



OPEN ACCESS

EDITED BY
Katia Pane,
IRCCS SYNLAB SDN, Italy

REVIEWED BY
Pavel Kuksa,
University of Pennsylvania, United States
Suhua Chang,
Peking University Sixth Hospital, China

*CORRESPONDENCE
Laura I. Furlong,
laura.furlong@upf.edu

SPECIALTY SECTION
This article was submitted to Genetics of
Common and Rare Diseases,
a section of the journal
Frontiers in Genetics

RECEIVED 29 July 2022
ACCEPTED 20 September 2022
PUBLISHED 05 October 2022

CITATION
Pérez-Granado J, Piñero J and
Furlong LI (2022), Benchmarking post-
GWAS analysis tools in major
depression: Challenges
and implications.
Front. Genet. 13:1006903.
doi: 10.3389/fgene.2022.1006903

COPYRIGHT
© 2022 Pérez-Granado, Piñero and
Furlong. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Benchmarking post-GWAS analysis tools in major depression: Challenges and implications

Judith Pérez-Granado¹, Janet Piñero^{1,2} and Laura I. Furlong^{1,2*}

¹Research Programme on Biomedical Informatics (GRIB), Hospital Del Mar Medical Research Institute (IMIM), Department of Medicine and Life Sciences (MELIS), Universitat Pompeu Fabra (UPF), Barcelona, Spain, ²MedBioinformatics Solutions SL, Barcelona, Spain

Our knowledge of complex disorders has increased in the last years thanks to the identification of genetic variants (GVs) significantly associated with disease phenotypes by genome-wide association studies (GWAS). However, we do not understand yet how these GV functionally impact disease pathogenesis or their underlying biological mechanisms. Among the multiple post-GWAS methods available, fine-mapping and colocalization approaches are commonly used to identify causal GV, meaning those with a biological effect on the trait, and their functional effects. Despite the variety of post-GWAS tools available, there is no guideline for method eligibility or validity, even though these methods work under different assumptions when accounting for linkage disequilibrium and integrating molecular annotation data. Moreover, there is no benchmarking of the available tools. In this context, we have applied two different fine-mapping and colocalization methods to the same GWAS on major depression (MD) and expression quantitative trait loci (eQTL) datasets. Our goal is to perform a systematic comparison of the results obtained by the different tools. To that end, we have evaluated their results at different levels: fine-mapped and colocalizing GV, their target genes and tissue specificity according to gene expression information, as well as the biological processes in which they are involved. Our findings highlight the importance of fine-mapping as a key step for subsequent analysis. Notably, the colocalizing variants, altered genes and targeted tissues differed between methods, even regarding their biological implications. This contribution illustrates an important issue in post-GWAS analysis with relevant consequences on the use of GWAS results for elucidation of disease pathobiology, drug target prioritization and biomarker discovery.

Abbreviations: CLPP, Colocalization Posterior Probability; eGenes, genes regulated by eQTLs; eQTL, expression quantitative trait loci; fastENLOC, Fast Enrichment Estimation Aided Colocalization Analysis; GTEx, Genotype-Tissue Expression; GV, genetic variant; GWAS, Genome-Wide Association Studies; LD, linkage disequilibrium; MD, major depression; pGenes, proximal genes; PICS, Probabilistic Identification of Causal SNPs; SNPs, single nucleotide polymorphisms; SSimp, Summary Statistics Imputation software.

KEYWORDS

fine-mapping, colocalization, post-GWAS, major depression, eQTLs

Introduction

More than 207,400 genetic variants (GVs) have been associated with complex diseases since the introduction of genome-wide association studies (GWAS) (Dehghan, 2018; Buniello et al., 2019). The vast majority of identified GV lie in non-coding regions of the genome with no clear impact on gene function and disease pathogenesis (Brandes et al., 2022), posing challenges in interpreting the association of the GV with the disease phenotype. Furthermore, these GV may not be the causal ones but may be in linkage disequilibrium (LD) with the true causal GV (Visscher et al., 2017; Brandes et al., 2022). We refer to causal GV to those with a biological impact. A variety of approaches are available to unravel the functional role of GV identified by GWAS (Kichaev et al., 2014; Amlie-Wolf et al., 2018; Wallace, 2021; Gazal et al., 2022). In addition, there are a plethora of different tools available that serve the same purpose but work with different types of data (e.g., genotype data *versus* full genome summary statistics), under different assumptions (e.g., one causal GV or more), and with diverse outcomes (e.g., causal GV or relevant gene-cell type combination) (Cano-Gamez and Trynka, 2020; Adebisi et al., 2021). There is, however, no guideline for determining which tool is best to use for each approach nor a gold standard for evaluating the validity of the results. Furthermore, in contrast to other areas where benchmarking evaluations of methods are in place, such as for protein structure prediction (Protein Structure Prediction Center, 2020) or disease module identification (Dream Challenges, 2022), among others, methods for GWAS data analysis have not been objectively benchmarked. Selecting the right tool is critical in post-GWAS analysis, to properly unravel the functional mechanisms by which the GV lead to disease, and where different performances can lead to different results (Wen et al., 2017; Rieger et al., 2018; LaPierre et al., 2021).

There is an absence of a benchmark dataset to assess the performance of post-GWAS analysis tools. Therefore, we propose a systematic and objective comparison of the results obtained by different tools when applied to the same datasets. We designed a fine-mapping and colocalization workflow with different tools running alternatively. Fine-mapping analysis identifies the causal GV and is a necessary step in most post-GWAS analyses. We used the tools Probabilistic Identification of Causal SNPs (PICS) (Taylor et al., 2021) and TORUS (Wen, 2016) as alternative tools for fine-mapping. Colocalization methods pinpoint the GV causally associated with a phenotype and a molecular trait of interest, such as expression or methylation. We focused our analysis on expression quantitative trait loci (eQTL), to identify GV with an effect on the expression of genes, from now on referred to as eGenes. We applied two methods for colocalization analysis: the

Colocalization Posterior Probability (CLPP) approach (Hormozdiari et al., 2016) and the Fast Enrichment Estimation Aided Colocalization analysis (fastENLOC) (Pividori et al., 2020; Hukku et al., 2021). We applied the fine-mapping and colocalization workflow to the same GWAS on major depression (MD) and eQTL datasets (Figure 1). The results obtained with each combination of tools were evaluated in terms of fine-mapped and colocalizing GV, the retrieved eGenes, the tissues in which this regulation of gene expression might take place, as well as the biological processes in which these genes are involved.

The results of the workflow reveal divergence across tools, pinpointing a relevant issue in post-GWAS analysis derived from the lack of method benchmarking. Our findings demonstrate how critical is the fine-mapping step to subsequent analysis and how colocalization outcomes are in turn highly impacted by the assumptions of each tool. As a consequence, the causal GV and eGenes identified are different and are involved in different biological processes. Overall, given the lack of agreement among tools, we highlight the need for an objective and unbiased assessment of post-GWAS analysis methods and tools to properly leverage GWAS data.

