Table of Contents

[Supplementary information 2](#_Toc152086328)

[1.1. Local genetic correlation analyses 2](#_Toc152086329)

[1.2. Functional annotation of genetically correlated regions 2](#_Toc152086330)

[1.3. Colocalisation analyses 2](#_Toc152086331)

[1.4. Functional annotation of 95% credible sets of shared causal variants 3](#_Toc152086332)

[References 5](#_Toc152086333)

# Supplementary information

## **Local genetic correlation analyses**

First, we generated 2,495 semi-independent genomic regions of ~1Mb by recursively splitting chromosomes while minimising linkage disequilibrium (LD) between them, following the approach reported by Werme et al. (2022). For this process, we used the 1000 Genomes Project, phase 3 (European ancestry, GRCh37/hg19 build) as our reference genome and the Local Analysis of [co]Variant Association (LAVA) partitioning algorithm (<https://github.com/cadeleeuw/lava-partitioning>; de Leeuw (2021)). Second, univariate analyses were conducted using LAVA to assess single-nucleotide polymorphism (SNP)-based heritability (h2SNP) at the level of each individual genomic region for each individual IR-related and neuropsychiatric phenotype. We selected only those genomic regions exhibiting significant local h2SNP values (univariate p-value<1x10-4) for the subsequent bivariate analyses. Third, local genetic correlations were estimated between IR-related and neuropsychiatric phenotype pairs at each individual genomic region passing univariate analyses. To account for any potential sample overlap in the input summary statistics, we suppled the intercepts from the bivariate cross-trait LD Score (LDSC) analyses to LAVA as these represent an estimate of the sampling correlation between the data sets (Bulik-Sullivan et al., 2015).

## **Functional annotation of genetically correlated regions**

For analysing the gene expression profiles, data from the Genotype-Tissue Expression (GTEx) project (Consortium, 2015) and the BrainSpan Atlas of the Human Brain were used, the latter including information from 11 general developmental stages and 29 distinct ages of brain samples (Kang et al., 2011; Miller et al., 2014). This analysis was complemented by an assessment of the tissue specificity of these genes, through the computation of the tissue-specificity index (tau) and identification of differentially expressed genes (DEGs) across tissues, offering further insight into their potential functional significance (Watanabe et al., 2017).

To investigate the involvement of the mapped genes in various biological pathways and processes, gene set enrichment analyses were carried out using hypergeometric tests. Enrichment analyses were conducted considering hallmark gene sets, which represent well-defined biological states or processes; ontology gene sets, which comprise genes annotated by the same ontology term; immunologic signature gene sets, representing cellular states and perturbations within the immune system; canonical pathway gene sets, which are collections of genes that work together to perform a specific biological function and are well-documented in scientific literature; and regulatory target gene sets, based on microRNA seed sequences and predicted transcription factor binding sites (Watanabe et al., 2017). Ontology gene sets were sourced from the Gene Ontology (GO) database, which collects predefined gene sets related to biological processes, cellular components, and molecular functions (The Gene Ontology, 2019). Canonical pathway gene sets were sourced from the BioCarta, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome databases (Fabregat et al., 2016; Kanehisa and Goto, 2000; Nishimura, 2001).

