MDPI

Article

Functional Genomics Analysis to Disentangle the Role of Genetic Variants in Major Depression

Judith Pérez-Granado [1], Janet Piñero [1,2], Alejandra Medina-Rivera [3,*] and Laura I. Furlong [1,2,*]

[1] Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Department of Medicine and Life Sciences (MELIS), Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, 08003 Barcelona, Spain; jperez2@imim.es (J.P.-G.); janet.pinero@upf.edu (J.P.)

[2] MedBioinformatics Solutions SL, Almogàvers 165, 08018 Barcelona, Spain

[3] Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Campus Juriquilla, Blvd Juriquilla 3001, Santiago de Querétaro 76230, Mexico

* Correspondence: amedina@liigh.unam.mx (A.M.-R.); laura.furlong@upf.edu (L.I.F.); Tel.: +52-555-623-4331 (A.M.-R.); +34-933-160-540 (L.I.F.)

Abstract: Understanding the molecular basis of major depression is critical for identifying new potential biomarkers and drug targets to alleviate its burden on society. Leveraging available GWAS data and functional genomic tools to assess regulatory variation could help explain the role of major depression-associated genetic variants in disease pathogenesis. We have conducted a fine-mapping analysis of genetic variants associated with major depression and applied a pipeline focused on gene expression regulation by using two complementary approaches: cis-eQTL colocalization analysis and alteration of transcription factor binding sites. The fine-mapping process uncovered putative causally associated variants whose proximal genes were linked with major depression pathophysiology. Four colocalizing genetic variants altered the expression of five genes, highlighting the role of SLC12A5 in neuronal chlorine homeostasis and MYRF in nervous system myelination and oligodendrocyte differentiation. The transcription factor binding analysis revealed the potential role of rs62259947 in modulating P4HTM expression by altering the YY1 binding site, altogether regulating hypoxia response. Overall, our pipeline could prioritize putative causal genetic variants in major depression. More importantly, it can be applied when only index genetic variants are available. Finally, the presented approach enabled the proposal of mechanistic hypotheses of these genetic variants and their role in disease pathogenesis.

Keywords: major depression; genetic variants; eQTL; colocalization analysis; transcription factors; genetic regulation

1. Introduction

Major Depression (MD) is the leading cause of impairment around the world [1]. It is mainly treated with both psychotherapy and drugs, but the latter is only effective in 40% of the patients [2]. Currently, there are no available biomarkers or tests that can aid in either MD diagnosis or personalized treatment. As a complex disease, multiple genetic variants (GVs) have been associated with MD in Genome-Wide Association Studies (GWAS), most of them falling within non-coding regions of the genome [3,4].

Functional follow-up studies to unravel the regulatory mechanisms by which these GVs play a role in the disease are key to understanding the molecular underpinnings of the disease and identifying biomarkers or new drug targets. Some authors propose that the efforts should be centered on the interpretation of GWAS signals to identify the causal GVs, meaning those with a biological effect on a disease, and their regulatory potential, instead of pursuing more GWAS [5].

In this study, we have focused on the GWAS meta-analysis on MD performed in 2019 by Howard et al. [3]. Full-genome summary statistics are not publicly available for

this GWAS, so we have leveraged available data on index GVs. Ninety-seven loci were identified as significantly associated with MD, and these underwent the classic post-GWAS analysis: a gene-set enrichment analysis, the computation of polygenic risk score, and genetic correlation with other traits, as well as drug-gene interaction analysis. In line with previous GWAS findings, most GVs lie in non-coding regions, thus having no obvious direct effect on a gene.

A necessary step forward to disentangle the role of GVs identified in GWAS requires the evaluation of functional regulatory variation. Here, we have pursued two complementary analytical approaches geared toward the use of index GVs: (1) identification of candidate susceptibility genes using expression quantitative trait loci in cis (cis-eQTLs), which are enriched among disease-associated loci [6], and (2) characterization of transcription factor (TF) binding sites modified by GVs, which are key to understanding their potential impact on regulatory mechanisms [6–8].

In the present study, we aim to advance the understanding of MD molecular underpinnings. We have designed and applied a regulatory variation analysis pipeline and conducted a functional enrichment analysis of the GVs, either acting as eQTLs or altering the transcription factor binding site (TFBS), along with the proximal (pGenes) and regulated genes (eGenes). Our findings provide biological insights into the functional role of MD GVs and enable the proposal of mechanistic hypotheses.

2. Materials and Methods

2.1. MD GWAS Dataset and LD expansion

In order to obtain a comprehensive and reliable set of genetic variants (GVs) associated with major depression (MD), we focused our analysis on the GWAS meta-analysis from Howard et al. [3]. This meta-analysis evaluated 807,553 European individuals (246,363 cases and 561,190 controls) and identified 102 genetic variants (GVs) associated with MD. We retrieved these data from the summary statistics available at GWAS Catalog [9] (Accession Study: GCST007342, note that the full-genome summary statistics for this GWAS were not publicly available; downloaded in December 2020). We filtered the GVs by genome-wide significance (p-value $\leq 5 \times 10^{-8}$ and proceeded with the analysis with this set. We then fine-mapped MD-associated GVs to prioritize the causal ones using the Probabilistic Identification of Causal SNPs (PICS) algorithm [10]. In brief, PICS takes the most significant variant per association locus and performs LD expansion using the 1000 Genomes Project linkage disequilibrium (LD) information data for the study population and then identifies the GVs more likely to be causal (PICS probabilities). Using the PICS2 Data portal, we downloaded the precomputed PICS GVs for this study. This data constituted our full dataset of GVs.

2.2. GVs Annotation: VEP, CADD and ENCODE

We annotated the full set of GVs with Variant Effect Predictor (VEP) [11] and Combined Annotation Dependent Depletion (CADD) [12]. VEP annotates GVs' consequence type using the Sequence Ontology, its allele frequency from the 1000 Genomes Project Phase 3 along with the genomic coordinates, chromosome, and mapped gene at ± 5000 bp distance (from now on pGenes). Combined Annotation Dependent Depletion (CADD) assesses GVs' potential pathogenicity by evaluating the PHRED-like scaled C-score; the recommended cut-off ≥ 15 was set to identify potentially pathogenic variants.

We analyzed the GVs with the Encyclopedia of DNA Elements (ENCODE) [13] to identify those potentially lying in transcription factor binding sites (TFBS). ENCODE data analysis was performed using SNPNexus [14], an online platform that allows a comprehensive annotation of GVs by integrating multiple tools.

2.3. Fine-Mapping and Colocalization of GWAS and cis-eQTLs

PICS2, in addition to GWAS PICS GVs, has precomputed PICS GVs for all Genotype-Tissue Expression (GTEx) V8 best eQTLs per gene, per tissue type. We overlapped the

extracted GWAS PICS for MD GVs with GTEx cis-eQTL PICS GVs, filtering both sets by a PICS probability greater than 10% to narrow down the set to the most likely causal GVs without being overly permissive, as previous applications of this method have done [15]. We performed a Fisher test to assess the enrichment of GVs in eQTL regions. Finally, to identify colocalizing GWAS and eQTL GVs, we computed the products of PICS probabilities following the colocalization posterior probability (CLPP) method, which assumes independence of causal probabilities for GWAS and eQTL GVs [16]. The genes regulated by these eQTLs from now on will be referred to as eGenes.

2.4. TF Binding Analysis with RSAT Variation Tools

We predicted those GVs affecting the TFBS using the Regulatory Sequence Analysis Tools (RSAT) suite, which evaluates cis-regulatory elements. First, we used ENCODE ChIP-seq data to keep only the GVs lying in TFBS and, therefore, have a more biologically relevant set of GVs and reduce the number of tests. However, ChIP-seq data retrieve regions of around 100–1000 bp, but the actual binding site corresponds to 9–15 bp [17,18]. Thus, we proceeded with the RSAT analysis for a more robust and accurate assessment of the GVs potentially altering the TFBS. RSAT provides tools that evaluate cis-regulatory elements to predict GVs affecting the TFBS by modifying the transcription factor (TF) binding affinity.

RSAT modular structure allowed the concatenation and independent execution of programs, each with a different goal. Before scanning the GVs and in order to account for their different nucleotide composition, we created four sets of background models according to the GV's functional impact obtained with VEP (i.e., intergenic and UTR, intronic, regulatory, and non-coding GVs). The subsequent steps were performed for each set separately. The module *create-background-model* was executed using the sequences obtained with *fetch-sequences-from-UCSC*, with the peak regions retrieved by ENCODE as input. In parallel, the module *retrieve-variation-sequence* was used to obtain the flanking sequence (30 bp per side) of the GVs of interest, using the dbSNP, genomic coordinates, reference, and alternative allele.