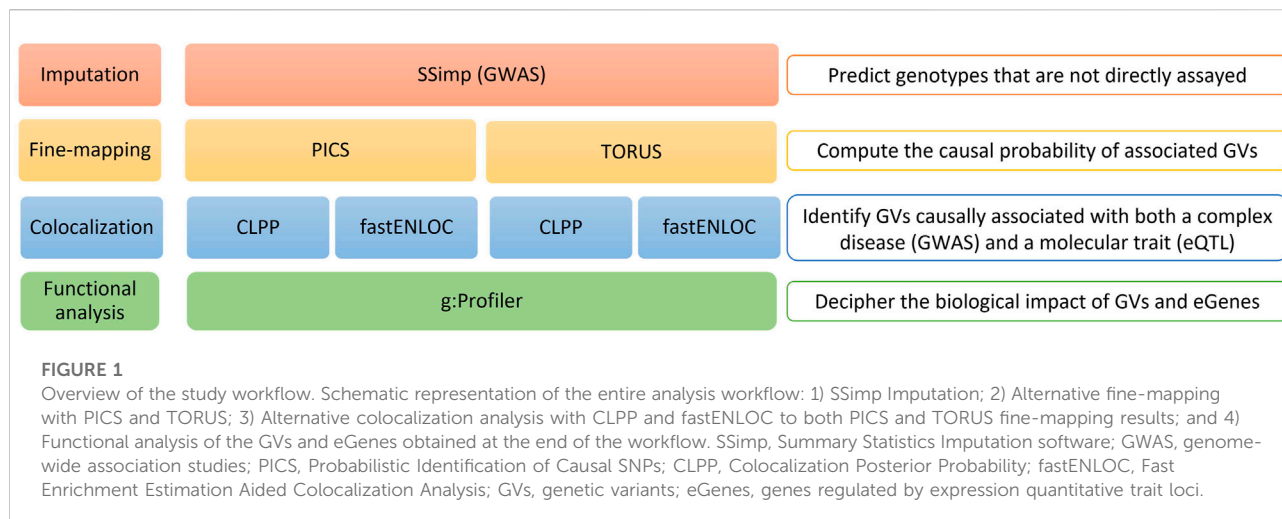
Materials and methods

Among the plethora of available methods for post-GWAS analysis and which have been reviewed elsewhere (Cano-Gamez and Trynka, 2020; Adebisi et al., 2021), we focused on fine-mapping and colocalization. Then, we conducted a tool selection based on: workability with full-genome summary statistics, documentation quality, software maturity and developer support availability.

The workflow we describe in this manuscript applies alternative tools for post-GWAS analysis to compare their outcomes (Figure 1). We begin with an imputation step, followed by a fine-mapping and colocalization analysis using two different tools for each of these processes, and finish with a functional analysis of the results obtained using different tools and databases. We present below a more detailed explanation of each step.

GWAS dataset and imputation

We have selected the latest genome-wide association study (GWAS) on major depression (MD) with publicly available full-genome summary statistics (GCST005902) (Howard et al., 2018). This GWAS evaluated 7,666,894 genetic variants (GV) in 322,580 European participants (113,769 cases and



208,811 controls). We used the harmonized version of this GWAS dataset. This implies the genomic position is reported against the latest genome build (GRCh38) and the orientation is checked by flipping the effect allele (i.e., the allele that confers the risk, which is not always the minor allele) and other alleles whenever appropriate. The beta and 95% confidence interval is also inverted accordingly [downloaded from: http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST005001-GCST006000/GCST005902/harmonised/29662059-GCST005902-EFO_0003761.h.tsv.gz].

The Genotype-Tissue Expression (GTEx) expression quantitative trait loci (eQTL) dataset contains single-tissue cis-eQTL data with eGene, meaning genes regulated by eQTL, and significant variant-gene associations for 49 tissues [downloaded from: <https://www.gtexportal.org/home/datasets>] (Genotype-Tissue Expression, 2017).

To match the GTEx eQTL panel, we imputed not genotyped GVs in the MD dataset with the Summary Statistics Imputation software (SSimp) (Rüeger et al., 2018). The parameters we used were GWAS full-genome summary statistics GVs with their matching z-scores, reference and effect alleles along with the European 1,000 genomes linkage disequilibrium (LD) reference panel [downloaded from: <http://hgdownload.cse.ucsc.edu/gbdb/hg19/1000Genomes/phase3/>]. We computed the z-scores by dividing GVs' effect size, understood as the effect of the risk allele relative to the reference allele, over the standard error (Shi, 2017). We then assessed the imputation quality returned by the SSimp software using the r^2 . pred parameter, which ranges between 0 -bad quality- and 1 -good quality-. Note that we only considered single nucleotide polymorphisms (SNPs) for this analysis.

Fine-mapping with PICS and TORUS

Before applying the Probabilistic Identification of Causal SNPs (PICS) and TORUS fine-mapping tools, we matched

GWAS GVs and eQTLs to their corresponding LD blocks using the European 1,000 Genomes LD reference panel (Berisa and Pickrell, 2016).

We run PICS by programmatically accessing its web application form. We used LD-based PICS (<https://pics2.ucsf.edu/pics2-LD.html>), which performs LD expansion and fine-mapping. In brief, PICS takes the most significant GV per association locus along with its associated p-value, performs LD expansion and then computes the probabilities by performing empirical permutations per GV. For GWAS, we submitted the data and obtained the computed PICS probabilities for the input GVs and those in LD, from now on linked GVs. As for eQTLs, we downloaded precomputed LD-based PICS for all GTEx best eQTLs per gene per tissue type [downloaded from: <https://pics2.ucsf.edu/Downloads/GTEx/>].

We executed TORUS software package using the parameters “-load_zval -dump_pip”. TORUS accepts full-genome summary statistics data, meaning all GVs analysed in the study, and their associated z-scores. Then it computes the causal probabilities using an expectation-maximization algorithm which assumes there is only one causal GV per locus. We obtained these probabilities for all GWAS GVs and GTEx eQTLs (v8) [downloaded from: https://storage.googleapis.com/gtex_analysis_v8/single_tissue_qtl_data/GTEx_Analysis_v8_eQTL.tar], per tissue.

The major histocompatibility complex region (chr6: 28,510,120–33,480,577, GRCh38) was excluded from all datasets for the analysis due to the complex LD structure of GVs, which may lead to inaccurate results (Ghoussaini et al., 2020).

Colocalization analysis via CLPP and fastENLOC

For the colocalization analysis, we implemented two different approaches in the workflow: the Colocalization Posterior

Probability (CLPP) approach and the Fast Enrichment Estimation Aided Colocalization Analysis (fastENLOC). We applied these two methods to the fine-mapping results obtained with both PICS and TORUS to identify the genes regulated by causal GVs, also known as eGenes. Both tools consider that there can be more than one causal GV per association locus. CLPP assumes independence between GWAS and eQTL data while fastENLOC does not and computes the enrichment of GWAS on eQTL data using an embedded function. In addition, fastENLOC not only computes SNP colocalization probabilities (SCP) but also regional colocalization probabilities (RCP) to overcome the inability to narrow down to a single causal SNP, common to all tools. Please note that the tools were run following the guidelines and parameters recommended by the authors. We conducted a CLPP approach by computing the product of PICS probabilities for GWAS and eQTL overlapping linked GVs. Based on previous experience in post-GWAS data analysis, we narrowed down the results to the most likely causal GVs (Farh et al., 2015; Ghousaini et al., 2020; Pérez-Granado et al., 2022) by filtering GWAS GVs and eQTLs PICS probabilities, as well as their product by >10%. We run fastENLOC with fine-mapped GWAS GVs and eQTLs per tissue using the following parameters: default shrinkage 1) and total variants (7,666,894). We filtered the results by RCP >0.5 and SCP >0.001 (Wen et al., 2017).

Proximal genes

Common gene mapping practices involve looking at the GVs' overlapping or nearest downstream and upstream genes, also known as proximal genes or pGenes. We retrieved this genetic information using Ensembl *via* SNPnexus (Oscanoa et al., 2020).

We first identified pGenes associated with GVs from PICS and TORUS fine-mapping results and performed gene-set enrichment analysis on both sets.

Then, for each fine-mapping and colocalization combination of tools, we obtained the pGenes to which the GVs mapped and compared them to the corresponding set of eGenes. We also evaluated each pGenes-eGenes set for their association with disease and performed a gene-set enrichment analysis.

Functional analysis

For the evaluation of association to disease, we followed two different approaches. When evaluating GVs from fine-mapping results, we used variant association data from DISGENET plus (Piñero et al., 2019; DISGENET plus, 2022). Note that the GWAS under evaluation (Howard et al., 2018) and a meta-analysis that it is a part of (Howard et al., 2019) were removed from DISGENET plus datasets to avoid circularity. As for genes, we used the R

package *disgenetplus2r* (disgenetplus2r, 2022), which contains gene-disease association data, and considered Medical Subject Headings (MeSH) disease classes system for disease grouping.