## **Colocalisation analyses**

The coloc method utilises a Bayesian framework to estimate the posterior probabilities (PPs) of five different hypotheses (i.e., H0: no association with either trait; H1: association with trait 1, not with trait 2; H2: association with trait 2, not with trait 1; H3: association with trait 1 and trait 2, two independent SNPs; H4: association with trait 1 and trait 2, one shared SNP) representing the shared or distinct causal variants between the two phenotypes (i.e., insulin resistance-related conditions and neuropsychiatric disorders). By integrating the summary statistics from two respective GWASs, the method calculates the PPs for each hypothesis and assesses the likelihood of colocalisation. A high PP for hypothesis H4 (PP.H4) indicates strong evidence for colocalisation, whereas a low PPH4 suggests distinct causal variants. These PPs are computed by taking into account the SNP-trait association statistics (e.g., p-values or Z-scores) and the LD structure among the SNPs. The LD structure of the correlated genomic regions was determined using PLINK 1.9 (Purcell et al., 2007). The coloc analysis was executed using the coloc.abf() function with default parameters (Giambartolomei et al., 2014). Genomic loci harbouring plausible causal variants affecting both traits were identified based on the PPs resulting from the colocalisation analyses (Huang et al., 2020; Schmiedel et al., 2021). In particular, genetically correlated loci were selected only if the cumulative PPs for hypotheses H3 and H4 exceeded 0.8, thereby ensuring a substantial degree of confidence that the identified loci host at least one causal variant for either or both traits under investigation. Moreover, we imposed a ratio constraint wherein the PP.H4 needed to be at least five-fold greater than PP.H3. This second criterion enriched our selection for genomic loci wherein the evidence supports a shared causal variant influencing both traits, as opposed to distinct causal variants for each trait (Huang et al., 2020; Schmiedel et al., 2021). After identifying genomic loci that met our pre-established condition (i.e., PP.H3+PP.H4≥0.8-PP.H4/PP.H3≥5), we pinpointed the most likely causal variant for the shared signal based on either a PP.H4 of the individual SNP (SNP.PP.H4)≥0.5 or the calculation of credible sets of variants, which are pivotal for narrowing down potential causal variants. Using Bayesian methods, we derived the minimal set of SNPs that may encompass all causal variants with 95% probability (α). To construct our α-credible sets, we ranked the SNPs within each locus according to their SNP.PP.H4 values, from highest to lowest. We then computed the cumulative sum of SNP.PP.H4s until it equalled or exceeded α. This procedure ensures that the resulting set of SNPs is the smallest subset accounting for at least 95% of the PP of containing the causal shared variant (Schaid et al., 2018). By applying this rigorous method, we were able to reduce the list of potentially causal variants, while retaining high confidence in the likelihood of these sets containing the true causal variants. Subsequent to the identification of the most likely causal variants, we employed LocusZoom to visually represent the genomic context and LD patterns surrounding these GWAS associated variants. This visualisation was conducted separately for each phenotype and then combined to illustrate the patterns for both phenotypes collectively (Pruim et al., 2010).

## **Functional annotation of 95% credible sets of shared causal variants**

Our annotation process involved an in-depth analysis using SNPnexus to characterise the functional significance of likely causal variants. Combined Annotation Dependent Depletion (CADD) scores integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated mutations, thereby predicting the functional impact of variants on protein-coding sequences (Kircher et al., 2014). For the purpose of our analysis, a CADD-PHRED score≥20 served as our threshold for identifying potentially deleterious variants falling within the top 1% of most deleterious substitutions in the human genome, while a CADD-PHRED≥10, which corresponds to the top 10% most deleterious substitutions, was also used as a less stringent threshold. The PHRED-like score is a transformation of the raw score that ranks the variant relative to all possible 8.6 billion variants in the human genome. Complementing the CADD scores, we analysed the transcription factor binding sites (TFBSs) related to our most likely causal SNPs (Wingender et al., 2001). TFBSs are specific DNA regions where transcription factors bind, influencing the regulation of gene expression. Moreover, the analysis incorporated information from miRBase, a comprehensive database of miRNA sequences and annotation. MicroRNAs (miRNAs) are pivotal non-coding RNAs that influence gene expression. We also considered the localisation of variants in predicted CpG (cytosine-phosphate-guanine) islands, which are genomic regions rich in cytosine-guanine dinucleotides, often linked to gene promoters. Other databases like TarBase and TargetScan also contributed to our analysis by providing experimentally validated data on miRNA targets and miRNA biological target predictions, respectively (Lewis et al., 2005; Papadopoulos et al., 2009). Other regulatory elements in the human genome were annotated using RegBuild, a resource from Ensembl (Zerbino et al., 2015). Additionally, the Roadmap Epigenomics Project data were considered to enrich our analysis by providing comprehensive epigenetic maps indicative of chromatin states that further elucidate potential functional elements in the regions of interest (Romanoski et al., 2015). The annotation process utilised the GRCh37/hg19 assembly of the human genome from Ensembl as the reference, supplementing it with additional databases including RefSeq, UCSC, GENCODE, and VEGA to ensure a thorough annotation (Oscanoa et al., 2020). Additionally, within the SNPnexus platform, pathway analysis was conducted focusing on 95% credible set variants. Within each query set, SNPnexus links genes to associated biological Reactome pathways. This process involves mapping variants in each credible set to their respective genes. Subsequently, the number of genes in the Reactome universe, as well as the number of genes in each pathway was determined. These parameters serve as inputs for Fisher’s Exact tests, which assess whether the number of affected genes in each set significantly deviates from what would be expected by random chance for each Reactome pathway. This approach allows us to identify pathways that are over-represented within credible set variants, providing further insights into potential causal mechanisms underlying neuropsychiatric-insulin resistance multimorbidity (Oscanoa et al., 2020).

# References

Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics, C., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., 2015. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47, 291-295.