To assess the TFBS alterations, position weight matrices (PSSM) for TFs expressed in brain tissues (filtering them using GTEx expression data, ≥ 2 transcripts per million (TPM)) [19] were retrieved from the following databases: JASPAR [20], ENCODE, HOCOMOCO [21], footprintDB [22], and hPDI [23]. In all cases, the non-redundant Homo sapiens database version was used.

Finally, the module *variation-scan* was run with the previously built background Markov models (order 2 to account for CpGs without overfitting), the PSSM matrices, the GVs with their flanking sequences (see above), and the following parameters: weight of predicted sites (>1), weight difference between variants (>1), *p*-value of predicted sites (<1 e-3), and *p*-value ratio between variants (>10). The weight represents the binding affinity and the *p*-value of a score is the probability of observing a score of at least weight given a background model.

In addition, two control datasets, one randomizing TF motifs and one randomizing GVs, were built to validate the results obtained running RSAT with the GVs of interest. On the one hand, the TF's PSSMs matrices were permuted using *permute-matrix -perm 5* to get randomized matrices with the same nucleotide composition and information content. On the other hand, a control set of GVs (1:10) was built using vSampler [24] with the following parameters: distance to closest transcription start site (TSS) deviation (± 5000), gene density deviation (± 5 in 100 kbp), number of variants in LD (± 50 and $r^2 = 0.1$), controlling for coding/non-coding match and variant type specificity, excluding for input GVs and sampling across the chromosome. Both controls were analyzed with the described RSAT pipeline.

We compared our set of GV-TF motif pair *p*-value ratios against the distribution of *p*-value ratios for the given motif in both control datasets. A Wilcox test was used to evaluate the results obtained from the controls because normality of *p*-value ratio distribution

could not be assumed for most motifs after running a Shapiro–Wilk test. The alternative hypothesis tested was "greater".

In addition, to further confirm statistically significant GVs, a larger negative control dataset of GVs (1:1000) was generated. Again, vSampler was used with relaxed parameters to get a bigger control set (i.e., controlling for coding/non-coding match and variant type specificity, excluding for input SNPs, and sampling across chromosomes). The same non-parametric test was used to evaluate the results.

2.5. Identification of TF Active Regions with ChromHMM

We used chromatin state annotations from ChromHMM [25,26], available from EN-CODE (v3), to evaluate whether GVs significantly altering the TFBS were lying in active transcription sites of brain regions. Under a 18-state ChromHMM model, we consider the following states annotations as active regulatory regions [26]: TssA, TssFlnk, TssFlnkU, TssFlnkD, Tx, TxWk, EnhG1, EnhG2, EnhA1, EnhA2, EnhWk, ZNF/Rpts. The available brain regions and cell types were: Brain Angular Gyrus, Brain Inferior Temporal Lobe, Brain Cingulate Gyrus, Brain Anterior Caudate, Brain Substantia Nigra, Brain Dorsolateral Prefrontal Cortex, Brain Hippocampus Middle, and Astrocytes. Additionally, the resulting TFs whose binding was altered were filtered by their expression in the specific brain region using GTEx matched data when available; otherwise, data for all brain regions were considered.

2.6. Retrieval of Regulation Evidence

We looked for evidence of gene expression regulation of TFs by matching GVs-TFs pairs from the TF binding analysis using RSAT with eQTL PICS GVs. We further explored the hTFtarget database [27] to identify specific mechanistic regulation evidence of those TFs whose binding is altered by our set of GVs to regulate the expression of the target eGenes. The hTFtarget database contains associations of TFs and their targets from chromatin immunoprecipitation sequencing (ChIP-seq) in a specific tissue. We considered evidences for mechanistic regulation when eQTL and ChIP-seq data tissues matched.

2.7. pGenes, eGenes, and GVs Characterization

We conducted a gene-set enrichment analysis using the tool g:Profiler via the R package gprofiler2 [28], which integrates different resources and annotates enriched terms at the following levels: (1) biological processes, molecular functions, and cellular processes annotated with the Gene Ontology (GO); (2) pathways from Reactome (REAC) and WikiPathways; (3) miRNA annotations from MIRNA; (4) phenotypic features associated to disease from Human Phenotype, which is mainly focused on rare Mendelian disorders. In addition, we included DISGENET plus [29,30] association data (v16) in this analysis to evaluate the annotation of complex diseases and phenotypic traits; note that the study by Howard et al. was removed from this dataset to avoid circularity. Variant-set functional enrichment analysis was performed using variant association data from DISGENET Plus. We considered the set of known genes as the domain scope for the analysis. Furthermore, we characterized tissue expression using GTEx gene expression data (v8).

We performed these analyses for the following two gene-sets: (1) genes mapped to by MD-associated GVs (pGenes) and (2) genes regulated by cis-eQTLs (eGenes), and two variant-sets: (1) causal GVs and (2) colocalizing GVs.

3. Results

3.1. Major Depression Associated Genetic Variants Lie in Non-Coding Regions of the Genome

The GWAS study by Howard et al., 2019, reported 102 risk loci associated with major depression (MD), 97 with a *p*-value $\leq 5 \times 10^{-8}$, which were the starting point of our analysis. After LD expansion, we obtained a set of 5723 potentially causal genetic variants (GVs) (Supplementary Scheme S1 and Table S1). We annotated these GVs with VEP [11] and CADD [12] (Supplementary Figure S1). The identification of probable causal GVs

using PICS fine-mapping GWAS data [10] revealed 172 GVs (PICS >10%) in LD with the 97 GWAS risk loci (Supplementary Table S2). These GVs are located in different regions of the genome, but most of them are in non-coding regions, being mainly annotated as intronic (30%), intergenic (30%), or located in non-coding transcript regions (17%) (Figure 1A). Only two GVs lie in exonic regions (i.e., synonymous and nonsynonymous consequence types). The median allele frequency of these GVs was 0.364 (with more deleterious consequence types having lower allele frequencies) (Figure 1B). Only 4% (7) of the GVs were predicted by CADD as potentially pathogenic (Figure 1C). The fine-mapped GVs were assigned to 95 proximal genes (±5000 bps), from now on referred to as pGenes. pGenes were classified based on their expression across tissues based on GTEx gene expression data [19]. Using hierarchical clustering, genes were divided into three roughly equally distributed clusters that seem to correspond to constitutively, lowly expressed in all tissues, and highly expressed in brain tissues (Supplementary Figure S2).
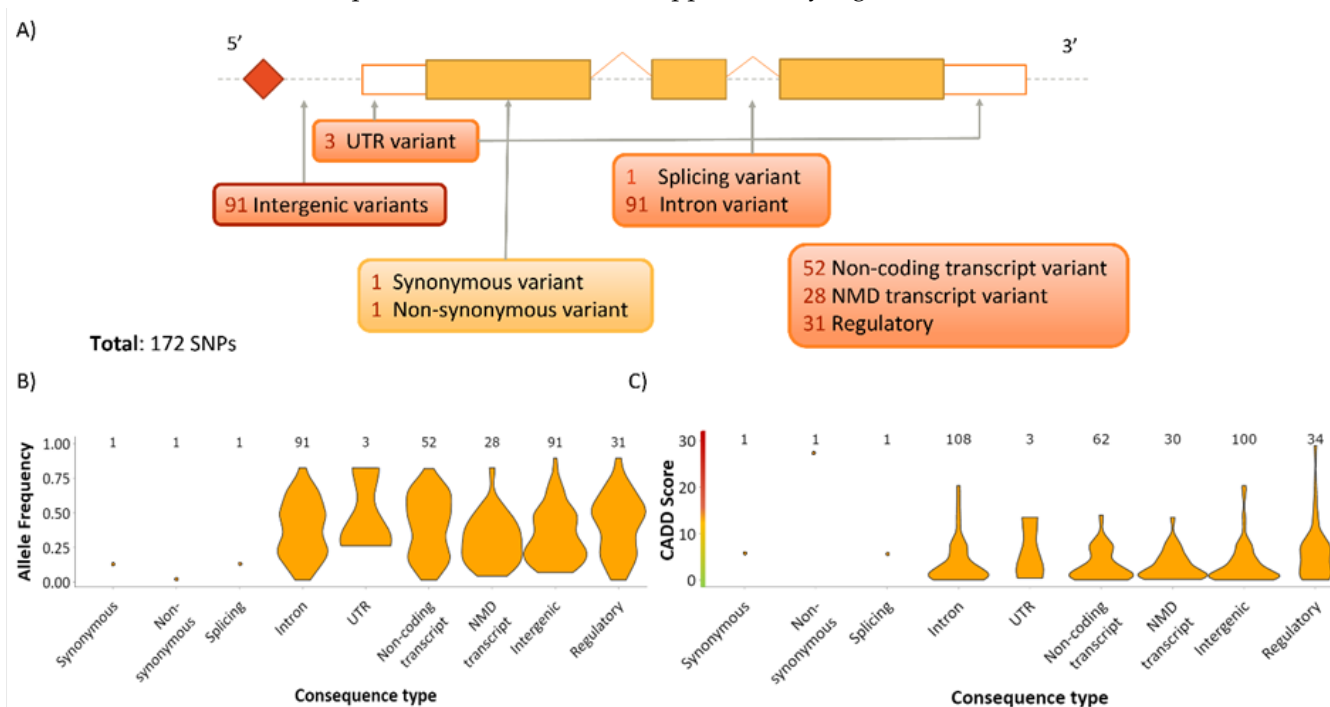