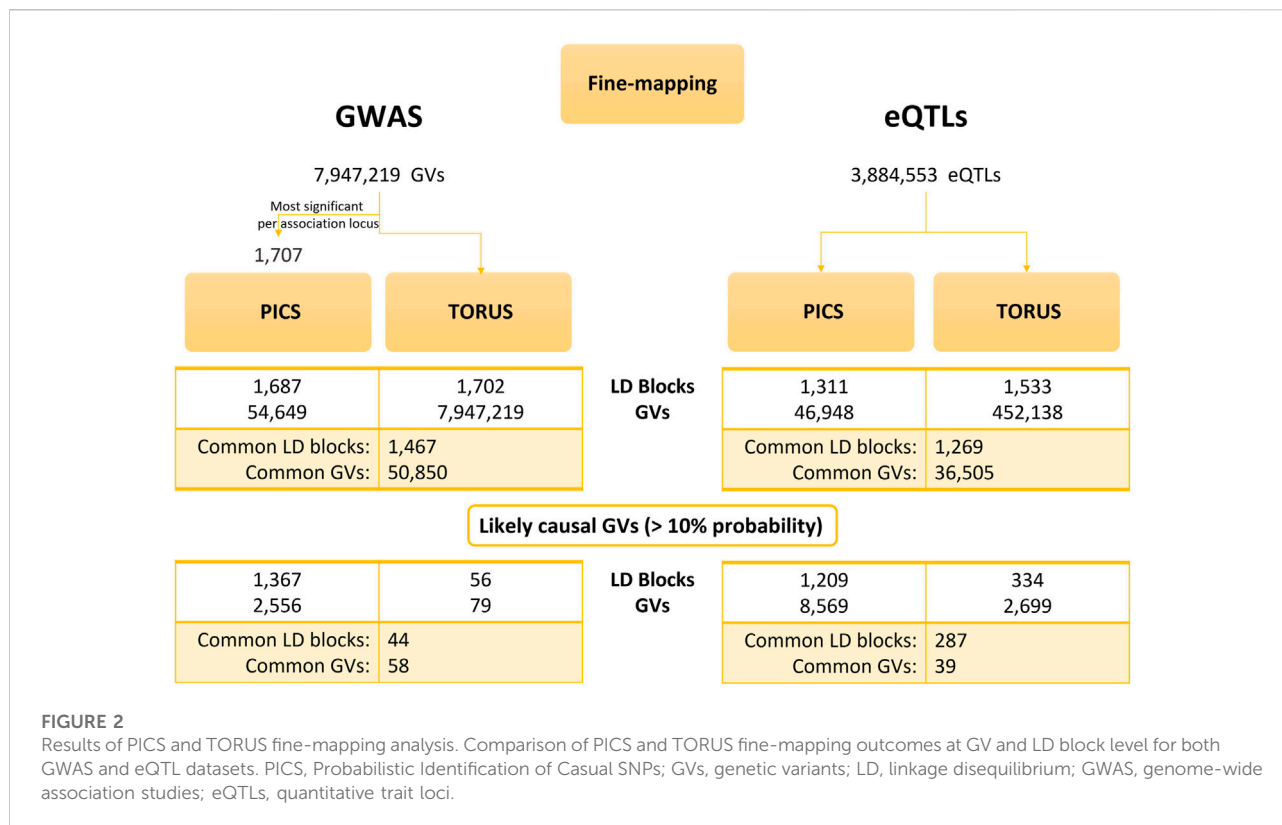
We performed the gene-set enrichment analysis using *g:Profiler* *via* the R package *gprofiler2* (Raudvere et al., 2019) and the following databases: 1) Gene Ontology (GO) biological processes, molecular functions and cellular processes; 2) Reactome and WikiPathways pathways; 3) miRNA annotations; 4) Human Phenotype Ontology, which focuses on rare Mendelian disorders, and has phenotypic features associated with disease; and 5) DISGENET plus, which has genes' association data to disease and phenotypic traits (v19). The whole set of known human genes was used as domain scope for the analysis and electronic GO annotations were not considered. Furthermore, to make the functional enrichment analysis more meaningful, we filtered the terms by their specificity using their term size (<1,500 genes), which corresponds to the number of genes associated with that term.

In addition, we applied a guilt-by-association approach to overcome the lack of functional information for some genes and assign the function of better-characterized neighbours in the interactome. Thus, we used molecular interaction data from IntAct (Orchard et al., 2014) clustered with MONET (Tomasoni et al., 2020) to evaluate whether different eGenes retrieved from the workflow could belong to the same cluster and thus affect the same molecular pathway. We performed a gene-set enrichment analysis of the retrieved clusters filtering by an eGene-cluster genes ratio of 1:50.

We evaluated the fine-mapping and colocalization results at different levels: the tissue specificity, colocalizing causal GVs, their target genes (eGenes) and their biological implications. We examined the results individually and then compared them across tools, with classic approaches (pGenes) and with the results reported in the original publication.

Results

This study evaluates and compares the outcomes of different fine-mapping and colocalization tools (Figure 1). To accomplish this, we have run our analysis using the same genome-wide association study (GWAS) on major depression (MD) and expression quantitative trait loci (eQTL) datasets. In addition, and in line with our goal, we address the results of each analytical step individually before getting into their biological implications. The workflow begins with an imputation phase (SSimp) to predict the genotypes not directly assayed in the original GWAS. Then, a fine-mapping step with Probabilistic Identification of Causal SNPs (PICS) and TORUS to identify the most likely causal genetic variants (GVs), meaning those likely to have a biological effect on the trait, and compute their causal probabilities. Next, a colocalization analysis using the Colocalization Posterior Probability (CLPP) approach and the



fast enrichment estimation aided colocalization (fastENLOC) software, to identify the GVs causally associated with both MD and a change in expression of a target gene. Finally, the functional analysis, leveraging a diversity of databases, aims to decipher the impact of the identified GVs and eGenes, meaning genes regulated by eQTLs.

GWAS dataset imputation

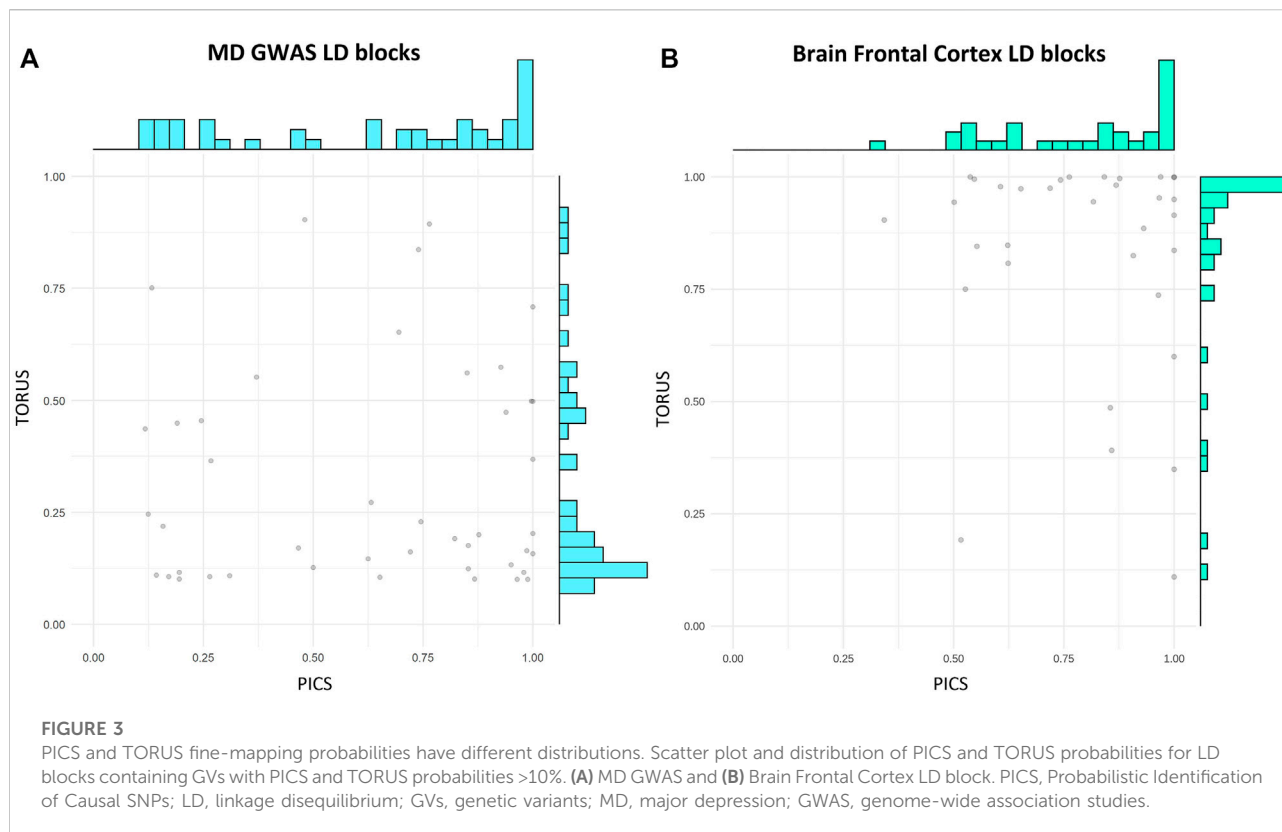
The original genome-wide association study (GWAS) consisted of 7,624,931 harmonised genetic variants (GVs) and after imputation to predict missing Genotype-Tissue expression (GTEx) eQTLs, we obtained 7,947,219 GVs (ie. a total of 554,824 imputed GVs). The estimated imputation quality provided by SSimp was generally good for all chromosomes (r^2 . pred >0.8) except for chromosome 17.