Consortium, G.T., 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648-660.

de Leeuw, C.A., 2021. LAVA partitioning algorithm (v1.0.0). Zenodo.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2016. The Reactome pathway Knowledgebase. Nucleic Acids Res 44, D481-487.

Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., Plagnol, V., 2014. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10, e1004383.

Huang, Q.Q., Tang, H.H.F., Teo, S.M., Mok, D., Ritchie, S.C., Nath, A.P., Brozynska, M., Salim, A., Bakshi, A., Holt, B.J., Khor, C.C., Sly, P.D., Holt, P.G., Holt, K.E., Inouye, M., 2020. Neonatal genetics of gene expression reveal potential origins of autoimmune and allergic disease risk. Nat Commun 11, 3761.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28, 27-30.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M.B., Krsnik, Z., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S.N., Vortmeyer, A., Weinberger, D.R., Mane, S., Hyde, T.M., Huttner, A., Reimers, M., Kleinman, J.E., Sestan, N., 2011. Spatio-temporal transcriptome of the human brain. Nature 478, 483-489.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., Shendure, J., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46, 310-315.

Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15-20.

Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., Arnold, J.M., Bennet, C., Bertagnolli, D., Brouner, K., Butler, S., Caldejon, S., Carey, A., Cuhaciyan, C., Dalley, R.A., Dee, N., Dolbeare, T.A., Facer, B.A., Feng, D., Fliss, T.P., Gee, G., Goldy, J., Gourley, L., Gregor, B.W., Gu, G., Howard, R.E., Jochim, J.M., Kuan, C.L., Lau, C., Lee, C.K., Lee, F., Lemon, T.A., Lesnar, P., McMurray, B., Mastan, N., Mosqueda, N., Naluai-Cecchini, T., Ngo, N.K., Nyhus, J., Oldre, A., Olson, E., Parente, J., Parker, P.D., Parry, S.E., Stevens, A., Pletikos, M., Reding, M., Roll, K., Sandman, D., Sarreal, M., Shapouri, S., Shapovalova, N.V., Shen, E.H., Sjoquist, N., Slaughterbeck, C.R., Smith, M., Sodt, A.J., Williams, D., Zollei, L., Fischl, B., Gerstein, M.B., Geschwind, D.H., Glass, I.A., Hawrylycz, M.J., Hevner, R.F., Huang, H., Jones, A.R., Knowles, J.A., Levitt, P., Phillips, J.W., Sestan, N., Wohnoutka, P., Dang, C., Bernard, A., Hohmann, J.G., Lein, E.S., 2014. Transcriptional landscape of the prenatal human brain. Nature 508, 199-206.

Nishimura, D., 2001. BioCarta. Biotech Software & Internet Report: The Computer Software Journal for Scient 2, 117-120.

Oscanoa, J., Sivapalan, L., Gadaleta, E., Dayem Ullah, A.Z., Lemoine, N.R., Chelala, C., 2020. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic Acids Res 48, W185-W192.

Papadopoulos, G.L., Reczko, M., Simossis, V.A., Sethupathy, P., Hatzigeorgiou, A.G., 2009. The database of experimentally supported targets: a functional update of TarBase. Nucleic Acids Res 37, D155-158.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J., 2010. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336-2337.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81, 559-575.

Romanoski, C.E., Glass, C.K., Stunnenberg, H.G., Wilson, L., Almouzni, G., 2015. Roadmap for regulation. Nature 518, 314-316.

Schaid, D.J., Chen, W., Larson, N.B., 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 19, 491-504.

Schmiedel, B.J., Rocha, J., Gonzalez-Colin, C., Bhattacharyya, S., Madrigal, A., Ottensmeier, C.H., Ay, F., Chandra, V., Vijayanand, P., 2021. COVID-19 genetic risk variants are associated with expression of multiple genes in diverse immune cell types. Nat Commun 12, 6760.

The Gene Ontology, C., 2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res 47, D330-D338.

Watanabe, K., Taskesen, E., van Bochoven, A., Posthuma, D., 2017. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 8, 1826.

Werme, J., van der Sluis, S., Posthuma, D., de Leeuw, C.A., 2022. An integrated framework for local genetic correlation analysis. Nat Genet 54, 274-282.

Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhauser, R., Pruss, M., Schacherer, F., Thiele, S., Urbach, S., 2001. The TRANSFAC system on gene expression regulation. Nucleic Acids Res 29, 281-283.

Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T., Flicek, P.R., 2015. The ensembl regulatory build. Genome Biol 16, 56.