Figure 1. MD GVs are mostly non-coding, common, and potentially not pathogenic. (**A**) GVs distributed along the genome according to their consequence type predicted with VEP. (**B**) Allele frequency density, according to GV's consequence type, also predicted with VEP. (**C**) Pathogenicity score (predicted by CADD) density per consequence type. Please note that a GV can have multiple consequence types; thus, the number of consequence types may not necessarily match the total number of GVs. MD: major depression; GV: genetic variant; VEP: Variant Effect Predictor; CADD: Combined Annotation Dependent Depletion; SNP: single nucleotide polymorphism; UTR: untranslated region; NMD: nonsense- mediated decay.

The pGenes are functionally enriched in GO terms related to nervous system development, neuron differentiation, synaptic signaling, and different cellular components of the neuron such as dendrite, axon, or synapse (Supplementary Table S3); these biological processes and molecular functions are involved in the pathophysiology of MD [31]. pGenes are associated with an abnormal nervous system morphology and physiology according to the Human Phenotype ontology. Disease enrichment analysis shows enrichment for the association of both pGenes and causal GVs with major depressive disorder and other related mental disorders such as schizophrenia or bipolar disorder (Figure 2 and Supplementary Table S4). pGenes are also associated with comorbid phenotypes and conditions in MD, such as smoking behavior, body mass index, and duration of sleep [32]. Notably, 37% of

pGenes and 42% of GVs have no previous evidence of association with depression or other mental disorders.
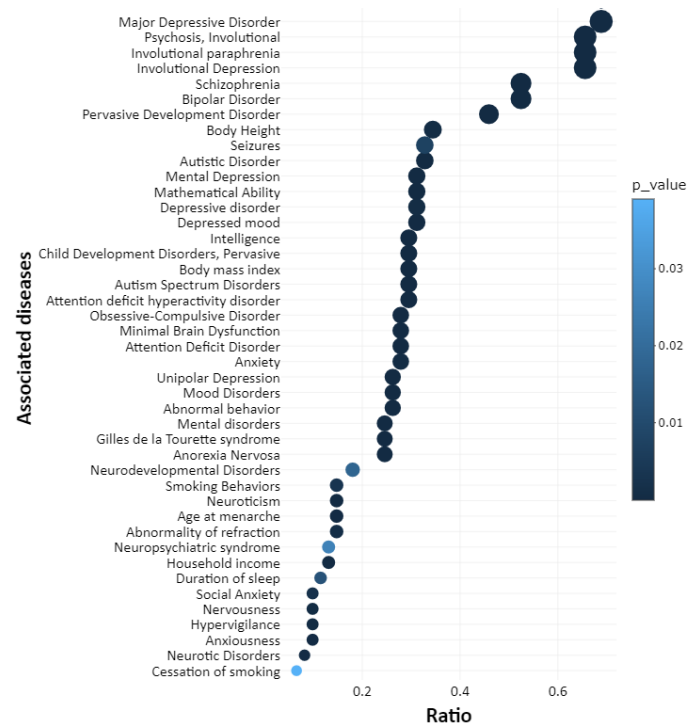
Figure 2. pGenes are associated with mental disorders. Result of the disease enrichment analysis. The ratio corresponds to the number of pGenes associated with each disease relative to all pGenes. Dot size is proportional to the number of pGenes associated with each disease. Gene enrichment analysis was performed using g:Profiler and the DISGENET plus database. pGenes: proximal genes.

Some of the pGenes are associated with processes related to MD pathogenesis, such as TLR4, involved in immune response [33], ESR2, a regulator of estrogen response [34], TCF4, with a role in nervous system development [35], DCC, in charge of axon guidance and neuronal connectivity [36], PAX5, which interferes in mouse neural stem cells proliferation and migration [37,38], and CYP7B1, that participates in the metabolism of the neurosteroids DHEA and pregnenolone [39]. Among the potentially pathogenic GVs, according to CADD, there are 3 intronic GVs lying in ZNF536, a gene involved in the negative regulation of neuron differentiation [40], a relevant process in MD pathogenesis and treatment [41]. rs1021362 lies in SORCS3, a gene previously associated with stress response associated with MD [37,42], rs3793577 lies in ELAVL2, whose silencing in animal models is associated with reduced behavioral despair [43]; the remaining GVs have been previously associated with major depression by several PheWAS studies [15].

3.2. Major Depression Causal Genetic Variants Regulate the Expression of Genes in Cis

The 172 fine-mapped GWAS GVs overlap with 13 GTEx PICS GVs (Scheme 1), revealing an enrichment of MD causal GVs in eQTLS (p-value = 7.392×10^{-10}). The colocalization analysis to identify GVs associated with both MD GWAS and cis-eQTLs resulted in 5 GV–eGenes pairs (i.e., genes whose expression is regulated by these GVs; rs10149470—BAG5, rs10149470—RP11-894P9.2 [ENSG00000258851.1], rs12624433—SLC12A5, rs198457—MYRF, rs301799—RP5-1115A15.1 [ENSG00000232912.5]), with a colocalization probability greater than 10% (Table 1). BAG5 and SLC12A5 are involved in neuron projection [44,45] and MYRF in central nervous system myelination [46]. In addition, all eQTLs have been previously associated with MD and other mental disorders according to DISGENET plus [30,47,48] (Supplementary Table S5). The eGenes BAG5, SLC12A5, and MYRF show higher expression

levels in brain regions according to GTEx (Supplementary Figure S3). Little is known about the function of the long non-coding RNAs RP11-894P9.2 and RP5-1115A15.1.
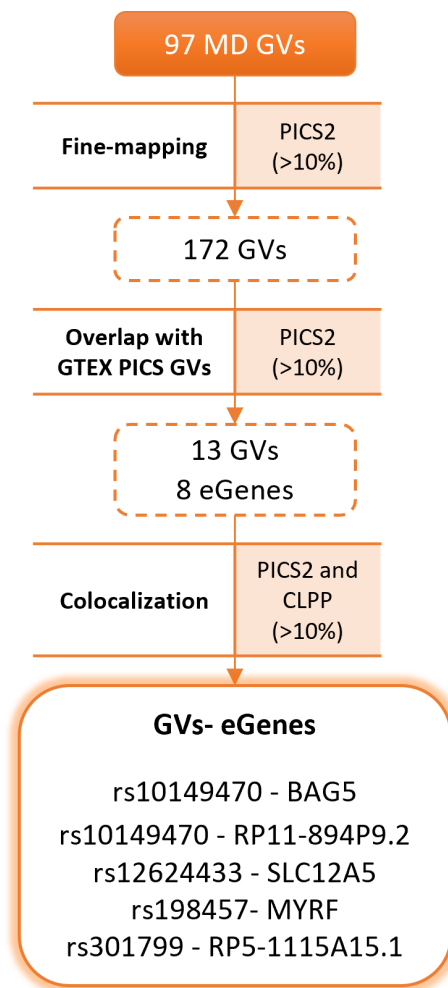
Scheme 1. Fine-mapping and colocalization analysis of MD GVs. MD GWAS GVs have been fine-mapped using PICS and overlapped with GTEx PICS GVs to ultimately perform a colocalization analysis identifying 4 colocalizing GVs affecting the expression of 5 eGenes. MD: major depression; GV: genetic variant; GWAS: genome-wide association studies; PICS: Probabilistic Identification of Causal SNPs; GTEx: Genotype-Tissue Expression; eGenes: genes regulated by expression quantitative trait loci; CLPP: colocalization posterior probability.

3.3. MD Associated GVs Affect the TFBS in Regulatory Regions of Genes Relevant for the Disease

The initial set of 5723 GVs associated with MD was first mapped to transcription factor binding sites (TFBS) using Chip-Seq data from ENCODE. A total of 955 GVs were identified as potentially altering the TFBS of 155 TFs (Scheme 2). The GVs' functional impact was assessed with VEP, and 4 sets were created: (a) intergenic and UTR GVs (333), (b) intronic GVs (562), (c) regulatory GVs (303), and (d) non-coding GVs (389). In addition, we further selected those transcription factors (TFs) that were expressed in brain tissues ($\geq$2 TPM), which left 115 TFs.

