset For PGC traits and the Whole Blood eQTL dataset for blood cell traits.

EUGENE

For all traits, we used the subset of downloaded variants with valid rsIDs as GWAS1. We downloaded the required input datasets for gene position from the EUGENE website[59], grouped GTEx brain tissues as the eQTL data for the PGC traits[39], and grouped whole blood eQTL data for the blood cell traits[39,60,61] after performing additional quality control on the whole blood eQTL data to remove discrepant rsIDs. We used Satterthwaite's approximation to calculate the gene-based summary statistics[62]. We then estimated the FDR thresholds using EUGENE and identified the p-value threshold closest to the FDR threshold of 0.05 to determine statistical significance (Supplemental Table 1).

JEPEG

We downloaded the SNP annotation data (v0.2.0) and reference panel (1000 Genomes EUR Phase 1 Release 3)[63] from the JEPEG website (https://dleelab.github.io/jepeg/). For all traits, we used the subset of downloaded variants with valid rsIDs. For blood cell traits, in the case of duplicate rsIDs, we retained the variant with the smaller p-value. Statistical significance of the results was determined by a Bonferroni correction applied to the JEPEG p-value.

MOLOC[25]

For PGC traits, our colocalization dataset of choice was the significant eQTL dataset for expression in the Nucleus Accumbens from GTEx, v7[39]. The methylation dataset used for PGC traits was downloaded from the processed data available on the GEO data repository at accession number GSE74193 and reflects the identification of meQTLs in the prefrontal cortex of 191 schizophrenia patients and 335 controls without psychiatric illness[64]. We defined a region to test for colocalized signal as +/- 200 kilobases upstream and downstream from start and stop positions of a single eQTL probe, and included all GWAS1 SNPs and meQTL probes contained within that region. This was repeated for all eQTL probes available in the downloaded dataset.

For blood cell traits, we repeated the same procedure using the significant Whole Blood GTEx v7 dataset. The methylation dataset used for blood cell traits was the methylation QTL results from the ALSPAC Accessible Resource for Integrated Epigenomics Studies (ARIES)[65] at the middle-aged timepoint (https://data.bris.ac.uk/data/dataset/r9bxayo5mmk510dczq6golkmb).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sherlock

Only variants with valid rsIDs were submitted to Sherlock for each GWAS. For the PGC traits, the Sherlock-provided eQTL data was chosen as GTEx v7 Brain – Nucleus accumbens. Sample sizes from Supplemental Table 2 were used, and disease prevalence was taken as 0.5% for schizophrenia[66], 1% for bipolar disorder[67], and 15% for major depressive disorder[68]. For the UK Biobank traits, the Sherlock-provided eQTL data was chosen as GTEx v7 Whole Blood, and the sample sizes from Supplemental Table 2 were entered for sample size. As Sherlock output is sometimes presented as a gene symbol and sometimes as an Ensembl gene ID (ENSG), we used the GENCODE annotations[50] to match gene symbols to Ensembl IDs and evaluated the overlap with our gold standards using Ensembl IDs.

SMR

For all GWAS1 inputs, we used the subset of downloaded variants with valid rsIDs and valid allele frequencies. Formatted eQTL data were downloaded from the SMR website (https://cnsgenomics.com/software/smr/#DataResource). For PGC traits, we evaluated the performance of SMR using three different eQTL datasets: GTEx v7 data from the Brain Nucleus Accumbens[39], the "lite" version of the GTEx v7 data from the Brain Nucleus Accumbens, and the Brain-eMeta eQTL data[69] derived from a meta-analysis of GTEx brain, Common Mind Consortium[70], and ROSMAP consortium[71] studies. For UK Biobank traits, we evaluated the performance of SMR using the GTEx v7 data from Whole Blood[39], both full and "lite" versions. For all datasets, we evaluated with and without the requirement for a Heidi p-value of < 0.05 to exclude trait-eQTL associations due to pleiotropy.

TWAS

For all GWAS1 inputs, we used the subset of downloaded variants with valid rsIDs. We downloaded reference LD data for the 1000 Genomes EUR samples provided by the Broad Institute Alkes Group (https://data.broadinstitute.org/alkesgroup/FUSION/). We downloaded (https://gusevlab.org/projects/fusion/) and applied pre-computed gene expression weights for GTEx v7 Brain Nucleus Accumbens for PGC traits and Whole Blood for blood traits[39].

UTMOST

For all GWAS1 inputs, we used the subset of downloaded variants with valid rsIDs. We used the pre-calculated covariance matrices using the 44 GTEx v7[39] tissues (https://github.com/Joker-Jerome/UTMOST). For all five GWAS1, we evaluated the full cross-tissue expression UTMOST results. We additionally evaluated the single-tissue UTMOST output for the Nucleus Accumbens for PGC traits and Whole Blood for blood traits.

2. Model Trait Selection

We evaluated the performance of the functional weighting methods using five traits that have published GWAS with publicly available summary statistics. We deliberately chose early phase GWAS (which we refer to as GWAS1) for each trait to allow for validation of results in subsequent GWAS for the traits (referred to as GWAS2 and/or GWAS3). We evaluated three traits with summary statistics available from the Psychiatric Genomics Consortium (PGC): schizophrenia[45] (SCZ), bipolar disorder[72] (BPD), and major depressive disorder[14] (MDD). We also evaluated two blood cell traits examined in the UK Biobank, mean platelet volume (MPV) and white blood cell count (WBC)[73], as examples of traits with a larger explained heritability, many genome-wide significant loci, minimal heterogeneity in phenotyping, and comprehensive tissue-specific functional annotations. Additional details of the GWAS used to test the functional weighting methods are presented in Supplemental Table 1.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We used Liftover[74] to convert the GWAS of the psychiatric traits from hg18 to hg19. For MPV, p-values were truncated at 7.41×10^{-323}, due to the extremely small p-values not being read into R (v3.6.0).

