medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Statistical examination of shared loci in neuropsychiatric diseases using genome-wide association study summary statistics

Thomas P Spargo[1,2,3,*], Lachlan Gilchrist[1,4,5], Guy P Hunt[2,5,6], Richard JB Dobson[2,3,7,8], Petroula Proitsi[1], Ammar Al-Chalabi[1,9], Oliver Pain[1,$], Alfredo Iacoangeli[1,2,3,$,*]

1. Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, UK
2. Department of Biostatistics and Health Informatics, King's College London, London, UK
3. NIHR Maudsley Biomedical Research Centre (BRC) at South London and Maudsley NHS Foundation Trust and King's College London, London, UK
4. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
5. Perron Institute for Neurological and Translational Science, Nedlands, WA 6009, Australia
6. Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Murdoch, WA 6150, Australia
7. Institute of Health Informatics, University College London, London, United Kingdom
8. NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, UK
9. King's College Hospital, Bessemer Road, London, SE5 9RS, UK

[$]co-senior authors

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

*Correspondence should be addressed to alfredo.iacoangeli@kcl.ac.uk and thomas.spargo@kcl.ac.uk

Abstract

Continued methodological advances have enabled numerous statistical approaches for the analysis of summary statistics from genome-wide association studies. Genetic correlation analysis within specific regions enables a new strategy for identifying pleiotropy. Genomic regions with significant 'local' genetic correlations can be investigated further using state-of-the-art methodologies for statistical fine-mapping and variant colocalisation. We explored the utility of a genome-wide local genetic correlation analysis approach for identifying genetic overlaps between the candidate neuropsychiatric disorders, Alzheimer's disease, amyotrophic lateral sclerosis, frontotemporal dementia, Parkinson's disease, and schizophrenia. The correlation analysis identified several associations between traits, the majority of which were loci in the human leukocyte antigen (HLA) region. Colocalisation analysis suggested that disease-implicated variants in these loci often differ between traits and, in one locus, indicated a shared causal variant between amyotrophic lateral sclerosis and Alzheimer's disease. Our study identified candidate loci that might play a role in multiple neuropsychiatric diseases and suggested the role of distinct mechanisms across diseases despite shared loci. The fine-mapping and colocalisation analysis protocol designed for this study has been implemented in a flexible analysis pipeline that produces HTML reports and is available at: https://github.com/ThomasPSpargo/COLOC-reporter.

Keywords: statistical genetics, GWAS, fine-mapping, colocalisation, neurodegeneration genetics, psychiatric disease

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

1. Introduction

The genetic spectrum of neuropsychiatric disease is diverse and various overlaps exist between traits. For instance, genetic pleiotropy between amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) is increasingly recognised, and ALS is genetically correlated with Alzheimer's disease (AD), Parkinson's disease (PD), and schizophrenia[1-3]. Improving understanding of the genetic architecture underlying these complex diseases could facilitate future treatment discovery.

Advances in genomic research techniques have accelerated discovery of genetic variation associated with complex traits. Genome-wide association studies (GWAS), in particular, have enabled population-scale investigations of the genetic basis of human diseases and anthropometric measures[4]. Summary-level results from GWAS are being shared alongside publications with increasing frequency over time[5], and a breadth of approaches now exist for downstream analysis based on summary statistics which can enable their interpretation and provide further biological insight.

Genetic correlation analysis allows estimation of genetic overlap between traits[6-10]. A 'global' genetic correlation approach gives a genome-wide average estimate of this overlap However, genetic relationships between traits can be obscured when correlations in opposing directions cancel out genome-wide[8]. Recent methods allow for a more nuanced analysis, of 'local' genetic correlations partitioned across the genome[8,9]. This stratified approach to genome-wide analysis could prove effective for identifying pleiotropic regions and designing subsequent analyses aiming to identify genetic variation shared between traits.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

A number of methods aim to disentangle causality within associated regions. This is important because the focus on single nucleotide polymorphisms (SNPs), which are markers of genetic variation, in GWAS produces results that can be difficult to interpret, and causal variants are typically unclear. More so, because of linkage disequilibrium (LD), GWAS associations often comprise large sets of highly correlated SNPs spanning large genomic regions. Statistical fine-mapping is a common approach for dissecting complex LD structures and finding variants with implications for a given trait among the tens or hundreds that might be associated in the region[11].