Using a pattern matching approach (*variation-scan*) [49], we identified GVs likely affecting TFBS. As negative controls, we permuted TF motifs and randomly selected variants matching GVs properties (see Methods). Using permuted motifs and randomly selected variants (1:10) as negative controls, we obtained a total of 306 GVs significantly altering the TFBS of 102 TFs (considering the 4 sets together). Ultimately, 289 GVs and 101 TFs passed the statistical analysis using randomly selected variants (1:1000) as negative

control. From this final set, 171 GVs are predicted to disrupt the TFBS of 89 TFs, whereas 143 GVs are predicted to create a TFBS for 82 TFs (Supplementary Table S6). Most of these GVs were not characterized as potentially pathogenic by CADD except for 11 GVs (score ≥ 15).

Table 1. GWAS-eQTL Colocalizing GVs in MD. MD associated GVs colocalizing with eQTLs. GWAS: genome-wide association studies; eQTL: expression quantitative trait loci; GV: Genetic variant; MD: major depression; eGene: gene regulated by eQTL; PICS: Probabilistic Identification of Causal SNPs.

| GV	eGene	Tissue	PICS Probability GWAS	PICS Probability eQTL	Colocalization Probability
rs10149470	BAG5	Artery Tibial	0.9657	0.633	0.6112881
rs10149470	RP11-894P9.2	Colon Sigmoid	0.9657	0.633	0.6112881
rs10149470	RP11-894P9.2	Esophagus Gastroesophageal Junction	0.9657	0.633	0.6112881
rs10149470	RP11-894P9.2	Esophagus Muscularis	0.9657	0.584	0.5639688
rs10149470	RP11-894P9.2	Artery Aorta	0.9657	0.499	0.4818843
rs10149470	RP11-894P9.2	Breast Mammary Tissue	0.9657	0.4494	0.43398558
rs12624433	SLC12A5	Brain Putamen Basal Ganglia	0.7355	0.303	0.2228565
rs10149470	RP11-894P9.2	Stomach	0.9657	0.1782	0.17208774
rs10149470	RP11-894P9.2	Adipose Subcutaneous	0.9657	0.1621	0.15653997
rs10149470	RP11-894P9.2	Colon Transverse	0.9657	0.1419	0.13703283
rs10149470	RP11-894P9.2	Adipose Visceral Omentum	0.9657	0.1412	0.13635684
rs198457	MYRF	Thyroid	0.9627	0.1258	0.12110766
rs10149470	RP11-894P9.2	Heart Left Ventricle	0.9657	0.1225	0.11829825
rs301799	RP5-1115A15.1	Whole Blood	0.6946	0.1542	0.10710732

A total of 270 GVs lie in active regulatory regions of the genome of brain-related tissues and cell types according to the epigenome annotation from the ENCODE project based on ChromHMM data (Supplementary Table S7) [25,26]. We then looked for evidence of their impact on gene expression regulation by evaluating their annotation to GTEx eQTLs fine-mapped via PICS. The only GV in this dataset of 270 GVs that also fulfills the criteria of being causal and overlapping GWAS and eQTL PICs in the brain with a probability greater than 10% was rs12624433, which is an eQTL for the gene SLC12A5. This GV is predicted to significantly alter the TFBS of USF1, USF2, and MYC. Both rs12624433 and SLC12A5 have been previously associated with major depression disorder and other mental disorders such as bipolar disorder or schizophrenia [48].

In addition, we also inspected the hTFtarget database [27], looking for evidence of a mechanistic association between the eGenes, considered the targets, and the TFs whose binding site is being altered by the GVs. Focusing on brain regions, we have evidence for two GV-TF-eGene/target associations (rs11227217: RAD21 -> ZNRD2-DT [ENSG00000260233.3]; rs62259947: YY1 -> P4HTM).

The GV rs62259947 has been annotated as an eQTL downregulating the expression of P4HTM in the Brain Cerebellar Hemisphere. We propose this effect is likely being mediated by the GV significantly changing the affinity for YY1 binding (weight difference = 14.98 and *p*-value ratio = 5058.82) (see Methods), a TF known to participate in gene regulation though looping of the DNA [50]. The eGene P4HTM has been associated with the hypoxia-

inducible factor HIF1α mediating calcium signaling [51], and its null mutation reduces behavioral despair [52] (Figure 3).
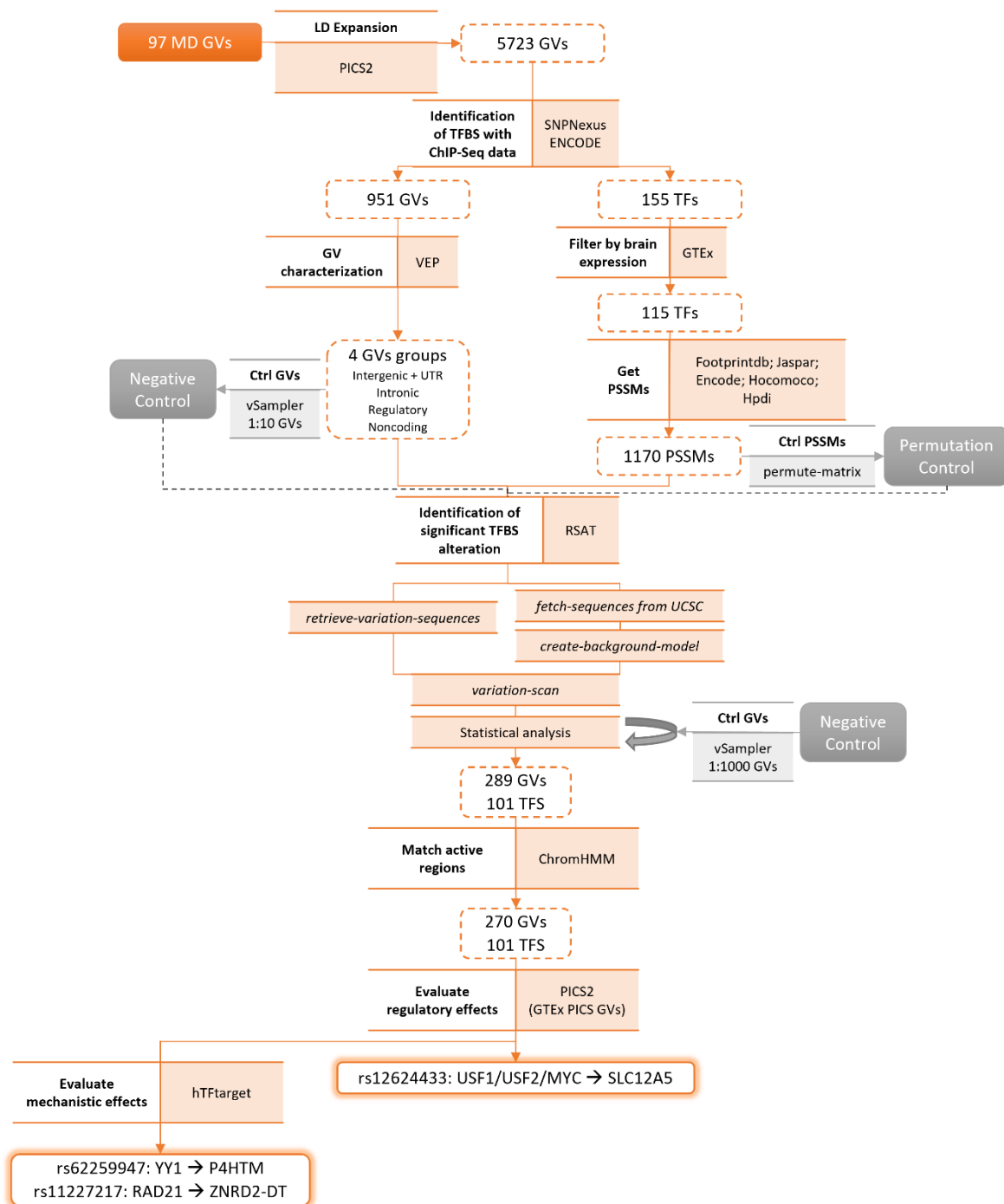
Scheme 2. Identification of altered TFBS using RSAT. Pipeline followed to identify GVs associated with MD that significantly alter TFBS. Methodologies are referred to in bold and along with them are the resources used. Highlighted in grey are the control datasets. TFBS: transcription factor binding site; RSAT: Regulatory Sequence Analysis Tools; GV: genetic variant; MD: major depression; LD: linkage disequilibrium; ENCODE: Encyclopedia of DNA Elements; VEP: Variant Effect Predictor; Ctrl: control; UTR: untranslated region; TF: transcription factor; GTEx: Genotype-Tissue Expression; PSSM: position weight matrix; PICS: Probabilistic Identification of Causal SNPs.
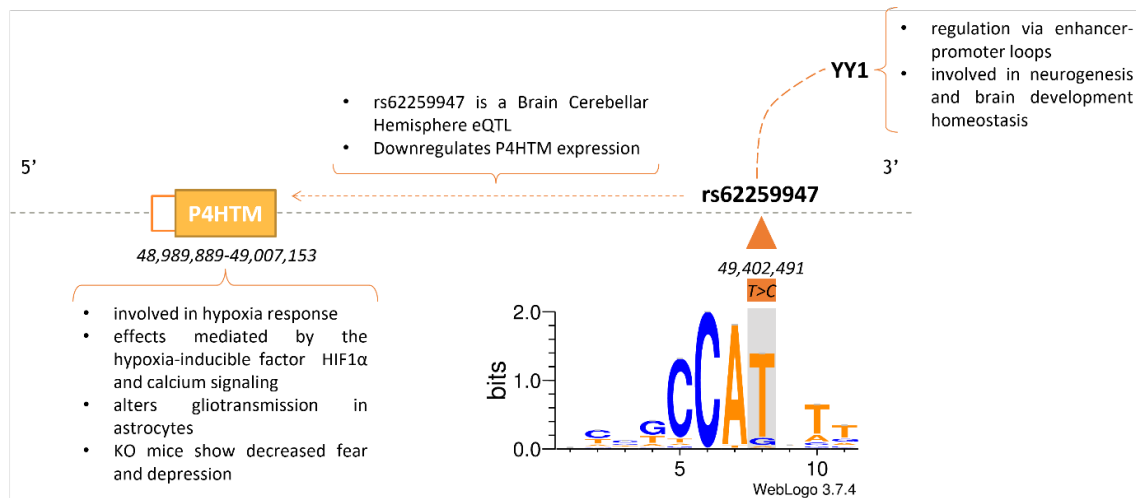
Figure 3. The GV rs62259947 might disrupt the binding of YY1, thus affecting the expression of P4HTM and resulting in behavioral alterations. rs62259947 is an eQTL downregulating the expression of P4HTM and is predicted to disrupt the binding of the TF YY1. This is represented by the sequence logo of the binding site with the nucleotide change highlighted in grey. YY1 is involved in neurogenesis and, in turn, controls the expression of P4HTM, which is mediated by HIF1α regulates calcium signaling and is also associated with behavior. GV: genetic variant; eQTL: expression quantitative trait loci; TF: transcription factor; KO: knockout.