3. Definition of a Gold Standard

For comparison to each GWAS1, we used as our "gold standard" a larger, more powerful GWAS, hereafter referred to collectively as GWAS2, performed on the same trait and by the same consortium to reduce variability in findings due to differences in trait definition, analytic strategies, or recruitment of study participants (Supplemental Table 2). Our gold standard "hits" for each GWAS2 were defined as those variants meeting the standard genome-wide significance threshold of 5×10^{-8}. We defined a significant locus as the region extending +/- 500 kilobases from the variant with the smallest p-value. Additional variants with genome-wide significant p-values within this region were included within the locus of the lead variant. This procedure was repeated in a stepwise fashion until all genome-wide significant variants were captured. As a final step, overlapping one-megabase intervals were combined into a single locus, and the extended HLA region was defined as the region spanning from base pair 25,000,000 to 35,000,000 on chromosome 6.

4. Exclusion of Significant GWAS1 Hits

All GWAS1 contained statistically significant loci except for MDD (Supplemental Table 2). To avoid giving credit to the functional weighting methods for "re-discovering" these significant loci, we excluded them from evaluation after applying the functional weighting method to GWAS1. We defined a significant locus in GWAS1 as the region extending +/- 500 kilobases from the variant with the smallest p-value. Additional variants with genome-wide significant p-values within this region were included within the locus of the variant of the smallest p-value. This procedure was repeated in a stepwise fashion until all genome-wide significant variants were accounted for. As a final step, overlapping one-megabase intervals were combined into a single locus and the extended HLA region was defined as above.

To exclude GWAS1 hits from the set of GWAS2 "gold standard" hits available for discovery, we used the GenomicRanges R package[75] to remove from GWAS2 any loci with any degree of overlap with the defined GWAS1 significant loci.

5. Evaluation Metrics

Because we focused on method performance to discover novel GWAS hits, our evaluations were based on calculating sensitivity (SN), positive predictive value (PPV), and the F1 score (F1, the harmonic mean of SN and PPV). Definitions can be found in Supplemental Table 3. Because variants with non-significant p-values in GWAS1 may be truly associated with the trait, but GWAS1 was not statistically powerful enough to uncover their associations, we avoided evaluation metrics that depend on the definition of a TN.

6. Evaluation of Variant-Level Methods

Nine functional weighting methods, including the use of a suggestive p-value threshold, provided results for individual genetic variants (Table 1). To evaluate the performance of these nine methods on a per-variant level, TP variants were defined as those with matching chromosome and position that were both genome-wide significant in GWAS2 and nominated as significant by the functional weighting method either by the threshold specified by the method or, if no threshold was explicitly stated, by a Bonferroni multiple testing-corrected threshold (Supplemental Table 1). To exclude variants that were statistically significant in GWAS1, we excluded variants within the +/- 500kb boundaries of GWAS1 hits defined above (Methods Section 4).

Because the functional weighting methods cannot account for secondary signals and some do not account for linkage disequilibrium, we also calculated SN, PPV, and F1 using a locus-based definition of statistical significance. In this evaluation, we defined each locus in the same manner as we identified GWAS1 significant hits (Methods Section 4). A TP was defined as an overlap of at least 250kb in the 1 MB flanking window of a top locus in GWAS2, as defined above, and a +/- 500kb window of a variant

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nominated by a functional weighting method. A FP was defined as a +/- 500kb window nominated by a functional weighting method with less than 250kb overlap among any GWAS2 loci. A FN was defined as a GWAS2 locus with less than 250kb overlap with any locus nominated by the functional weighting method being evaluated. This locus-to-locus comparison was performed assessing any degree of overlap between the gold standard GWAS2 loci, excluding the significant GWAS1 loci, and the loci calculated from the results of the functionally weighted GWAS1 using the GenomicRanges package[75] in R.

7. Evaluation of eQTL-Level Methods

To comparably evaluate methods that yield results on the level of eQTL or gene, we calculated transcript- or gene-based p-values using MAGMA[76]. As most of the eQTL data used in these comparisons came from GTEx, we downloaded and used their GENCODE annotations[50] for transcript/gene names and genomic locations. Statistical significance for GWAS2 was determined at a GWAS-specific Bonferroni correction to the MAGMA p-value after excluding eQTL-based gene results that did not yield a MAGMA p-value.

For methods that did not provide a significance threshold, we first excluded any results that did not result in a valid statistic, then performed a Bonferroni correction based on the number of remaining tests. To exclude established significant loci from GWAS1, we excluded nominated transcripts/genes where the midpoint of the genomic location was within +/- 500kb of the GWAS1 loci, defined above (Methods Section 4).

For MAGMA-based evaluations, TP, FP, and FN were determined by matching either the Ensembl ID or gene symbol, depending on what was used by the particular functional weighting method (Supplemental Table 1), to the output of our modified MAGMA analysis to each GWAS2. A TP was defined as an eQTL/gene that was nominated as significant by the functional weighting method and identified as statistically significant by MAGMA as described above. FP and FN were defined analogously, and we calculated SN, PPV, and F1.

We also conducted locus-based evaluations in two other ways. The first was to use the boundaries of the nominated eQTL/gene, either defined by the functional weighting method when provided or the GENCODE annotation boundaries used to generate the MAGMA p-values (Supplemental Table 1). The second approach was to define the locus boundaries for a functionally weighted eQTL/gene as +/- 500kb from the midpoint of the previously stated boundaries. To avoid possible double-counting, we merged overlapping eQTL/genes into a single locus. Loci were determined in a similar fashion as before (Methods Section 4) using the midpoint of the GENCODE-defined start and end positions, with no truncation at the ends of chromosomes or centromeres, with the exception of EUGENE, where we used the chromosome and position defined by EUGENE output.

We performed a locus-to-locus comparison by looking for a minimum of 250kb of overlap when nominations were defined as +/- 500kb from the midpoint, and 2500 bases of overlap when nominations were defined by the start and end positions using the GENCODE annotation boundaries between the gold standard loci calculated from GWAS2 (Methods Section 3) and the loci calculated from the results of the functionally weighted GWAS1 using the GenomicRanges package[75] in R.