Fine-mapping with PICS and TORUS

We run linkage disequilibrium (LD)-based PICS by inputting the most significant GWAS GVs per LD block (1,707 GVs) along with their p-values (Figure 2). After PICS LD expansion and fine-mapping, we obtained 54,649 GVs with their corresponding PICS probabilities. As for the GTEx eQTLs, we downloaded

the precomputed LD-based PICS per tissue from the data portal. In parallel, we computed the z-scores for all GWAS GVs and GTEx eQTLs and along with the LD block specification, we used them as input for TORUS.

We compared PICS and TORUS initial fine-mapping results (Supplementary Figure S1) and then filtered GVs by a probability >10% to keep the most likely causal GVs. Because each tool has its own assumptions and different GVs could be identified, but these may be in LD, the comparison was done considering the probabilities per LD block. In addition, we examined the distribution of PICS and TORUS sum of probabilities for all LD blocks with likely causal GVs (GWAS: 1,367 and 56, respectively; GTEx: 1,209 and 334, respectively) (Supplementary Figure S1A) as well as the common ones (44 and 287, respectively) (Figure 3A). PICS probabilities for GWAS GVs are biased towards higher values in all cases, with 74% of GWAS LD blocks having a probability greater than 50%. Meanwhile, TORUS probability distribution is skewed towards lower values with only 21% of LD blocks surpassing the 50% probability. Regarding GTEx eQTLs, PICS and TORUS results generally follow a more similar distribution with probabilities biased towards higher values, especially when only GVs with probabilities greater than 10% are considered (Figure 3B and Supplementary Figure S1B; note that we focused on Brain



Frontal Cortex region because it is relevant to MD and for illustrative purposes).

The analysis of PICS and TORUS most likely causal GVs (probability >10%) revealed that both sets are enriched in GVs associated with MD, bipolar disorder and other psychiatric disorders (Supplementary Tables S1, S2). PICS causal GVs are also enriched in metabolic-related traits such as triglycerides measurement.

Additionally, we applied classic gene-mapping approaches to PICS and TORUS fine-mapping results, yielding 1,277 and 1,248 proximal genes or pGenes, respectively. Both sets were enriched in genes associated with neurogenesis as well as neuron differentiation and development (Supplementary Tables S3, S4).

Colocalization analysis via CLPP and fastENLOC

The colocalization results from CLPP approach using PICS fine-mapping results yielded 44 GVs and 43 genes regulated by eQTLs, also known as eGenes, affecting 28 tissues (Supplementary Table S5), whereas no results were obtained when using TORUS causal GVs. In parallel, fastENLOC applied to causal GVs identified by PICS resulted in 24 GVs and 17 eGenes across 13 tissues (Supplementary Table S6), while

when applied on TORUS probabilities yielded 10 GVs and 3 eGenes in 2 tissues (Supplementary Table S7).

When comparing methods, the use of different colocalization tools after fine-mapping with PICS yields the most similar results. When using PICS, all tissues and eGenes identified by fastENLOC are also obtained by CLPP, with differences found at the GV level, and CLPP retrieving additional eGenes compared to fastENLOC (Supplementary Table S8 and Figure 4). Meanwhile, when comparing the use of PICS or TORUS fine-mapping probabilities followed by fastENLOC, we only identified one common tissue but with different eGenes and GVs. Similarly, PICS+CLPP and TORUS+fastENLOC yielded common findings only at the tissue level. Among the tissues with causal GVs and eGenes retrieved when using PICS and either colocalization tools, we can find diverse brain regions like the frontal cortex or hypothalamus.

Proximal genes and functional analysis

We compared eGenes from fine-mapping and colocalization workflow to pGenes from PICS and TORUS fine-mapping results. Only 3 genes overlapped between pGenes from PICS fine-mapping and PICS+CLPP eGenes (KTN1, PXMP4 and ESYT2) and one with PICS+fastENLOC

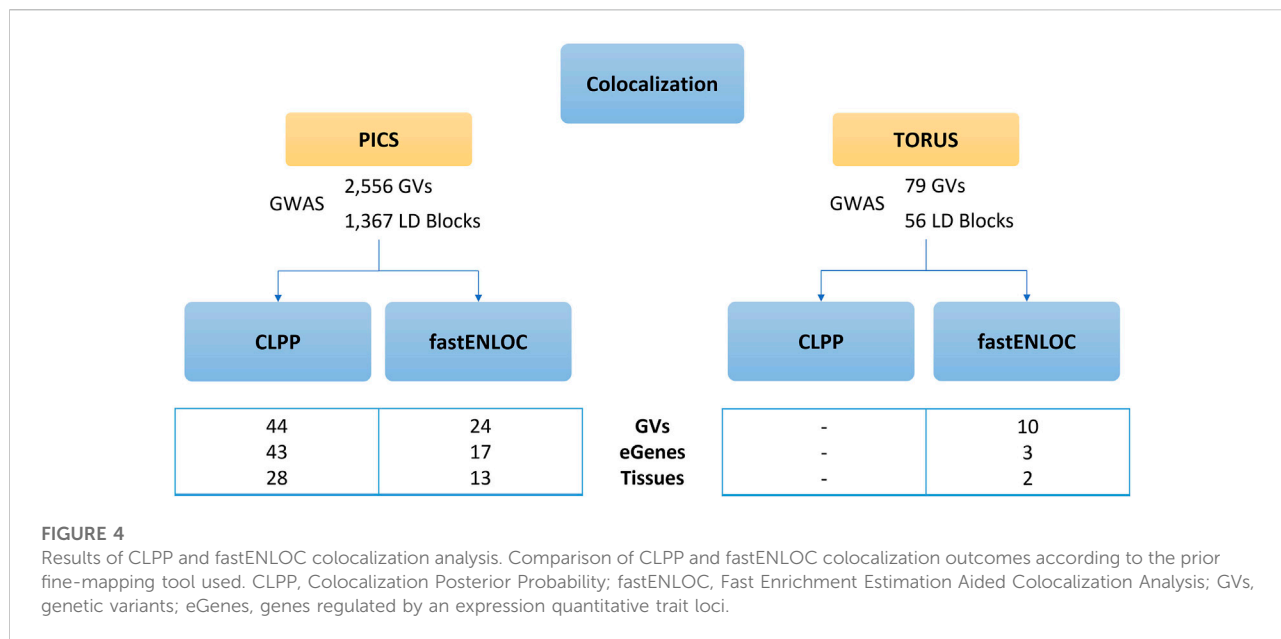


TABLE 1 Identified eGenes differ from classic gene mapping (pGenes).

	eGenes	pGenes	Matches
PICS + CLPP	43	64	10
PICS + fastENLOC	17	31	3
TORUS + fastENLOC	3	8	0

eGenes, genes regulated by expression quantitative trait loci; pGenes, proximal genes; PICS, Probabilistic Identification of Causal SNPs; CLPP, Colocalization Posterior Probability; fastENLOC, fast enrichment estimation aided colocalization analysis. The number of eGenes, also known as genes regulated by eQTLs, retrieved from the fine-mapping and colocalization analysis; the number of pGenes or proximal genes, that is overlapping or nearest upstream and downstream genes; and matches between eGenes and pGenes. The information is shown for each combination of fine-mapping and colocalization tools used.