4. Discussion

Despite the large volume of genetic information available, the pathogenesis and etiology of MD are not yet fully understood, probably because most GVs lie in non-coding regions with no obvious direct effect on a gene or function. In this context, leveraging multiple omics data is key for moving forward in the understanding of the influence of genomic variants in MD disease development. On top of that, full-genome summary statistics are not readily available due to study sharing policies (especially for private–public research partnerships) hampering the usage of most post-GWAS data analysis tools. This study aims to unravel the role of MD GVs in genetic regulation by focusing on regulatory variation following two complementary approaches: cis-eQTLs and TF binding alterations. Both are key to identifying potentially causal genes and understanding gene expression regulation [6,8], as reported by supporting evidence for its association with other mental disorders [53–55] and with MD in particular [56–58]. The regulatory variation analysis pipelines we have implemented involve fine-mapping, cis-eQTL colocalization, transcription factor binding analysis, and chromatin accessibility data, specially designed to perform well when full-genome summary statistics are not available [59]. These pipelines are in line with other approaches that leverage available omics data, and as such, they could be applied to other complex disorders with a similar genetic architecture and similar data access issues [53,60,61].

Multiple GVs have been associated with MD, most of them characterized as not potentially pathogenic in addition to being common and mostly in non-coding regions of the genome according to CADD and VEP, respectively (Figure 1). The fine-mapping of MD GVs identified 172 causal GVs and 95 pGenes (Supplementary Table S2). The functional enrichment analysis of pGenes stands along with hypotheses of MD pathogenesis such as alterations in neurogenesis and neuroplasticity or the circadian rhythm theory [31]. Additionally, these are also enriched for other phenotypes frequently co-occurring with MD, such as alterations of body mass index or smoking [32]. While most pGenes (63%) and GVs (58%) have previous evidence for association with MD, our study pinpoints novel pGenes and GVs (Supplementary Table S3 and S4). Additionally, existing literature supports the role of pGenes in processes related to MD pathogenesis, such as immune response, nervous system development, response to stress, or behavioral despair.

MD causally associated GVs are those most likely to be causal and functioning as eQTLs and, indeed, proved to be significantly enriched in cis-eQTLs from GTEx, in line with previous findings on MD and other psychiatric disorders [53,62]. The colocalizing eGenes are involved in processes relevant to MD, such as neuron projection [63], and have been associated with MD and related phenotypes according to DISGENET plus [47,48]. BAG5 is constitutively expressed in all tissues, while MYRF and SLC12A5 show higher levels in brain tissues (Supplementary Figure S2). BAG5 has been previously identified as associated with MD [64]. We characterize SLC12A5, involved in chloride homeostasis in neurons, as a pGene, also, and its downregulation has been described as an effect of stress leading to the activation of the hypothalamic–pituitary–adrenal axis, which ultimately can lead to MD-like symptoms [31,65]. However, rs12624433 is an eQTL in the Brain Putamen basal ganglia associated with the upregulation of SLC12A5. Thus, more research is needed to unravel the exact mechanism by which rs12624433 exerts its role in the regulation of the expression of SLC12A5. This eGene has been described as a potential drug target for mental disorders, but considerations should be taken given its important role in brain development; besides, it is highly influenced by exercise and environmental factors [65]. rs198457 mediates the downregulation of MYRF expression, which plays a role in myelination and oligodendrocyte differentiation [46]. These, in turn, require thyroid hormones for their differentiation and maturation [66]. Furthermore, oligodendrocytes have been stated as crucial for psychological functions likely involved in mental disorders such as MD [67].

The analysis of TF regulation with RSAT enabled a precise prediction of TF binding alterations. Although TF expression is not highly tissue-specific [7,68], for this type of analysis, it is key to pick meaningful sets of TFs and GVs [69]. We focused on TF expressed in brain-related tissues as it has been previously reported that genes involved in depression are highly expressed in brain regions [4,32,37,47]. Our analysis resulted in the prediction of 270 GVs lying in active regulatory regions of the genome of brain-related tissues based on chromatin accessibility data. These GVs alter the binding of 101 TFs, roughly equally distributed as disrupting or creating a binding site. The activating or repressing role of these TFs is hard to interpret since it will always depend on the sequence context and the cofactors involved [68]. Thus, further analysis is required to elucidate the impact of these GVs on gene expression regulation. Our pipeline enabled us to filter and prioritize the large number of candidate GVs by combining different omics data and ultimately propose mechanistic hypotheses.

By using eQTL data, we were able to identify the GV rs12624433, which regulates the expression of SLC12A5. This GV, previously referred to as colocalizing, is predicted to alter the binding of the TFs USF1, USF2, and MYC; these belong to the bHLH family involved in neural development [70]. USF1 and USF2 generally exert activating effects [71], with USF1 being a risk gene for Alzheimer's disease and relevant for brain cholesterol metabolism involving its transport from astrocytes to neurons [72].

Additionally, we found mechanistic evidence for 2 GV-TF-eGene/target associations (rs11227217: RAD21 → ZNRD2-DT; rs62259947: YY1 → P4HTM) when combining pattern matching results, chromatin accessibility data, GTEx eQTLs PICS, and hTFtarget data. Variant rs11227217 is more than 20 kbp away from ZNRD2-DT, but RAD21 is a member of the cohesion complex, which enables genes and enhancers to interact via loop formation [73,74]. NRD2-DT is a lncRNA, and interestingly, our findings include several lncRNAs in the set of pGenes as well as related with regulatory variations, either colocalizing with cis-eQTLS (RP11-894P9.2 and RP5-1115A15.1) or with mechanistic evidence for its association with gene expression regulation (ZNRD2-DT). Although not their exact role in MD pathophysiology is not clear, ncRNAs have been described as promising biomarkers and drug targets for MD [75,76].