Interpretation of regions associated with multiple traits can also be challenging, since it is often unclear whether these overlaps are driven by the same causal variant. Statistical colocalisation analysis can disentangle association signals across traits to suggest whether the overlaps result from shared or distinct causal genetic factors[12-14]. Traditionally this analysis was restricted by the assumption of at most one causal variant for each trait in the region. However, recent extensions to the method now permit analysis based on univariate fine-mapping results for the traits compared and, therefore, analysis of regions with multiple causal variants.

Accordingly, we conducted genome-wide local genetic correlation analysis across 5 neuropsychiatric traits with recognised phenotypic and genetic overlap[2, 3, 15-17]: AD, ALS, FTD, PD, and schizophrenia. Although previous studies have performed global genetic correlation analyses between various combinations of these traits[1, 2, 18, 19], this is the first to compare them at a genome-wide scale using a local genetic correlation approach. Loci highly

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

correlated between trait pairs were further investigated with univariate fine-mapping and bivariate colocalisation techniques to examine variants driving these associations.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2. Methods

2.1. Sampled GWAS summary statistics

We leveraged publicly-accessible summary statistics from European ancestry GWAS meta-analyses of risk for AD[20], ALS[1], FTD[21], PD[22], and schizophrenia[23]. European ancestry data were selected to avoid LD mismatch between the GWAS sample and reference data from an external European population.

2.2. Procedure

Figure 1 summarises the analysis protocol for this study; further details are provided below.

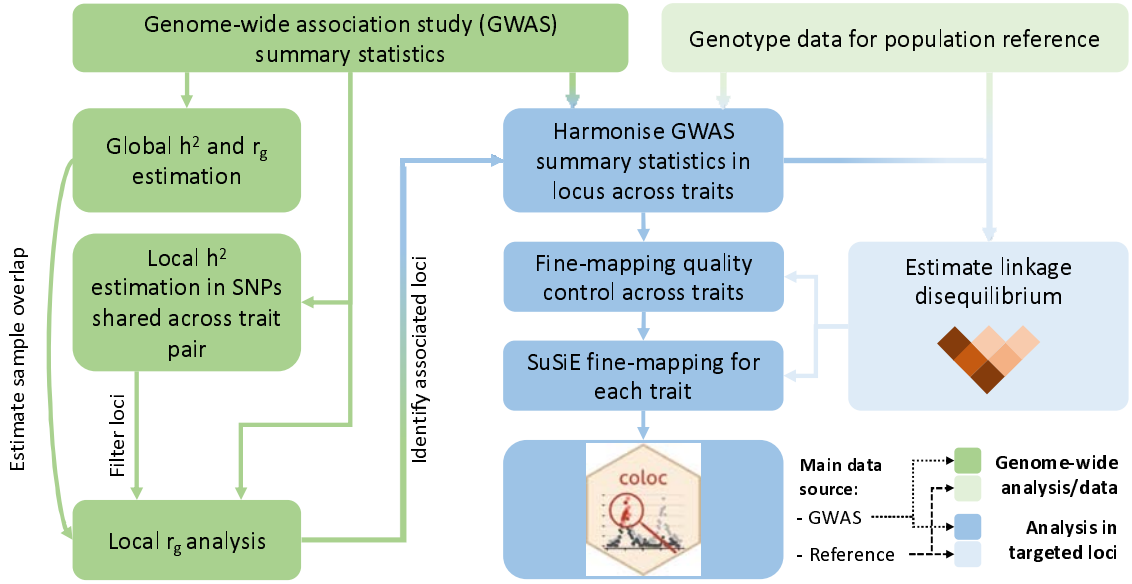

Figure 1. Overview of the analysis procedure for this study

SuSiE (sum of single effects) is a univariate fine-mapping approach implemented within the R package susieR. 'coloc' is an R package for bivariate colocalisation analysis between pairs of traits. h^2 = Heritability, r_g= bivariate genetic correlation. The analysis steps shaded in blue have been implemented within a readily applied analysis pipeline available on GitHub: https://github.com/ThomasPSpargo/COLOC-reporter.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2.2.1. Processing of GWAS summary statistics