8. Generation of UpSet Plots

To identify an optimal ensemble approach, we examined the overlap among nominations across functional weighting methods for each trait by generating UpSet plots. Plots were generated using the ComplexUpset package [77,78] in R. To construct the UpSet plot, for each trait, functional weighting GWAS methods were ordered from largest to smallest number of nominated loci, defined using +/- 500kb from either the top variant or gene midpoint. For methods with multiple options, the top performing option was selected based on largest PPV. A matrix of nominated loci vs the fwGWAS methods was created in a stepwise fashion. The method nominating the largest number of novel loci was populated first, and then each of its nominated loci was tested for overlap of at least 250 kilobases with loci nominated by all other methods and these overlaps populated the matrix. For each

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

subsequent functional weighting GWAS method, only nominated loci that had not been found to overlap with loci from previously examined methods were added to the matrix. These new additions were then checked for overlap with all remaining method nominations, and all methods nominating a new locus were noted on the matrix.

9. Application of 18 Functional Weighting Methods to Model Traits

Full details of the application of each functional weighting method can be found in the Supplemental Methods, with details of significance cutoffs and functional databases presented in Supplemental Table 1. Briefly, we used the default inputs, external databases, and statistical significance cutoffs recommended by the method developers to the full extent that they were provided. When statistical significance cutoffs were not provided, we applied a standard threshold of either a Bonferroni-corrected p-value or a false discovery rate cutoff of 0.05, as appropriate for the statistics calculated by the functional weighting method.

For the choice of functional database to use with each method, our default was to use a preformatted database provided by the method developers (e.g., TWAS/FUSION). When multiple databases were made available (e.g., SMR), we chose the largest database representing a tissue type appropriate to the model trait being evaluated.

When no functional database was made available by the method authors (e.g., COLOC), we used the statistically significant GTEx v7 nucleus accumbens data downloaded from the GTEx data portal to apply the functional weighting methods to the three psychiatric traits[79–81] and the corresponding statistically significant GTEx v7 whole blood data for the two blood cell traits[39].

We investigated the performance of the pleiotropy-based methods GPA and MTAG in each psychiatric trait using contemporaneous GWAS of the other two psychiatric traits. For the blood cell traits, we used a variety of potentially omnigenetic traits: SCZ[45], HDL cholesterol[48,49], BMI[46], height[47], and the other blood cell trait[73].

10. Sensitivity Analyses

To determine whether the 250kb overlap between a nomination and a novel GWAS2 locus impacted our results, we tested overlaps of 1 base, 500kb, and 750kb, used to replace the +/- 500kb and Ensembl locus definitions.

We investigated whether a less stringent cutoff resulted in better performance by applying an FDR significance cutoff for those methods (suggestive, MTAG, Weighted eQTL, JEPEG, TWAS/FUSION, and UTMOST) for which we used a Bonferroni multiple testing correction. The FDR correction was implemented using the fdrtool R package[82] and a cutoff of $q < 0.05$ was used to determine statistical significance.

We sought to determine the accuracy of our FP definition by using wave 3 GWAS, hereafter referred to as GWAS3, recently released by the PGC for the three psychiatric traits. Known loci from GWAS1 were excluded as described above.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Citations

1. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
2. Watanabe, K. *et al.* Author Correction: A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **52**, 353–353 (2020).
3. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism* **8**, 21 (2017).
4. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
5. Arvanitis, M. *et al.* Genome-wide association and multi-omic analyses reveal ACTN2 as a gene linked to heart failure. *Nat. Commun.* **11**, 1122 (2020).
6. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
7. Ormel, J., Hartman, C. A. & Snieder, H. The genetics of depression: successful genome-wide association studies introduce new challenges. *Transl. Psychiatry* **9**, 114 (2019).
8. Zhou, H. *et al.* Association of OPRM1 Functional Coding Variant With Opioid Use Disorder. *JAMA Psychiatry* **77**, 1072 (2020).
9. Gelernter, J. *et al.* Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Mol. Psychiatry* **19**, 717–723 (2014).
10. Pasman, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal effect of schizophrenia liability. *Nat. Neurosci.* **21**, 1161–1170 (2018).
11. Hancock, D. B., Markunas, C. A., Bierut, L. J. & Johnson, E. O. Human Genetics of Addiction: New Insights and Future Directions. *Curr. Psychiatry Rep.* **20**, 8 (2018).
12. Zhang, C. *et al.* Common genetic variation and risk of osteosarcoma in a multi-ethnic pediatric and adolescent population. *Bone* **130**, 115070 (2020).
13. Mol, C. L. *et al.* Polygenic Multiple Sclerosis Risk and <scp>Population-Based</scp> Childhood Brain Imaging. *Ann. Neurol.* **87**, 774–787 (2020).
14. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
15. Bonevski, B. *et al.* Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Med. Res. Methodol.* **14**, 42 (2014).
16. Hatzikotoulas, K., Gilly, A. & Zeggini, E. Using population isolates in genetic association studies. *Brief. Funct. Genomics* **13**, 371–377 (2014).
17. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
18. Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *Am. J. Psychiatry* **175**, 15–27 (2018).
19. Liu, J. Z. *et al.* A Versatile Gene-Based Test for Genome-wide Association Studies. *Am. J. Hum. Genet.* **87**, 139–145 (2010).
20. Wojcik, G. L., Kao, W. L. & Duggal, P. Relative performance of gene- and pathway-level methods as secondary analyses for genome-wide association studies. *BMC Genet.* **16**, 34 (2015).
21. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
22. Wu, C. & Pan, W. Integration of methylation QTL and enhancer–target gene maps with schizophrenia GWAS summary results identifies novel genes. *Bioinformatics* **35**, 3576–3583 (2019).
23. Wang, J., Zheng, J., Wang, Z., Li, H. & Deng, M. Inferring Gene-Disease Association by an

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Integrative Analysis of eQTL Genome-Wide Association Study and Protein-Protein Interaction Data. *Hum. Hered.* **83**, 117–129 (2018).

24. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, e1004383 (2014).
25. Giambartolomei, C. *et al.* A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**, 2538–2545 (2018).
26. Pickrell, J. K. Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
27. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
28. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
29. Levey, D. F. *et al.* Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat. Neurosci.* **24**, 954–963 (2021).
30. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
31. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
32. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
33. Atkinson, E. G. *et al.* Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* **53**, 195–204 (2021).
34. Hsiao, C.-L., Lian, I.-B., Hsieh, A.-R. & Fann, C. S. Modeling expression quantitative trait loci in data combining ethnic populations. *BMC Bioinformatics* **11**, 111 (2010).
35. Shang, L. *et al.* Genetic Architecture of Gene Expression in European and African Americans: An eQTL Mapping Study in GENOA. *Am. J. Hum. Genet.* **106**, 496–512 (2020).
36. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
37. Glanville, K. P. *et al.* Classical Human Leukocyte Antigen Alleles and C4 Haplotypes Are Not Significantly Associated With Depression. *Biol. Psychiatry* **87**, 419–430 (2020).
38. Gerring, Z. F., Gamazon, E. R. & Derks, E. M. A gene co-expression network-based analysis of multiple brain tissues reveals novel genes and molecular pathways underlying major depression. *PLOS Genet.* **15**, e1008245 (2019).
39. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
40. Lu, Q. *et al.* A Statistical Framework to Predict Functional Non-Coding Regions in the Human Genome Through Integrated Analysis of Annotation Data. *Sci. Rep.* **5**, 10576 (2015).
41. Lu, Q., Yao, X., Hu, Y. & Zhao, H. GenoWAP: GWAS signal prioritization through integrated analysis of genomic functional annotation. *Bioinformatics* **32**, 542–548 (2016).
42. Lu, Q., Powles, R. L., Wang, Q., He, B. J. & Zhao, H. Integrative Tissue-Specific Functional Annotations in the Human Genome Provide Novel Insights on Many Complex Traits and Improve Signal Prioritization in Genome Wide Association Studies. *PLOS Genet.* **12**, e1005947 (2016).
43. Li, L. *et al.* Using eQTL weights to improve power for genome-wide association studies: a genetic study of childhood asthma. *Front. Genet.* **4**, (2013).
44. Chung, D., Yang, C., Li, C., Gelernter, J. & Zhao, H. GPA: A Statistical Approach to Prioritizing GWAS Results by Integrating Pleiotropy and Annotation. *PLoS Genet.* **10**, e1004787 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).
46. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
47. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
48. Willer, C. J. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**, 161–169 (2008).
49. Neale, B. M. No Title. http://www.nealelab.is/uk-biobank/.
50. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
51. Sveinbjornsson, G. *et al.* Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat. Genet.* **48**, 314–317 (2016).
52. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* **6**, 80–92 (2012).
53. Ming, J. *et al.* LSMM: a statistical approach to integrating functional annotations with genome-wide association studies. *Bioinformatics* **34**, 2788–2796 (2018).
54. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
55. Lu, Q. *et al.* Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late-onset Alzheimer's disease. *PLOS Genet.* **13**, e1006933 (2017).
56. Wen, X., Pique-Regi, R. & Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLOS Genet.* **13**, e1006646 (2017).
57. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (80-.).* **369**, 1318–1330 (2020).
58. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
59. Ferreira, M. A. R. *et al.* Gene-based analysis of regulatory variants identifies 4 putative novel asthma risk genes related to nucleotide synthesis and signaling. *J. Allergy Clin. Immunol.* **139**, 1148–1157 (2017).
60. Battle, A. *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* **24**, 14–24 (2014).
61. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
62. Bakshi, A. *et al.* Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Sci. Rep.* **6**, 32894 (2016).
63. Lee, D. *et al.* JEPEG: a summary statistics based tool for gene-level joint testing of functional variants. *Bioinformatics* **31**, 1176–1182 (2015).
64. Jaffe, A. E. *et al.* Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* **19**, 40–47 (2016).
65. Gaunt, T. R. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* **17**, 61 (2016).
66. WU, E. Q., SHI, L., BIRNBAUM, H., HUDSON, T. & KESSLER, R. Annual prevalence of diagnosed schizophrenia in the USA: a claims data analysis approach. *Psychol. Med.* **36**, 1535–1540 (2006).
67. Craddock, N. & Sklar, P. Genetics of bipolar disorder: successful start to a long journey. *Trends Genet.* **25**, 99–105 (2009).
68. Kessler, R. C. & Bromet, E. J. The Epidemiology of Depression Across Cultures. *Annu. Rev. Public*

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Health **34**, 119–138 (2013).

69. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).
70. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
71. Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
72. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43**, 977–983 (2011).
73. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429.e19 (2016).
74. Hinrichs, A. S. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
75. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
76. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
77. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).
78. Krassowski, M. ComplexUpset. doi:10.5281/zenodo.3700590.
79. McCollum, L. A. & Roberts, R. C. Uncovering the role of the nucleus accumbens in schizophrenia: A postmortem analysis of tyrosine hydroxylase and vesicular glutamate transporters. *Schizophr. Res.* **169**, 369–373 (2015).
80. Xu, L., Nan, J. & Lan, Y. The Nucleus Accumbens: A Common Target in the Comorbidity of Depression and Addiction. *Front. Neural Circuits* **14**, (2020).
81. Whittaker, J. R., Foley, S. F., Ackling, E., Murphy, K. & Caseras, X. The Functional Connectivity Between the Nucleus Accumbens and the Ventromedial Prefrontal Cortex as an Endophenotype for Bipolar Disorder. *Biol. Psychiatry* **84**, 803–809 (2018).
82. Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