Regarding the association rs62259947: YY1 → P4HTM, P4HTM has been related to neurological disorders and social behavior (Figure 3) [51,52]. It is involved in Ca^{2+} signaling mediated by the hypoxia-inducible factor HIF1α altering astrocytes gliotransmis-

sion [51]. Indeed, hypoxia has been associated with mental disorders in general and MD in particular [77–80]. In addition, P4HTM null mutation results in a reduction in fear and depression [52]. In turn, rs62259947 downregulates the expression of P4HTM and changes the binding affinity of YY1 in the Brain Cerebellar Hemisphere. Additionally, YY1 regulates transcription by forming loops, although its specific role as activator or repressor is not yet fully understood [50]. Furthermore, P4HTM and HIF1α have been reported as potential drug targets for MD [52,81]. rs11227217 and RAD21 are associated with red blood cell and reticulocyte count, respectively, by PheWAS [15]. Indeed, red blood cell parameters have been described as altered in patients with mental disorders [82].

5. Conclusions

Overall, we have successfully developed and applied a regulatory variation analysis pipeline including fine-mapping, colocalization, TF regulation analysis, and chromatin accessibility data, which overcomes the limitation of the lack of full-genome summary statistics. We have identified causal GVs, pGenes, and eGenes and proposed hypotheses for their role in MD pathogenesis, highlighting the role of chloride homeostasis and myelination. We also found mechanistic evidence involving hypoxia response mediated by altered TF binding. Our findings include GVs and genes supported by the literature on MD and mental disorders, as well as novel molecular mechanisms underlying MD pathogenesis.

RNA-Seq data can be downloaded from https://www.gtexportal.org/home/datasets (accessed on 18 February 2021) (filename: GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct. gz); ENCODE data can be accessed from the following Accession Numbers: ENCSR674KAN, ENCSR801APH, ENCSR826BFW, ENCSR658SFK, ENCSR082KYZ, ENCSR363VGK, ENCSR738WFF and ENCSR860PXK. The data from JASPAR, ENCODE, HOCOMOCO, footprintDB and hPDI is available at (http://rsat.sb-roscoff.fr/retrieve-matrix_form.cgi (accessed on 24 February, 2 and 4 March 2021), see View matrix descriptions and download full collections); and hTFtarget data can be downloaded from http://bioinfo.life.hust.edu.cn/hTFtarget#!/download (accessed on 15 November 2021) (filename: TF-Target-information.txt). The data generated by the current study as a result of the analysis of the above-mentioned datasets are available at Zenodo (https://doi.org/10.5281/zenodo.6838470) and are also available in the Supplementary Tables File with this manuscript.

Abbreviations

CADD	Combined Annotation Dependent Depletion
CLPP	colocalization posterior probability
eGenes	genes regulated by eQTLs
ENCODE	Encyclopedia of DNA Elements
eQTLs	expression quantitative trait loci
GO	Gene Ontology
GV	genetic variant
GTEx	Genotype-Tissue Expression
GWAS	genome-wide association studies
LD	linkage disequilibrium
MD	major depression
pGenes	proximal genes
PICS	Probabilistic Identification of Causal SNPs
PSSM	Position-specific scoring matrix or position weight matrix
REAC	Reactome
RSAT	Regulatory Sequence Analysis Tools
TF	transcription factor
TPM	transcripts per million
TFBS	transcription factor binding site
VEP	variant effect predictor

References

1. World Health Organization: Depression. Available online: https://www.who.int/news-room/fact-sheets/detail/depression (accessed on 21 December 2021).
2. Preskorn, S.H. Drug Development in Psychiatry: The Long and Winding Road from Chance Discovery to Rational Development. In *Handbook of Experimental Pharmacology*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 250, pp. 307–324. [CrossRef]
3. Howard, D.M.; Adams, M.J.; Clarke, T.-K.; Hafferty, J.D.; Gibson, J.; Shirali, M.; Coleman, J.R.I.; Hagenaars, S.P.; Ward, J.; Wigmore, E.M.; et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **2019**, *22*, 343–352. [CrossRef] [PubMed]
4. Wray, N.R.; Ripke, S.; Mattheisen, M.; Trzaskowski, M.; Byrne, E.M.; Abdellaoui, A.; Adams, M.J.; Agerbo, E.; Air, T.M.; Andlauer, T.M.F.; et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **2018**, *50*, 668–681. [CrossRef] [PubMed]
5. Cano-Gamez, E.; Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front. Genet.* **2020**, *11*, 424. [CrossRef] [PubMed]
6. Umans, B.D.; Battle, A.; Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends Genet.* **2020**, *37*, 109–124. [CrossRef] [PubMed]
7. Hu, H.; Miao, Y.-R.; Jia, L.-H.; Yu, Q.-Y.; Zhang, Q.; Guo, A.-Y. AnimalTFDB 3.0: A comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **2018**, *47*, D33–D38. [CrossRef] [PubMed]

8.   Perdomo-Sabogal, Á.; Nowick, K. Genetic Variation in Human Gene Regulatory Factors Uncovers Regulatory Roles in Local Adaptation and Disease. *Genome Biol. Evol.* **2019**, *11*, 2178–2193. [CrossRef]

9.   Buniello, A.; MacArthur, J.A.L.; Cerezo, M.; Harris, L.W.; Hayhurst, J.; Malangone, C.; McMahon, A.; Morales, J.; Mountjoy, E.; Sollis, E.; et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **2019**, *47*, D1005–D1012. [CrossRef]

10.   Taylor, K.E.; Ansel, K.M.; Marson, A.; Criswell, L.A.; Farh, K.K.-H. PICS2: Next-generation fine mapping via probabilistic identification of causal SNPs. *Bioinformatics* **2021**, *37*, 3004–3007. [CrossRef]

11.   McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.S.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122. [CrossRef]

12.   Rentzsch, P.; Witten, D.; Cooper, G.M.; Shendure, J.; Kircher, M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **2019**, *47*, D886–D894. [CrossRef]

13.   The ENCODE Project Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **2012**, *489*, 57–74. [CrossRef] [PubMed]

14.   Oscanoa, J.; Sivapalan, L.; Gadaleta, E.; Ullah, A.Z.D.; Lemoine, N.R.; Chelala, C. SNPnexus: A web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res.* **2020**, *48*, W185–W192. [CrossRef] [PubMed]

15.   Ghoussaini, M.; Mountjoy, E.; Carmona, M.; Peat, G.; Schmidt, E.M.; Hercules, A.; Fumis, L.; Miranda, A.; Carvalho-Silva, D.; Buniello, A.; et al. Open Targets Genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **2020**, *49*, D1311–D1320. [CrossRef] [PubMed]

16.   Hormozdiari, F.; van de Bunt, M.; Segrè, A.V.; Li, X.; Joo, J.W.J.; Bilow, M.; Sul, J.H.; Sankararaman, S.; Pasaniuc, B.; Eskin, E. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **2016**, *99*, 1245–1260. [CrossRef] [PubMed]

17.   Landt, S.G.; Marinov, G.K.; Kundaje, A.; Kheradpour, P.; Pauli, F.; Batzoglou, S.; Bernstein, B.E.; Bickel, P.; Brown, J.B.; Cayting, P. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **2012**, *22*, 1813–1831. [CrossRef] [PubMed]

18.   Jayaram, N.; Usvyat, D.; Martin, A.C.R. Evaluating tools for transcription factor binding site prediction. *BMC Bioinform.* **2016**, *17*, 547. [CrossRef] [PubMed]

19.   GTEx Portal. Available online: https://www.gtexportal.org/home/datasets (accessed on 22 December 2021).

20.   Fornes, O.; Castro-Mondragon, J.A.; Khan, A.; Van Der Lee, R.; Zhang, X.; Richmond, P.A.; Modi, B.P.; Correard, S.; Gheorghe, M.; Baranašić, D.; et al. JASPAR 2020: Update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **2020**, *48*, D87–D92. [CrossRef] [PubMed]

21.   Kulakovskiy, I.V.; Vorontsov, I.E.; Yevshin, I.S.; Sharipov, R.N.; Fedorova, A.D.; Rumynskiy, E.I.; Medvedeva, Y.A.; Magana-Mora, A.; Bajic, V.B.; Papatsenko, D.A.; et al. HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **2017**, *46*, D252–D259. [CrossRef] [PubMed]

22.   Sebastian, A.; Contreras-Moreira, B. footprintDB: A database of transcription factors with annotated cis elements and binding interfaces. *Bioinformatics* **2013**, *30*, 258–265. [CrossRef] [PubMed]

23.   Xie, Z.; Hu, S.; Blackshaw, S.; Zhu, H.; Qian, J. hPDI: A database of experimental human protein-DNA interactions. *Bioinformatics* **2009**, *26*, 287–289. [CrossRef] [PubMed]

24.   Huang, D.; Wang, Z.; Zhou, Y.; Liang, Q.; Sham, P.C.; Yao, H.; Li, M.J. vSampler: Fast and annotation-based matched variant sampling tool. *Bioinformatics* **2020**, *37*, 1915–1917. [CrossRef] [PubMed]