(KTN1). There was no overlap with pGenes when comparing to TORUS. The eGenes from PICS+CLPP are enriched in their association with miRNAs and the eGenes from PICS+fastENLOC in RNA Polymerase I Promoter Escape (Supplementary Table S9). Considering all eGenes together (46), these are functionally enriched in terms related to transcription factor regulation and miRNA. We also assessed the distribution of the eGenes in a clustered human interactome. The three sets of eGenes (i.e., PICS+CLPP, PICS+fastENLOC and TORUS+fastENLOC) belonged to different clusters, except for eGenes shared across tools results (ie. 17 shared eGenes which are located in 10 clusters). Some of these clusters were associated with transcription factor regulation, inflammation or neurogenesis (Supplementary Tables S10, S11). No clusters identified for TORUS+fastENLOC passed the functional analysis filters, that

is a ratio of eGenes over cluster genes higher than 1:50 and enriched term size <1,500 genes.

When we applied traditional gene mapping approaches to the GV's that were found to regulate the expression of those eGenes, we discovered a total of 74 pGenes. The vast majority of eGenes identified do not match pGenes, which holds true across all workflows (Table 1). In addition, most matches derive from GV's lying in an intronic region of the genome (Supplementary Table S12). Nonetheless, all sets of pGenes and eGenes are associated with mental disorders, behaviour and behaviour mechanisms as well as psychological phenomena and processes and nervous system disease (Supplementary Figures S2, S3). Additionally, pGenes are enriched in GO terms associated with diverse signalling pathways (Supplementary Table S13).

Furthermore, we compared the results obtained with the original publication where 14 GV's and 7 pGenes were reported. The latter are functionally associated with synapsis (Supplementary Table S13) and 6 of them have a prior association with mental disorders (Supplementary Figure S4). Only 5 fine-mapped GV's from PICS and 7 GV's from TORUS overlapped with the GV's reported in the original publication, and 1 pGene (SGIP1), which is in both sets of fine-mapped pGenes. However, none of the GV's and pGenes obtained by colocalization with any combination of tools evaluated in our pipeline overlapped with the GV's and pGenes reported in the original publication.

Discussion

Currently, there are a plethora of strategies available for post-GWAS analysis (Cano-Gamez and Trynka, 2020; Adebiyi et al.,

2021). Here, we have focused on two main approaches: fine-mapping, which aims to identify the likely causal GVs, and colocalization, aimed at identifying which genes are regulated by the GVs at the expression level (eGenes). Furthermore, while many tools address the same goal, there is no standard set of causal GVs that have been experimentally validated for benchmarking to determine and compare which one is the most adequate (Brandes et al., 2022). Thus, we have designed an evaluation exercise to assess the outcome of different fine-mapping and colocalization tools using the same MD GWAS and eQTL dataset. To the best of our knowledge, no study goes beyond the comparison of the different tool's assumptions and thus the evaluation of the biological implications of their findings (Wen et al., 2017; Cano-Gamez and Trynka, 2020).

Our main premise throughout this analysis has been to use each tool as it was intended by following developers' recommendations and guidelines as closely as possible. This way we could get the most out of them and compare their optimised outcomes. Furthermore, one of the primary reasons behind the tools' selection was their ability to work with full-genome summary statistics instead of individual genotype data, which can be difficult to obtain due to privacy concerns. Other criteria for tool selection included the quality of documentation, the maturity of the software, and the availability of developer support.

Prior to post-GWAS analysis, the imputation process using SSimp yielded very good quality results except for chromosome 17. One possible explanation is that SSimp provides hg19 1000Genomes phase3 as the reference panel. This version of the genome has some gaps, most of which are found in telomeres and centromeres, having a strong impact on chromosome 17 (Rashid-Kolvear et al., 2007; Genome Reference Consortium, 2022). We then proceeded with the fine-mapping and colocalization workflow, keeping the previously mentioned issue in mind when evaluating their results.

Fine-mapping is significantly influenced by LD patterns and the used tools, PICS and TORUS, which work under different assumptions (see Methods). Therefore, to have comparable results we considered the probabilities obtained at the LD block level, because the most likely causal GVs may differ or not be discernible due to high LD between GVs. In addition, to account for the difference in the number of GVs which could be driving the observed inverse distribution of probabilities between tools (Supplementary Figure S1), only the most likely causal GVs were considered in the comparison (Figure 2). In general, TORUS retrieves GVs with lower probabilities compared to PICS. This could be explained by the algorithm's conservative nature and its assumption of one causal GV per association locus, with probabilities biased towards zero when the locus contains multiple causal GVs (Wen, 2016). Indeed, the one causal GV assumption has been debated, with multiple GVs acting together resulting in a more reasonable theory (Burgess, 2022). Nonetheless, both PICS and TORUS most likely causal GVs

are enriched in their association with MD, bipolar disorder and other psychiatric disorders (Supplementary Tables S1, S2). This suggests that both fine-mapping approaches identify likely causal GVs associated with MD. GVs fine-mapped by PICS are also enriched in diseases and traits usually comorbid with MD such as alcohol consumption (Gémes et al., 2019) and metabolic traits like serum total cholesterol measurement (Gold et al., 2020). Classic gene mapping of PICS and TORUS fine-mapping results (2,556 GVs and 79 GVs respectively, common- 58 GVs) (Figure 2), yielded 1,277 and 1,248 pGenes, respectively, with all TORUS pGenes included in PICS. These results could be explained because, compared to TORUS, PICS computes higher probability values and may retrieve more than one likely causal GV per locus. But provided the set probability threshold, some of these GVs may be in LD and therefore mapping to the same genes. Both sets of pGenes are enriched in genes associated with neurogenesis, highly affected in MD (Li Z. et al., 2021). All in all, fine-mapping is a critical step in post-GWAS analysis, with high divergence observed between different methods, particularly at the level of GVs and their associated probabilities, which will highly impact subsequent colocalization analysis.

CLPP and fastENLOC colocalization approaches were applied to both fine-mapping results from PICS and TORUS. Following the same logic, given that TORUS computed lower probability values, PICS yielded more colocalization findings (Supplementary Tables S5–S7). Furthermore, we have similar results under CLPP assumption of independence between GWAS and eQTLs compared to fastENLOC built-in function to compute their enrichment, with fastENLOC being more stringent as previously described (Hukku et al., 2021). Interestingly, when focusing on a single tissue, the results do not match at the GV level but do so at the eGene level (Supplementary Table S8). This suggests that there might be different GVs that have an effect on the expression of the same eGenes. It also highlights the importance of the identification of eGenes to determine how GVs may ultimately impact the disease phenotype.