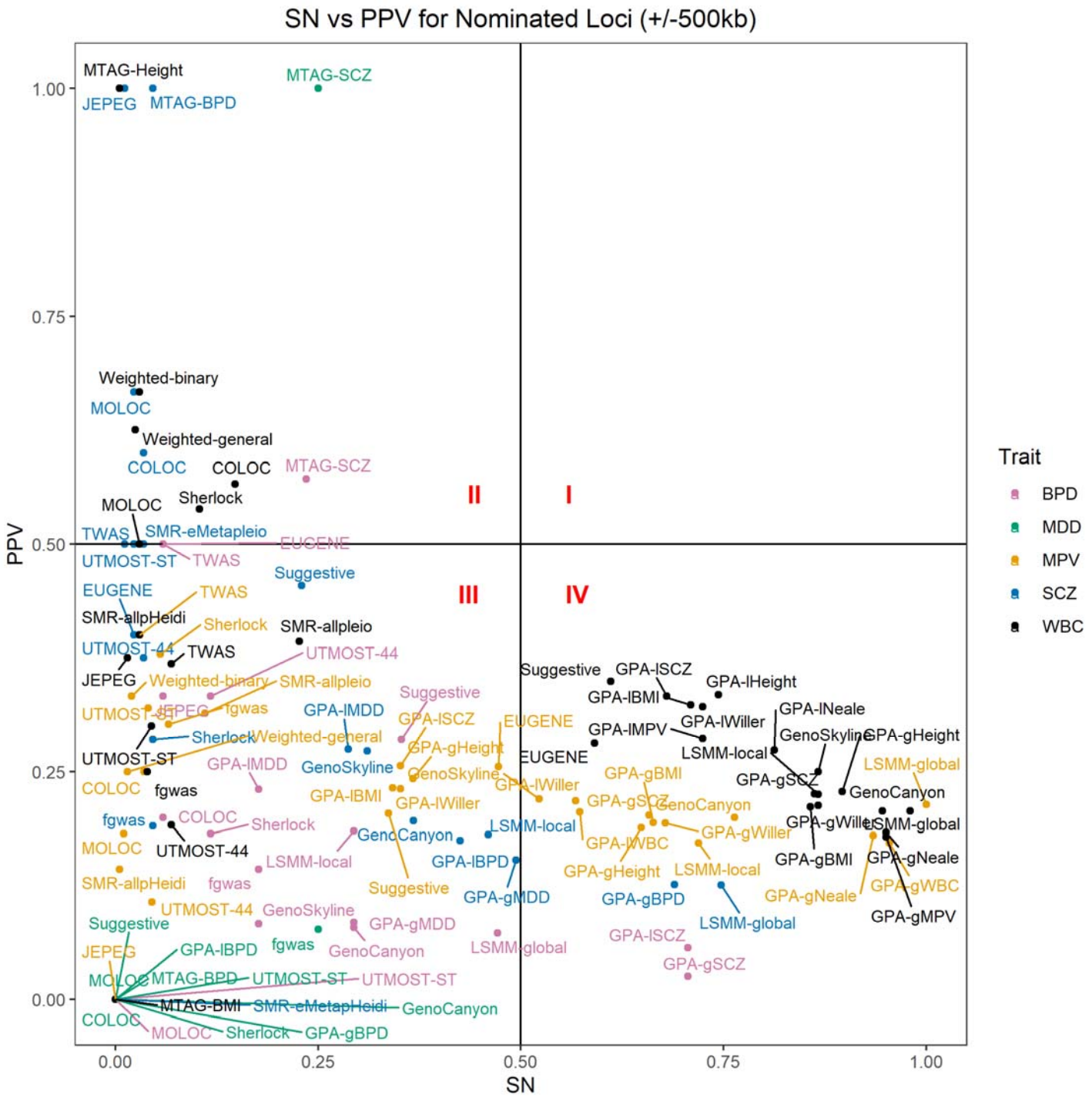
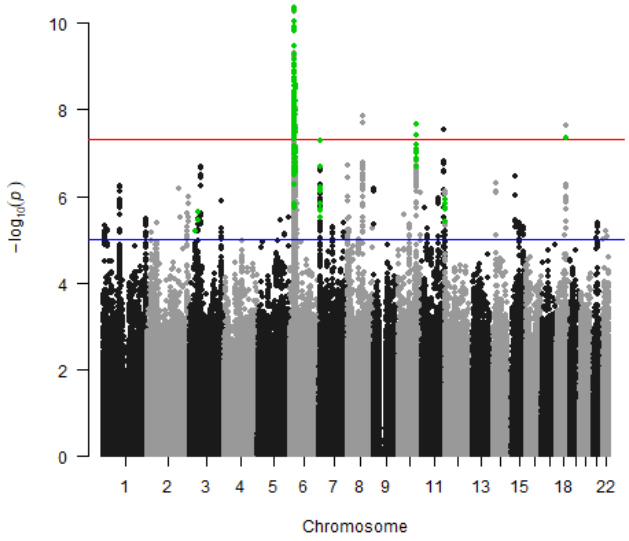


Figure 1. Scatterplot of the relationship between sensitivity (SN; proportion of loci that are significant exclusively in the second wave GWAS [GWAS2] that are also nominated by a given method when applied to GWAS1) and positive predictive value (PPV; proportion of all nominated loci by a given method when applied to GWAS1 that are also significant exclusively in GWAS2) for all method-trait combinations whose results contained at least one gene or locus that was nominated as trait-associated by each method, respectively, after excluding loci identified in GWAS1. SN and PPV were calculated using the +/- 500kb locus-based evaluation and requiring a minimum overlap of 250kb between nominated loci and GWAS2 significant loci. Horizontal and vertical lines denote PPV and SN of 50%, respectively.