25.   Ernst, J.; Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **2017**, *12*, 2478–2492. [CrossRef] [PubMed]

26.   Annotation of the non-coding genome. *Nature* **2015**. [CrossRef]

27.   Zhang, Q.; Liu, W.; Zhang, H.-M.; Xie, G.-Y.; Miao, Y.-R.; Xia, M.; Guo, A.-Y. hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. *Genom. Proteom. Bioinform.* **2020**, *18*, 120–128. [CrossRef] [PubMed]

28.   Raudvere, U.; Kolberg, L.; Kuzmin, I.; Arak, T.; Adler, P.; Peterson, H.; Vilo, J. g: Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **2019**, *47*, W191–W198. [CrossRef] [PubMed]

29.   Piñero, J.; Ramírez-Anguita, J.M.; Saüch-Pitarch, J.; Ronzano, F.; Centeno, E.; Sanz, F.; Furlong, L.I. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **2019**, *48*, D845–D855. [CrossRef] [PubMed]

30.   Disgenet Plus. Available online: https://beta.disgenetplus.com/ (accessed on 21 December 2021).

31.   Shadrina, M.; Bondarenko, E.A.; Slominsky, P.A. Genetics Factors in Major Depression Disease. *Front. Psychiatry* **2018**, *9*, 334. [CrossRef]

32.   McIntosh, A.M.; Sullivan, P.F.; Lewis, C.M. Uncovering the Genetic Architecture of Major Depression. *Neuron* **2019**, *102*, 91–103. [CrossRef]

33.   Zhang, K.; Lin, W.; Zhang, J.; Zhao, Y.; Wang, X.; Zhao, M. Effect of Toll-like receptor 4 on depressive-like behaviors induced by chronic social defeat stress. *Brain Behav.* **2020**, *10*, e01525. [CrossRef]

34.   Keyes, K.; Agnew-Blais, J.; Roberts, A.L.; Hamilton, A.; De Vivo, I.; Ranu, H.; Koenen, K. The role of allelic variation in estrogen receptor genes and major depression in the Nurses Health Study. *Soc. Psychiatry* **2015**, *50*, 1893–1904. [CrossRef]

35.   Mossakowska-Wójcik, J.; Orzechowska, A.; Talarowska, M.; Szemraj, J.; Gałecki, P. The importance of TCF4 gene in the etiology of recurrent depressive disorders. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2018**, *80*, 304–308. [CrossRef] [PubMed]

36. Berrio, A.T.; Lopez, J.P.; Bagot, R.C.; Nouel, D.; Bo, G.D.; Cuesta, S.; Zhu, L.; Manitt, C.; Eng, C.; Cooper, H.M.; et al. DCC Confers Susceptibility to Depression-like Behaviors in Humans and Mice and Is Regulated by miR-218. *Biol. Psychiatry* **2016**, *81*, 306–315. [CrossRef] [PubMed]

37. Hyde, C.L.; Nagle, M.W.; Tian, C.; Chen, X.; Paciga, S.A.; Wendland, J.R.; Tung, J.Y.; Hinds, D.A.; Perlis, R.H.; Winslow, A.R. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **2016**, *48*, 1031–1036. [CrossRef] [PubMed]

38. Wu, Q.; Tang, W.; Luo, Z.; Li, Y.; Shu, Y.; Yue, Z.; Xiao, B.; Feng, L. DISC1 Regulates the Proliferation and Migration of Mouse Neural Stem/Progenitor Cells through Pax5, Sox2, Dll1 and Neurog2. *Front. Cell. Neurosci.* **2017**, *11*, 261. [CrossRef] [PubMed]

39. Rudzinskas, S.; Hoffman, J.F.; Martinez, P.; Rubinow, D.R.; Schmidt, P.J.; Goldman, D. In vitro model of perimenopausal depression implicates steroid metabolic and proinflammatory genes. *Mol. Psychiatry* **2020**, *26*, 3266–3276. [CrossRef]

40. Qin, Z.; Ren, F.; Xu, X.; Ren, Y.; Li, H.; Wang, Y.; Zhai, Y.; Chang, Z. ZNF536, a Novel Zinc Finger Protein Specifically Expressed in the Brain, Negatively Regulates Neuron Differentiation by Repressing Retinoic Acid-Induced Gene Transcription. *Mol. Cell. Biol.* **2009**, *29*, 3633–3643. [CrossRef]

41. Laifenfeld, D.; Klein, E.; Ben-Shachar, D. Norepinephrine alters the expression of genes involved in neuronal sprouting and differentiation: Relevance for major depression and antidepressant mechanisms. *J. Neurochem.* **2002**, *83*, 1054–1064. [CrossRef]

42. Lanshakov, D.A.; Sukhareva, E.V.; Bulygina, V.V.; Bannova, A.V.; Shaburova, E.V.; Kalinina, T.S. Single neonatal dexamethasone administration has long-lasting outcome on depressive-like behaviour, Bdnf, Nt-3, p75ngfr and sorting receptors (SorCS1-3) stress reactive expression. *Sci. Rep.* **2021**, *11*, 8092. [CrossRef]

43. Sanna, M.D.; Quattrone, A.; Galeotti, N. Antidepressant-like actions by silencing of neuronal ELAV-like RNA-binding proteins HuB and HuC in a model of depression in male mice. *Neuropharmacology* **2018**, *135*, 444–454. [CrossRef]

44. Beilina, A.; Rudenko, I.N.; Kaganovich, A.; Civiero, L.; Chau, H.; Kalia, S.K.; Kalia, L.V.; Lobbestael, E.; Chia, R.; Ndukwe, K.; et al. Unbiased screen for interactors of leucine-rich repeat kinase 2 supports a common pathway for sporadic and familial Parkinson disease. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2626–2631. [CrossRef]

45. Dzhala, V.I.; Talos, D.M.; Sdrulla, D.A.; Brumback, A.; Mathews, G.C.; Benke, T.; Delpire, E.; Jensen, F.E.; Staley, K.J. NKCC1 transporter facilitates seizures in the developing brain. *Nat. Med.* **2005**, *11*, 1205–1213. [CrossRef] [PubMed]

46. Bujalka, H.; Koenning, M.; Jackson, S.; Perreau, V.M.; Pope, B.; Hay, C.M.; Mitew, S.; Hill, A.F.; Lu, Q.R.; Wegner, M.; et al. MYRF Is a Membrane-Associated Transcription Factor That Autoproteolytically Cleaves to Directly Activate Myelin Genes. *PLoS Biol.* **2013**, *11*, e1001625. [CrossRef] [PubMed]

47. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **2019**, *179*, 1469–1482.e11. [CrossRef] [PubMed]

48. Yao, X.; Glessner, J.T.; Li, J.; Qi, X.; Hou, X.; Zhu, C.; Li, X.; March, M.E.; Yang, L.; Mentch, F.D.; et al. Integrative analysis of genome-wide association studies identifies novel loci associated with neuropsychiatric disorders. *Transl. Psychiatry* **2021**, *11*, 69. [CrossRef] [PubMed]

49. Garcia, W.S.; Rocha-Acevedo, M.; Ramirez-Navarro, L.; Mbouamboua, Y.; Thieffry, D.; Thomas-Chollier, M.; Contreras-Moreira, B.; van Helden, J.; Medina-Rivera, A. RSAT variation-tools: An accessible and flexible framework to predict the impact of regulatory variants on transcription factor binding. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 1415–1428. [CrossRef] [PubMed]

50. Verheul, T.C.J.; Van Hijfte, L.; Perenthaler, E.; Barakat, T.S. The Why of YY1: Mechanisms of Transcriptional Regulation by Yin Yang 1. *Front. Cell Dev. Biol.* **2020**, *8*, 592164. [CrossRef]

51. Byts, N.; Sharma, S.; Laurila, J.; Paudel, P.; Miinalainen, I.; Ronkainen, V.-P.; Hinttala, R.; Törnquist, K.; Koivunen, P.; Myllyharju, J. Transmembrane Prolyl 4-Hydroxylase is a Novel Regulator of Calcium Signaling in Astrocytes. *ENeuro* **2020**, *8*, 1–23. [CrossRef]

52. Leinonen, H.; Koivisto, H.; Lipponen, H.-R.; Matilainen, A.; Salo, A.M.; Dimova, E.Y.; Hämäläinen, E.; Stavén, S.; Miettinen, P.; Myllyharju, J.; et al. Null mutation in P4h-tm leads to decreased fear and anxiety and increased social behavior in mice. *Neuropharmacology* **2019**, *153*, 63–72. [CrossRef]