The overlap between eGenes and pGenes from PICS and TORUS fine-mapping was very small, with 3 genes in total. Among them, KTN1 has also been associated with MD (Dall'Aglio et al., 2021) and ESYT2 is involved in neurodevelopmental pathways and may be associated with suicidal behaviour trends in MDD although more research is needed (Calabrò et al., 2018). eGenes from PICS+CLPP were functionally enriched with miRNAs. These have been recently reported as relevant in MD pathogenesis and treatment (Dwivedi, 2014). Specifically, hsa-miR-23a-3p has repeatedly been associated with duloxetine treatment response assessment in MD (Kim et al., 2019). Moreover, the eGene GMPPB identified from TORUS+fastENLOC has already been associated with MD pathogenesis in proteome-wide association studies (Wingo et al., 2021). GMPPB is involved in glycosylation, which has been reported as relevant and even hypothesized as a

potential biomarker for MD (Yamagata and Nakagawa, 2020). Considering all eGenes together, they are enriched in their association with transcription factor regulation (Supplementary Table S9), which has already been related to MD (Zhong et al., 2019; Li X. et al., 2021; Pérez-Granado et al., 2022). The mapping of eGenes to protein interaction clusters indicated that the three sets of genes (i.e., PICS+CLPP, PICS+fasENLOC and TORUS+fastENLOC) belonged to distinct clusters and are thus likely to be involved in different biological processes. Nevertheless, PICS+CLPP and PICS+fastENLOC associated sets of clusters were enriched with genes associated with processes involving TF regulation as well as inflammation or neurogenesis (Supplementary Tables S10, S11). All these processes are associated with MD pathogenesis (Shadrina et al., 2018; Zhong et al., 2019; Li X. et al., 2021, Li et al., 2021 Z.; Pérez-Granado et al., 2022). In general, the identified eGenes are poorly characterized yet the cluster analysis shades some light on their potential molecular associations.

Fine-mapping and colocalization analysis successfully identified eGenes associated with mental disorders (Supplementary Figures S2–S4) that differed from the set of pGenes, particularly when focusing on non-coding regions of the genome (Table 1 and Supplementary Table S12). Accordingly, pGenes are enriched in their association with pathways that have been reported as disrupted in MD such as MAPK (Wang et al., 2020), ErbB (Ledonne and Mercuri, 2020), PI3K/AKT (Matsuda et al., 2019) and ERK (Wang and Mao, 2019) signalling pathways (Supplementary Table S13); as well as MD potential causes like stress or inflammation (Shadrina et al., 2018; Li Z. et al., 2021). When comparing the results from our workflow to the original manuscript, there were only matches when considering the fine-mapped PICS and TORUS results but not after colocalization analysis. The common pGene between the three datasets was SGIP1, which has been involved in mood regulation (Dvorakova et al., 2021).

Brain regions are of particular interest in MD and as such, we focused the evaluation of our results on them. The brain frontal cortex, hypothalamus, pituitary and brain cerebellar hemisphere have common findings between PICS and both colocalization tools. MD and myoclonus-dystonia are usually comorbid, and their association has typically been studied in relation to SGCE mutation and its potential pleiotropic effect (Peall et al., 2013; Kim et al., 2017; Cazurro-Gutiérrez et al., 2021). However, whether SGCE plays a role in MD manifestation has been debated. On the one hand, animal studies have shown that knocking out this gene causes myoclonus, motor coordination deficits, and depression-like behaviour (Cazurro-Gutiérrez et al., 2021) which is consistent with the lower expression levels reported by GTEx. On the other hand, a similar frequency of MD has been reported in SCGE mutated and wild-type myoclonus dystonia patients (Kim et al., 2017). Focusing on the hypothalamus, one of the most common causes of MD is

stress, which affects the hypothalamic-pituitary-adrenal axis by increasing glucocorticoid levels (Karger et al., 2018; Oliva et al., 2018). These have an impact on various signalling pathways, including the Wnt pathway, in which FZD5 plays a role, and neurogenesis (Karger et al., 2018). However, the changes in gene expression caused by rs77678807 reported by GTEx are the inverse of what we would expect (Genotype-Tissue Expression, 2017). PCOLCE2 is highly expressed in the pituitary and there is evidence of reduced levels in depression-like behaviours in mice (Yamawaki et al., 2018), consistent with rs9757063 effect. Indeed, it has already been associated with psychiatric disorders by GWAS studies (Martínez-Magaña et al., 2021). However, how exactly they play a role in MD pathogenesis is still unknown. Little is known about the eGenes and GVs identified in the brain cerebellar hemisphere. Additionally, in the brain frontal cortex and hypothalamus, two different lncRNAs have been identified, LINC01159 and RP11-838N2.5 respectively. Even though little is known about them, lncRNAs seem to play a relevant role in MD pathogenesis and therapeutics (Shi et al., 2021; Hao et al., 2022). PICS+CLPP identified rs1480432 as upregulating the expression of DTNA, which is associated with neurogenesis and underregulating the maturation and stability of postsynaptic density (Chen et al., 2022). MAO B has been found to be overexpressed in postmortem brain tissue from MD patients, while DTNA is found to be underexpressed in MAO B knockout mice. The colon is another tissue whose associations with MD have produced intriguing results. ACTL8 is both associated with the microbiome composition and MD, but it is still unclear whether and/or which role the gut microbiome may have in a person's susceptibility to MD (Martins-Silva et al., 2021).

In general, both classic gene mapping approaches and colocalization analysis identified genes associated with MD or associated relevant processes. Colocalization analysis can provide insights about the effect of GVs located in non-coding regions of the genome, pinpointing the genes they regulate and the relevant tissues. As it has previously been reported the closest gene may not always be the causal one (Brodie et al., 2016; Zhu et al., 2016). These results would need further evaluation with other types of functional genomics data and ultimately experimental validation to verify the role of these regulatory mechanisms in disease pathogenesis (Dehghan, 2018).

Our goal was to illustrate the impact of the lack of standards on the selection of the most adequate post-GWAS analysis method using a fine-mapping and colocalization workflow that compared different tools. The results revealed a high divergence between fine-mapping methods due to their assumptions, which in turn highly impacted the next steps. TORUS one causal variant assumption may tip the balance in favour of PICS considering fine-mapping and posterior analytical steps. Colocalization results seem to diverge in the amount of GVs and eGenes identified, with fastENLOC being more stringent by considering the enrichment of GWAS on eQTLs.

All in all, despite the potential of combining GWAS data with molecular profiling datasets to guide in the interpretation of the functional impact of GVs located in non-coding regions of the genome, the results of our analysis revealed shortcomings related with the analytic tools. We propose that objective evaluation and benchmarking of post-GWAS analysis tools is required in order to fully leverage GWAS data for precision medicine and drug R&D applications.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

JP-G, JP, and LF designed the analysis. JP-G implemented the different analytical tools and wrote the manuscript with the support and guidance of JP and LF. All authors reviewed the manuscript. The authors read and approved the final manuscript.

Funding

IMI2-JU resources which are composed of financial contributions from the European Union's Horizon 2020 Research and Innovation Programme and EFPIA (GA: 116030 TransQST and GA: 777365 eTRANSafe), and the EU H2020 Programme 2014–2020 (GA: 676559 Elixir-Excelerate); Project 001-P-001647—Valorisation of EGA for Industry and Society funded by the European Regional Development Fund (ERDF) and Generalitat de Catalunya; Agència de Gestió d'Ajuts Universitaris i de Recerca Generalitat de Catalunya