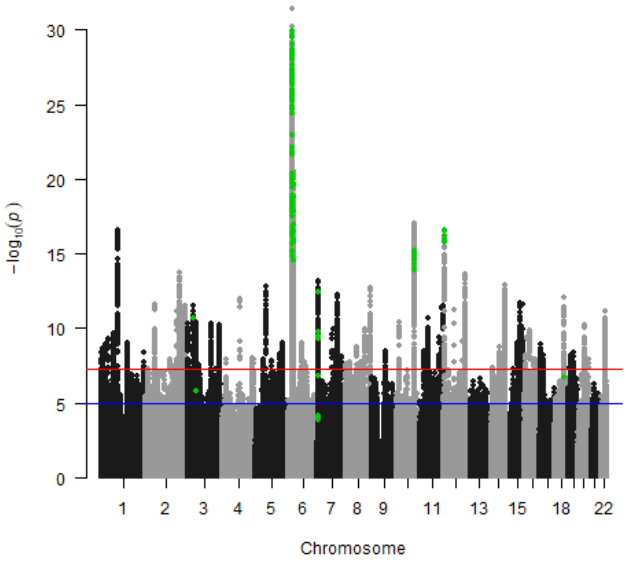
bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MTAG-SCZ1-nominated variants in SCZ1



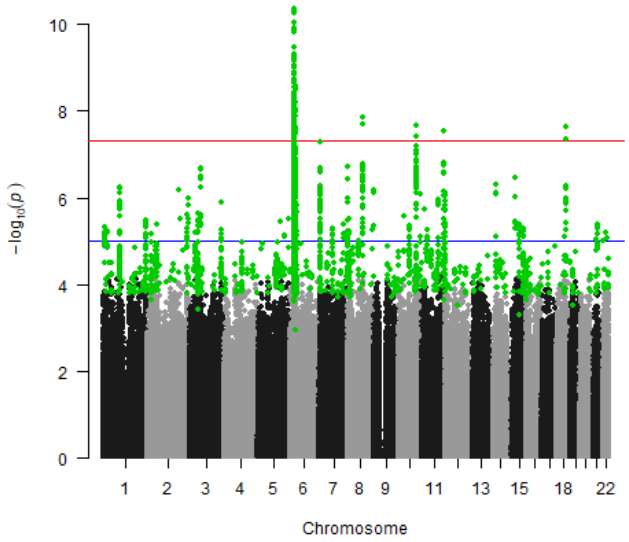
(a)

MTAG-SCZ1-nominated variants in SCZ2



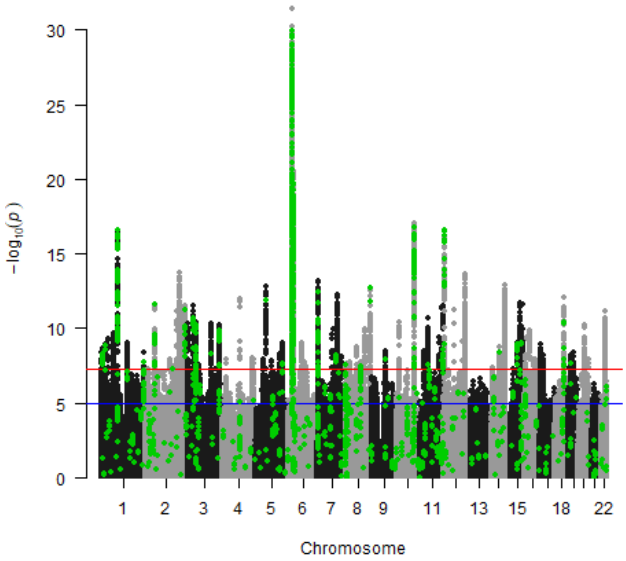
(b)

LSMM-SCZ1-nominated variants in SCZ1



(c)

LSMM-SCZ1-nominated variants in SCZ2



(d)

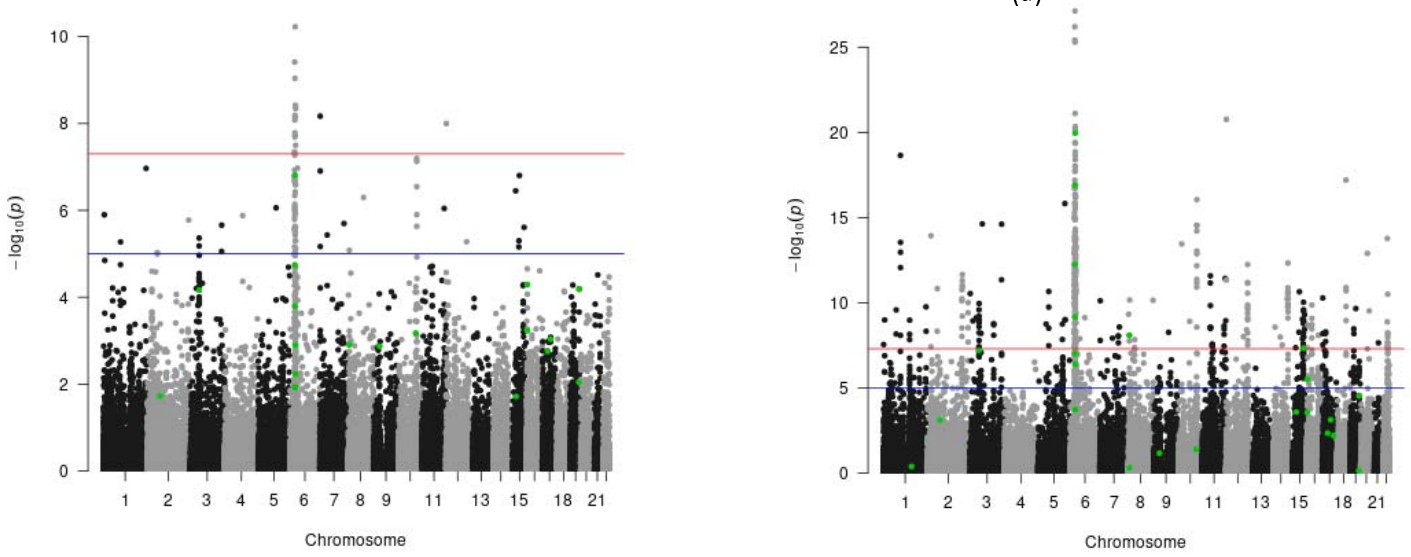
Figure 2. Manhattan plots of schizophrenia GWAS waves 1 (SCZ1; a and c) and 2(SCZ2; b and d) with the variants nominated by MTAG (a and b) using bipolar disorder GWAS wave 1 (BPD1) as the pleiotropic trait and the LSMM method using a global FDR (LSMM) (c and d) highlighted in green. P-values were derived from the publicly available downloads of SCZ1 and SCZ2 provided by the Psychiatric Genomics Consortium, respectively. These plots include the full downloadable GWAS summary statistics for both SCZ GWAS waves, without excluding significant GWAS1 regions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