53. Bhalala, O.G.; Nath, A.P.; Inouye, M.; Sibley, C.R. UK Brain Expression Consortium Identification of expression quantitative trait loci associated with schizophrenia and affective disorders in normal brain tissue. *PLoS Genet.* **2018**, *14*, e1007607. [CrossRef]

54. Li, S.; Li, X.; Liu, J.; Huo, Y.; Li, L.; Wang, J.; Luo, X.-J. Functional variants fine-mapping and gene function characterization provide insights into the role of ZNF323 in schizophrenia pathogenesis. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **2021**, *186*, 28–39. [CrossRef]

55. Roksana, Z. Transcription Factors in Schizophrenia: A Current View of Genetic Aspects. *Sci. J. Genet. Gene Ther.* **2016**, *2*, 17–21. [CrossRef]

56. Li, X.; Su, X.; Liu, J.; Li, H.; Li, M.; Li, W.; Luo, X.-J. Transcriptome-wide association study identifies new susceptibility genes and pathways for depression. *Transl. Psychiatry* **2021**, *11*, 306. [CrossRef] [PubMed]

57. Zhong, J.; Li, S.; Zeng, W.; Li, X.; Gu, C.; Liu, J.; Luo, X.-J. Integration of GWAS and brain eQTL identifies FLOT1 as a risk gene for major depressive disorder. *Neuropsychopharmacology* **2019**, *44*, 1542–1551. [CrossRef]

58. Santos-Terra, J.; Deckmann, I.; Fontes-Dutra, M.; Schwingel, G.B.; Bambini-Junior, V.; Gottfried, C. Transcription factors in neurodevelopmental and associated psychiatric disorders: A potential convergence for genetic and environmental risk factors. *Int. J. Dev. Neurosci.* **2021**, *81*, 545–578. [CrossRef] [PubMed]

59. Burt, C.; Munafò, M. Has GWAS lost its status as a paragon of open science? *PLoS Biol.* **2021**, *19*, e3001242. [CrossRef] [PubMed]

60. Lee, B.; Yao, X.; Shen, L. Integrative analysis of summary data from GWAS and eQTL studies implicates genes differentially expressed in Alzheimer's disease. *BMC Genom.* **2022**, *23*, 414. [CrossRef] [PubMed]

61. Brooks-Warburton, J.; Modos, D.; Sudhakar, P.; Madgwick, M.; Thomas, J.P.; Bohar, B.; Fazekas, D.; Zoufir, A.; Kapuy, O.; Szalay-Beko, M.; et al. A systems genomics approach to uncover patient-specific pathogenic pathways and proteins in ulcerative colitis. *Nat. Commun.* **2022**, *13*, 2299. [CrossRef]

62. O'Brien, H.E.; Hannon, E.; Hill, M.J.; Toste, C.C.; Robertson, M.J.; Morgan, J.E.; McLaughlin, G.; Lewis, C.M.; Schalkwyk, L.C.; Hall, L.S.; et al. Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol.* **2018**, *19*, 194. [CrossRef]

63. Hare, B.D.; Duman, R.S. Prefrontal cortex circuits in depression and anxiety: Contribution of discrete neuronal populations and target regions. *Mol. Psychiatry* **2020**, *25*, 2742–2758. [CrossRef]

64. Amare, A.T.; Vaez, A.; Hsu, Y.-H.; Direk, N.; Kamali, Z.; Howard, D.; McIntosh, A.; Tiemeier, H.; Bültmann, U.; Snieder, H.; et al. Bivariate genome-wide association analyses of the broad depression phenotype combined with major depressive disorder, bipolar disorder or schizophrenia reveal eight novel genetic loci for depression. *Mol. Psychiatry* **2019**, *25*, 1420–1429. [CrossRef]

65. Pozzi, D.; Rasile, M.; Corradini, I.; Matteoli, M. Environmental regulation of the chloride transporter KCC2: Switching inflammation off to switch the GABA on? *Transl. Psychiatry* **2020**, *10*, 349. [CrossRef] [PubMed]

66. Saponaro, F.; Sestito, S.; Runfola, M.; Rapposelli, S.; Chiellini, G. Selective Thyroid Hormone Receptor-Beta (TRβ) Agonists: New Perspectives for the Treatment of Metabolic and Neurodegenerative Disorders. *Front. Med.* **2020**, *7*, 331. [CrossRef] [PubMed]

67. Zhou, B.; Zhu, Z.; Ransom, B.R.; Tong, X. Oligodendrocyte lineage cells and depression. *Mol. Psychiatry* **2020**, *26*, 103–117. [CrossRef] [PubMed]

68. Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The Human Transcription Factors. *Cell* **2018**, *172*, 650–665. [CrossRef] [PubMed]

69. Andersen, M.C.; Engström, P.G.; Lithwick, S.; Arenillas, D.; Eriksson, P.; Lenhard, B.; Wasserman, W.W.; Odeberg, J. In Silico Detection of Sequence Variations Modifying Transcriptional Regulation. *PLoS Comput. Biol.* **2008**, *4*, e5. [CrossRef] [PubMed]

70. Dennis, D.J.; Han, S.; Schuurmans, C. bHLH transcription factors in neural development, disease, and reprogramming. *Brain Res.* **2019**, *1705*, 48–65. [CrossRef]

71. Rada-Iglesias, A.; Ameur, A.; Kapranov, P.; Enroth, S.; Komorowski, J.; Gingeras, T.R.; Wadelius, C. Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders. *Genome Res.* **2008**, *18*, 380–392. [CrossRef]

72. Sertbaş, M.; Ülgen, K.; Çakır, T. Systematic analysis of transcription-level effects of neurodegenerative diseases on human brain metabolism by a newly reconstructed brain-specific metabolic network. *FEBS Open Bio* **2014**, *4*, 542–553. [CrossRef]

73. Grubert, F.; Srivas, R.; Spacek, D.V.; Kasowski, M.; Ruiz-Velasco, M.; Sinnott-Armstrong, N.; Greenside, P.; Narasimha, A.; Liu, Q.; Geller, B.; et al. Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **2020**, *583*, 737–743. [CrossRef]

74. Brodie, A.; Azaria, J.R.; Ofran, Y. How far from the SNP may the causative genes be? *Nucleic Acids Res.* **2016**, *44*, 6046–6054. [CrossRef]

75. Shi, Y.; Wang, Q.; Song, R.; Kong, Y.; Zhang, Z. Non-coding RNAs in depression: Promising diagnostic and therapeutic biomarkers. *EBioMedicine* **2021**, *71*, 103569. [CrossRef] [PubMed]

76. Żurawek, D.; Turecki, G. The miRNome of Depression. *Int. J. Mol. Sci.* **2021**, *22*, 11312. [CrossRef] [PubMed]

77. Bian, Z.; Li, H.; Liu, Y.; Cao, Y.; Kang, Y.; Yu, Y.; Zhang, F.; Li, C.; Kang, Y.; Wang, F. The Association Between Hypoxia Improvement and Electroconvulsive Therapy for Major Depressive Disorder. *Neuropsychiatr. Dis. Treat.* **2021**, *17*, 2987–2994. [CrossRef] [PubMed]

78. Li, G.; Zhao, M.; Cheng, X.; Zhao, T.; Feng, Z.; Zhao, Y.; Fan, M.; Zhu, L. FG-4592 Improves Depressive-Like Behaviors through HIF-1-Mediated Neurogenesis and Synapse Plasticity in Rats. *Neurotherapeutics* **2019**, *17*, 664–675. [CrossRef] [PubMed]

79. Ding, F.-S.; Cheng, X.; Zhao, T.; Zhao, Y.; Zhang, G.-B.; Wu, H.-T.; Zhu, L.-L.; Wu, K.-W. Intermittent hypoxic preconditioning relieves fear and anxiety behavior in post-traumatic stress model mice. *Sheng Li Xue Bao* **2019**, *71*, 537–546.

80. Shibata, T.; Yamagata, H.; Uchida, S.; Otsuki, K.; Hobara, T.; Higuchi, F.; Abe, N.; Watanabe, Y. The alteration of hypoxia inducible factor-1 (HIF-1) and its target genes in mood disorder patients. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2013**, *43*, 222–229. [CrossRef] [PubMed]

81. Kang, I.; Kondo, D.; Kim, J.; Lyoo, I.K.; Yurgelun-Todd, D.; Hwang, J.; Renshaw, P.F. Elevating the level of hypoxia inducible factor may be a new potential target for the treatment of depression. *Med. Hypotheses* **2020**, *146*, 110398. [CrossRef]

82. Szczepocka, E.; Wysokiński, A. Red Blood Cells Parameters in Patients with Acute Schizophrenia, Unipolar Depression and Bipolar Disorder. *Psychiatr. Danub.* **2018**, *30*, 323–330. [CrossRef]