References

- Adebiyi, E., Adam, Y., Samtal, C., Brandenburg, J. T., and Falola, O. (2021). Performing post-genome-wide association study analysis: Overview, challenges and recommendations. *F1000Res.* 10, 1002. doi:10.12688/f1000research.53962.1
- Amlic-Wolf, A., Tang, M., Mlynarski, E. E., Kuska, P. P., Valladares, O., Katanic, Z., et al. (2018). INFERN0: inferring the molecular mechanisms of noncoding genetic variants. *Nucleic Acids Res.* 46, 8740–8753. doi:10.1093/NAR/GKY686
- Berisa, T., and Pickrell, J. K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32, 283–285. doi:10.1093/BIOINFORMATICS/BTV546
- Brandes, N., Weissbrod, O., and Linal, M. (2022). Open problems in human trait genetics. *Genome Biol.* 23, 131. doi:10.1186/s13059-022-02697-9
- Brodie, A., Azaria, J. R., and Ofran, Y. (2016). How far from the SNP may the causative genes be? *Nucleic Acids Res.* 44, 6046–6054. doi:10.1093/NAR/GKW500
- Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi:10.1093/NAR/GKY1120
- Burgess, D. J. (2022). Fine-mapping causal variants — Why finding 'the one' can be futile. *Nat. Rev. Genet.* 23, 261. doi:10.1038/s41576-022-00484-7
- Calabrò, M., Mandelli, L., Crisafulli, C., Lee, S. J., Jun, T. Y., Wang, S. M., et al. (2018). Neuroplasticity, neurotransmission and brain-related genes in major depression and bipolar disorder: Focus on treatment outcomes in an asiatic sample. *Adv. Ther.* 35, 1656–1670. doi:10.1007/s12325-018-0781-2
- Cano-Gamez, E., and Trynka, G. (2020). From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* 11, 424. doi:10.3389/fgene.2020.00424
- Cazurro-Gutiérrez, A., Marcé-Grau, A., Correa-Vela, M., Salazar, A., Vanegas, M. I., Macaya, A., et al. (2021). e-Sarcoglycan: Unraveling the myoclonus-dystonia gene. *Mol. Neurobiol.* 58, 3938–3952. doi:10.1007/s12035-021-02391-0
- Chen, K., Palagashvili, T., Hsu, W., Chen, Y., Tabakoff, B., Hong, F., et al. (2022). Brain injury and inflammation genes common to a number of neurological diseases and the genes involved in the Genesis of GABAergic neurons are altered in monoamine oxidase B knockout mice. *Brain Res.* 1774, 147724. doi:10.1016/J.BRAINRES.2021.147724

(2017SGR00519), and the Institute of Health Carlos III (project IMPaCT-Data, exp. IMP/00019), co-funded by the European Union, European Regional Development Fund (ERDF, “A way to make Europe”). The Research Programme on Biomedical Informatics (GRIB) is a member of the Spanish National Bioinformatics Institute (INB), funded by ISCIII and ERDF (PRB2-ISCIII (PT13/0001/0023, of the PE I + D + i 2013–2016)). The MELIS is a ‘Unidad de Excelencia María de Maeztu’, funded by the MINECO (MDM-2014-0370). JP-G was supported by Instituto de Salud Carlos III-Fondo Social Europeo (FI18/00034). This statement is a requirement from our funding agencies and therefore has to be included in the Funding section.

Conflict of interest

Competing interest reported. LF and JP are co-founders and hold shares of Medbioinformatics Solutions SL.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1006903/full#supplementary-material>

- Dall'Aglio, L., Lewis, C. M., and Pain, O. (2021). Delineating the genetic component of gene expression in major depression. *Biol. Psychiatry* 89, 627–636. doi:10.1016/J.BIOPSYCH.2020.09.010
- Dehghan, A. (2018). Genome-wide association studies. *Methods Mol. Biol.* 1793, 37–49. doi:10.1007/978-1-4939-7868-7_4
- DISGENET plus (2022). DISGENET plus. Available at: <https://beta.disgenetplus.com/> (Accessed December 21, 2021).
- disgenetplus2r (2022). disgenetplus2r: An R package to explore the molecular underpinnings of human diseases. Available at: <https://medbio.gitlab.io/disgenetplus2r/> (Accessed December 21, 2022).
- Dream Challenges (2022). DREAM Challenges use crowd-sourcing to solve complex biomedical research questions. Available at: <https://dreamchallenges.org/> (Accessed May 26, 2022).
- Dvorakova, M., Kubik-Zahorodna, A., Straiker, A., Sedlacek, R., Hajkova, A., Mackie, K., et al. (2021). SGIP1 is involved in regulation of emotionality, mood, and nociception and modulates *in vivo* signalling of cannabinoid CB1 receptors. *Br. J. Pharmacol.* 178, 1588–1604. doi:10.1111/BPH.15383
- Dwivedi, Y. (2014). Emerging role of microRNAs in major depressive disorder: diagnosis and therapeutic implications. *Dialogues Clin. Neurosci.* 16, 43–61. doi:10.31887/DCNS.2014.16.1/YDWIVEDI
- Farh, K. K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., et al. (2015). Genetic and epigenetic fine-mapping of causal autoimmune disease variants. *Nature* 518, 337–343. doi:10.1038/NATURE13835
- Gazal, S., Weissbrod, O., Hormozdiari, F., Dey, K. K., Nasser, J., Jagadeesh, K. A., et al. (2022). Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nat. Genet.* 54, 827–836. doi:10.1038/s41588-022-01087-y
- Gémes, K., Forsell, Y., Janszky, I., László, K. D., Lundin, A., Ponce De Leon, A., et al. (2019). Moderate alcohol consumption and depression – A longitudinal population-based study in Sweden. *Acta Psychiatr. Scand.* 139, 526–535. doi:10.1111/ACPS.13034
- Genome Reference Consortium (2022). The genome reference Consortium. Available at: <https://www.ncbi.nlm.nih.gov/grc> (Accessed April 30, 2022).
- Genotype-Tissue Expression (2017). *GTEx portal*. Available at: <https://www.gtexportal.org/home/datasets> (Accessed February 18, 2021).
- Ghousaini, M., Moutnjoy, E., Carmona, M., Peat, G., Schmidt, E. M., Hercules, A., et al. (2020). Open targets genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* 49, D1311–D1320. doi:10.1093/nar/gkaa840
- Gold, S. M., Köhler-Forsberg, O., Moss-Morris, R., Mehnert, A., Miranda, J. J., Bullinger, M., et al. (2020). Comorbid depression in medical diseases. *Nat. Rev. Dis. Prim.* 6, 69–22. doi:10.1038/s41572-020-0200-2
- Hao, W. Z., Chen, Q., Wang, L., Tao, G., Gan, H., Deng, L. J., et al. (2022). Emerging roles of long non-coding RNA in depression. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 115, 110515. doi:10.1016/J.PNPBP.2022.110515
- Hormozdiari, F., van de Bunt, M., Segrè, A. v., Li, X., Joo, J. W. J., Bilow, M., et al. (2016). Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* 99, 1245–1260. doi:10.1016/J.AJHG.2016.10.003
- Howard, D. M., Adams, M. J., Shirali, M., Clarke, T.-K., Marioni, R. E., Davies, G., et al. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.* 9, 1470. doi:10.1038/s41467-018-03819-3
- Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* 22, 343–352. doi:10.1038/s41593-018-0326-7
- Hukku, A., Pividori, M., Luca, F., Pique-Regi, R., Im, H. K., and Wen, X. (2021). Probabilistic colocalization of genetic variants from complex and molecular traits: Promise and limitations. *Am. J. Hum. Genet.* 108, 25–35. doi:10.1016/j.ajhg.2020.11.012
- Karger, S., Parhar, I. S., Teo, C. H., and Soga, T. (2018). Brain beta-catenin signalling during stress and depression. *Neurosignals* 26, 31–42. doi:10.1159/000487764
- Kichaev, G., Yang, W. Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A. L., et al. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* 10, e1004722. doi:10.1371/JOURNAL.PGEN.1004722
- Kim, J. Y., Lee, W. W., Shin, C. W., Kim, H. J., Park, S. S., Chung, S. J., et al. (2017). Psychiatric symptoms in myoclonus-dystonia syndrome are just concomitant features regardless of the MGCE gene mutation. *Park. Relat. Disord.* 42, 73–77. doi:10.1016/J.PARKRELDIS.2017.06.014
- Kim, H. K., Tyryshkin, K., Elmi, N., Dharsee, M., Evans, K. R., Good, J., et al. (2019). Plasma microRNA expression levels and their targeted pathways in patients with major depressive disorder who are responsive to duloxetine treatment. *J. Psychiatr. Res.* 110, 38–44. doi:10.1016/J.JPSYCHIRES.2018.12.007
- LaPierre, N., Taraszka, K., Huang, H., He, R., Hormozdiari, F., and Eskin, E. (2021). Identifying causal variants by fine mapping across multiple studies. *PLoS Genet.* 17, e1009733. doi:10.1371/JOURNAL.PGEN.1009733
- Ledonne, A., and Mercuri, N. B. (2020). On the modulatory roles of neuregulins/ErbB signaling on synaptic plasticity. *Int. J. Mol. Sci.* 21, 275. doi:10.3390/IJMS21010275
- Li, X., Su, X., Liu, J., Li, H., Li, M., Li, W., et al. (2021a). Transcriptome-wide association study identifies new susceptibility genes and pathways for depression. *Transl. Psychiatry* 11, 306. doi:10.1038/s41398-021-01411-W
- Li, Z., Ruan, M., Chen, J., and Fang, Y. (2021b). Major depressive disorder: Advances in neuroscience research and translational applications. *Neurosci. Bull.* 37, 863–880. doi:10.1007/s12264-021-00638-3
- Martínez-Magaña, J. J., Genis-Mendoza, A. D., Villatoro Velázquez, J. A., Bustos-Gamiño, M., Juárez-Rojop, I. E., Tovilla-Zarate, C. A., et al. (2021). Genome-wide association study of psychiatric and substance use comorbidity in Mexican individuals. *Sci. Rep.* 11, 6771. doi:10.1038/s41598-021-85881-4
- Martins-Silva, T., Salatino-Oliveira, A., Genro, J. P., Meyer, F. D. T., Li, Y., Rohde, L. A., et al. (2021). Host genetics influences the relationship between the gut microbiome and psychiatric disorders. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 106, 110153. doi:10.1016/J.PNPBP.2020.110153
- Matsuda, S., Ikeda, Y., Murakami, M., Nakagawa, Y., Tsuji, A., and Kitagishi, Y. (2019). Roles of PI3K/AKT/GSK3 pathway involved in psychiatric illnesses. *Diseases* 7, 22. doi:10.3390/DISEASES7010022
- Oliva, C. A., Montecinos-Oliva, C., and Inestrosa, N. C. (2018). Wnt signaling in the central nervous system: New insights in Health and disease. *Prog. Mol. Biol. Transl. Sci.* 153, 81–130. doi:10.1016/BS.PMBTS.2017.11.018
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42, D358–D363. doi:10.1093/NAR/GKT1115
- Oscanoa, J., Sivapalan, L., Gadaleta, E., Dayem Ullah, A. Z., Lemoine, N. R., and Chelala, C. (2020). SNPNexus: A web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res.* 48, W185–W192. doi:10.1093/NAR/GKAA420
- Peall, K. J., Smith, D. J., Kurian, M. A., Wardle, M., Waite, A. J., Hedderly, T., et al. (2013). SGCE mutations cause psychiatric disorders: clinical and genetic characterization. *Brain* 136, 294–303. doi:10.1093/BRAIN/AWS308
- Pérez-Granado, J., Piñero, J., Medina-Rivera, A., and Furlong, L. I. (2022). Functional genomics analysis to disentangle the role of genetic variants in major depression. *Genes (Basel)* 13, 1259. doi:10.3390/GENES13071259
- Piñero, J., Piñero, P., Manuel Ramírez-Anguita, J., Sá Uch-Pitarch, J., Ronzano, F., Centeno, E., et al. (2019). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48, D845–D855. doi:10.1093/nar/gkz1021
- Pividori, M., Rajagopal, P. S., Barbeira, A., Liang, Y., Melia, O., Bastarache, L., et al. (2020). PhenomeXcan: Mapping the genome to the phenome through the transcriptome. *Sci. Adv.* 6, eaba2083. doi:10.1126/SCIADV.ABA2083
- Protein Structure Prediction Center (2020). Available at: <https://predictioncenter.org/> (Accessed May 26, 2022).
- Rashid-Kolvear, F., Pintilie, M., and Done, S. J. (2007). Telomere length on chromosome 17q shortens more than global telomere length in the development of breast cancer. *Neoplasia* 9, 265–270. doi:10.1593/NEO.07106
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. doi:10.1093/nar/gkz369
- Rüeger, S., McDaid, A., and Kotalik, Z. (2018). Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet.* 14, e1007371. doi:10.1371/JOURNAL.PGEN.1007371
- Shadrina, M., Bondarenko, E. A., and Slominsky, P. A. (2018). Genetics factors in major depression disease. *Front. Psychiatry* 9, 334. doi:10.3389/fpsy.2018.00334
- Shi, Y., Wang, Q., Song, R., Kong, Y., and Zhang, Z. (2021). Non-coding RNAs in depression: Promising diagnostic and therapeutic biomarkers. *EBioMedicine* 71, 103569. doi:10.1016/J.EBIOM.2021.103569
- Shi, H. (2017). Tips for formatting A lot of GWAS summary association statistics data. Available at: <https://huwenboshi.github.io/data%20management/2017/11/23/tips-for-formatting-gwas-summary-stats.html> (Accessed March 30, 2022).