EUGENE-SCZ1-nominated variants in SCZ1

EUGENE-SCZ1-nominated variants in SCZ2

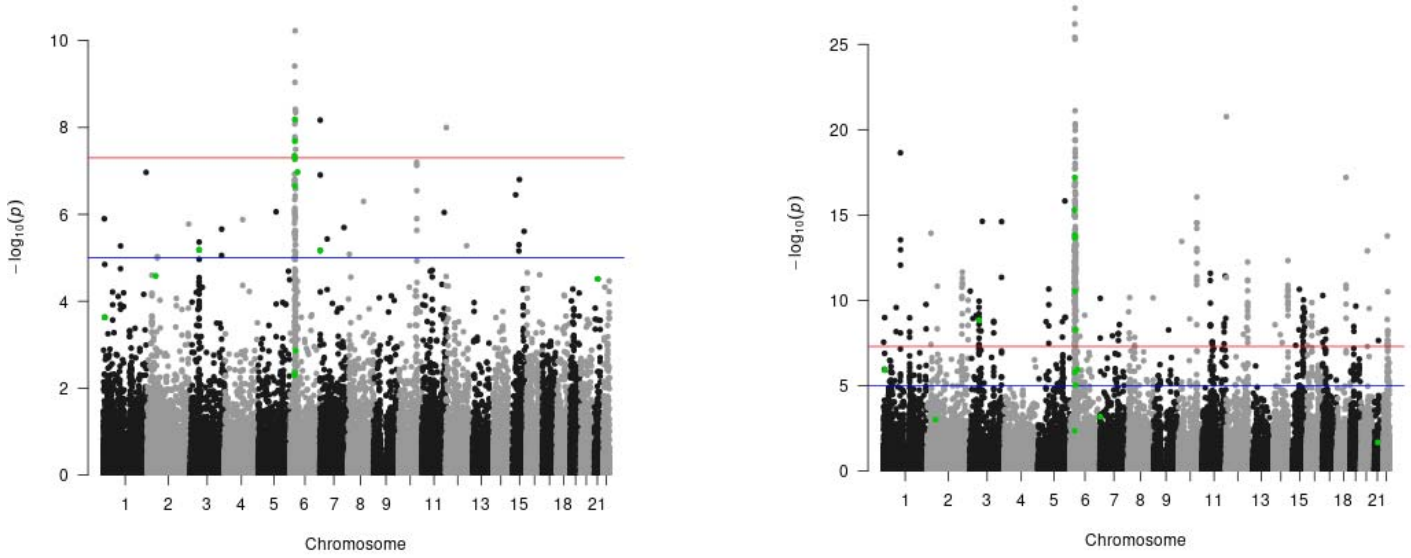
(a)



(b)

SMR-SCZ1-nominated variants in SCZ1

SMR-SCZ1-nominated variants in SCZ2



(c)

(d)

Figure 3. Manhattan plots of schizophrenia GWAS waves 1 (SCZ1; a and c) and 2 (SCZ2; b and d) with the variants nominated by the EUGENE (a and b) and SMR (c and d) methods using Brain eMETA cohort annotations (SMR2) highlighted in green. P-values were derived from applying MAGMA to SCZ1 and SCZ2, respectively. These plots include the full downloadable GWAS summary statistics for both SCZ GWAS waves, without excluding significant GWAS1 regions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