- Taylor, K. E., Ansel, K. M., Marson, A., Criswell, L. A., and Farh, K. K.-H. (2021). PICS2: next-generation fine mapping via probabilistic identification of causal SNPs. *Bioinformatics* 37, 3004–3007. doi:10.1093/BIOINFORMATICS/BTAB122
- Tomasoni, M., Gómez, S., Crawford, J., Zhang, W., Choobdar, S., Marbach, D., et al. (2020). MONET: a toolbox integrating top-performing methods for network modularization. *Bioinformatics* 36, 3920–3921. doi:10.1093/BIOINFORMATICS/BTAA236
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 Years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22. doi:10.1016/j.ajhg.2017.06.005
- Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* 17, e1009440. doi:10.1371/JOURNAL.PGEN.1009440
- Wang, J. Q., and Mao, L. (2019). The ERK pathway: Molecular mechanisms and treatment of depression. *Mol. Neurobiol.* 56, 6197–6205. doi:10.1007/S12035-019-1524-3
- Wang, X. L., Yuan, K., Zhang, W., Li, S. X., Gao, G. F., and Lu, L. (2020). Regulation of circadian genes by the MAPK pathway: Implications for rapid antidepressant action. *Neurosci. Bull.* 36, 66–76. doi:10.1007/s12264-019-00358-9
- Wen, X., Pique-Regi, R., and Luca, F. (2017). Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet.* 13, e1006646. doi:10.1371/JOURNAL.PGEN.1006646
- Wen, X. (2016). Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. *Ann. Appl. Stat.* 10, 1619–1638. doi:10.1214/16-AOAS952
- Wingo, T. S., Liu, Y., Gerasimov, E. S., Gockley, J., Logsdon, B. A., Duong, D. M., et al. (2021). Brain proteome-wide association study implicates novel proteins in depression pathogenesis. *Nat. Neurosci.* 24, 810–817. doi:10.1038/s41593-021-00832-6
- Yamagata, H., and Nakagawa, S. (2020). Glycosylation and depression — a review. *Trends Glycosci. Glycotechnol.* 32, 157–160. doi:10.4052/tigg.2002.1E
- Yamawaki, Y., Yoshioka, N., Nozaki, K., Ito, H., Oda, K., Harada, K., et al. (2018). Sodium butyrate abolishes lipopolysaccharide-induced depression-like behaviors and hippocampal microglial activation in mice. *Brain Res.* 1680, 13–38. doi:10.1016/J.BRAINRES.2017.12.004
- Zhong, J., Li, S., Zeng, W., Li, X., Gu, C., Liu, J., et al. (2019). Integration of GWAS and brain eQTL identifies FLOT1 as a risk gene for major depressive disorder. *Neuropsychopharmacology* 44, 1542–1551. doi:10.1038/s41386-019-0345-4
- Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487. doi:10.1038/ng.3538