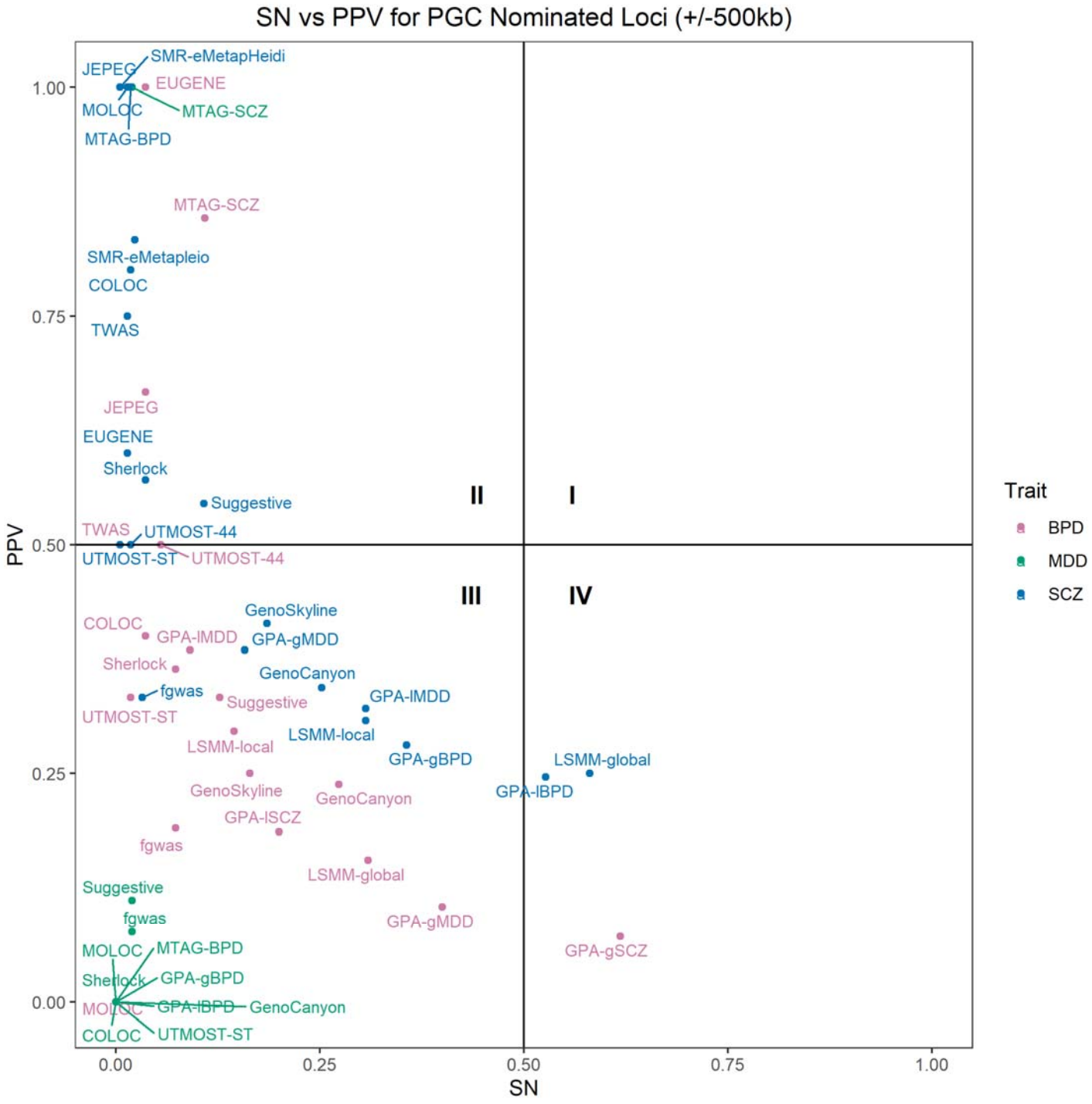


Figure 4. Scatterplot of the relationship between sensitivity (SN) and positive predictive value (PPV) for method-psychiatric trait combinations that return nominated variants. SN and PPV were calculated using +/- 500kb overlap criteria and compared to GWAS3 as the gold standard. Horizontal and vertical lines denote SN and PPV of 50%, respectively.

Table 1. Description of functional weighting methods

Method Name	Classification	Level	Citation	Significance Threshold
Suggestive	NA	Variant	NA	p < 1e-5
GenoCanyon10K	annotation	Variant	Lu, et al. Sci Rep 2015	Prediction Score > 0.5
GenoSkyline	annotation	Variant	Lu, et al. PLoS Genetics 2017	Prediction Score > 0.5
Naïve	annotation	Variant	Sveinbjornsson, et al. 2016	Annotation-based threshold
LSMM	annotation	Variant	Ming, et al. Bioinformatics 2018	FDR < 0.05
GPA	pleiotropy	Variant	Chung, et al. PLoS Genetics 2014	FDR < 0.05
MTAG	pleiotropy	Variant	Turley, et al. Nature Genetics 2018	p < 5e-8
fGWAS	eQTL	Variant	Pickrell. AJHG 2014	PPA > 0.9
Weighted eQTL	eQTL	Variant	Li, et al. Front Genet 2013	p < 5e-8
COLOC	eQTL	eQTL	Giambartolomei, et al. PLoS Genetics 2014	Approximate Bayes Factor > 0.75
MOLOC	eQTL	eQTL	Giambartolomei, et al. Bioinformatics 2018	Posterior Probability > 0.80
Jepeg	eQTL	eQTL	Lee, et al. Bioinformatics 2014	Bonferroni-adjusted Jepeg p-value
Sherlock	eQTL	eQTL	He, et al. AJHG 2013	Log Bayes Factor >= 4.0
SMR	eQTL	eQTL	Zhu, et al. Nature Genetics 2016	FDR q-value < 0.05; Heidi p-value < 0.05
TWAS/FUSION	eQTL	eQTL	Gusev, et al. 2016	Bonferroni-adjusted TWAS p-value
fastENLOC	eQTL	eQTL	Wen, et al. PLoS Genetics 2017	RCP >= 0.50
EUGENE	eQTL	eQTL	Ferreira, et al. JACI 2017	p-value corresponding to largest FDR < 0.05
UTMOST	eQTL	eQTL	Hu, et al. Nat Genet 2019	Bonferroni-adjusted UTMOST p-value

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Ranking of all methods by best performing PPV, as measured by locus (+/- 500kb)

	BPD		MDD		SCZ		MPV		WBC			
Method	Best PPV	Rank	Best PPV	Rank	Best PPV	Rank	Best PPV	Rank	Best PPV	Rank	Median Rank	Median Rank excluding MDD
Suggestive	0.286	6	0.000	9	0.455	7	0.204	11	0.349	9	9	8
Weighted eQTL	NA	18	NA	18	NA	18	0.333	3	0.667	2	18	3
fgwas	0.143	10	0.077	2	0.190	13	0.302	4	0.250	14	10	9
GenoSkyline	0.083	12	NA	18	0.273	11	0.220	9	0.250	14	12	10
GenoCanyon	0.079	13	0.000	9	0.196	12	0.200	12	0.207	16	12	12
GPA	0.231	7	0.000	9	0.275	10	0.256	6	0.335	10	9	8
MTAG	0.571	1	1.000	1	1.000	2	NA	18	1.000	1	1	2
Naïve	NA	18	NA	18	NA	18	NA	18	NA	18	18	18
LSMM	0.185	9	NA	18	0.181	14	0.214	10	0.225	15	14	12
Jepeg	0.333	5	NA	18	1.000	2	0.000	15	0.375	7	7	5
TWAS/FUSION	0.500	3	NA	18	0.500	6	0.400	1	0.368	8	6	4
EUGENE	0.500	3	NA	18	0.400	8	0.255	7	0.282	12	8	8
SMR (best performing)	NA	18	NA	18	0.000	15	0.143	14	0.400	6	15	14
COLOC	0.200	8	0.000	9	0.600	4	0.250	8	0.566	3	8	4
UTMOST- (best result)	0.333	5	0.000	9	0.500	6	0.320	5	0.300	11	6	6
Sherlock	0.182	11	0.000	9	0.286	9	0.379	2	0.538	4	9	3
fastENLOC	NA	18	NA	18	NA	18	NA	18	NA	18	18	18
moloc	0.000	14	0.000	9	0.667	3	0.182	13	0.500	5	9	4

Ties between methods resolved using the Olympic method.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.10.475153; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.