A standard data cleaning protocol was applied to each set of summary statistics[24]. We retained only single nucleotide polymorphisms (SNPs), excluding any non-SNP or strand-ambiguous variants. SNPs were filtered to those present within the 1000 Genomes phase 3 (1KG) European ancestry population reference dataset[25] (N = 503). They were matched to the 1KG reference panel by GRCh37 chromosomal position using *bigsnpr* (version 1.11.6)[26], harmonising allele order with the reference and assigning SNP IDs.

If not reported, and where possible, effective sample size (N_{eff}) was calculated from per-SNP case and control sample sizes. When this could not be determined per-SNP, all variants were assigned a single N_{eff}, calculated as a sum of N_{eff} values for each cohort within the GWAS meta-analysis[27].

Further processing was performed where possible, excluding SNPs with imputation INFO <0.9, p-values ≤0 or >1, and N_{eff} >3 standard deviations from the median N_{eff}. We filtered to include only variants with minor allele frequency (MAF) ≥0.005 in both the reference and GWAS samples and excluded SNPs with an absolute MAF difference of >0.2 between the two.

2.2.2. Genome-wide analyses

2.2.2.1. Global heritability and genetic correlations

LDSC (version 1.0.1)[6,7] was applied to estimate genome-wide univariate heritability (h^2) for each trait on the liability scale. The software was also applied to derive 'global' (i.e., genome-wide) genetic correlation estimates between trait pairs and estimate sample

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

overlap from the bivariate intercept. The latter of these outputs was taken forward as an input for the local genetic correlation analysis using LAVA (see 2.2.2.2). Since global genetic correlation analysis across the traits studied here is not novel and associations reported in past studies are congruent across different tools[18], the compatibility between LDSC and LAVA motivated our use of LDSC for this analysis.

These analyses were performed using the HapMap3[28] SNPs and the LD score files provided with the software, calculated in the 1KG European population. No further MAF filter was applied (therefore variants with MAF >0.005 were included) and the other settings were left to their defaults.

2.2.2.2. Local genetic correlation analysis

LAVA (version 0.1.0)[8] was applied to obtain local genetic correlation estimates across 2,495 approximately independent blocks delineating the genome, based on patterns in LD. We used the blocks provided alongside the LAVA software which were derived from the 1KG European cohort. Bivariate intercepts from LDSC were provided to LAVA to estimate sample overlap between trait pairs.

LAVA was the most appropriate local genetic correlation approach for this study for several reasons[8]. First, unlike SUPERGNOVA[9] and rho-HESS[10], LAVA makes specific accommodations for analysis of binary traits. Second, other tools focus on bivariate correlation between traits whilst LAVA offers this alongside multivariate tests such as multiple regression and partial correlation, enabling rigorous testing of pleiotropic effects. Lastly, LAVA is shown to provide results which are less biased than those from other tools.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

In accordance with prior studies, genetic correlation analysis was performed following an initial filtering step. Univariate heritability was estimated for each genomic block across SNPs in-common between a pair of traits, and only loci with local h^2 p-values below a threshold of 2.004×10^{-5} (0.05/2,495) in both traits continued to the bivariate analysis. This step ensures that univariate heritability is sufficient in both traits for a robust correlation estimate.

2.2.3. Targeted genetic analyses

2.2.3.1. Fine-mapping and colocalisation analysis

Statistical fine-mapping and colocalisation techniques were applied to further analyse associations between trait pairs in regions where the false discovery rate (FDR) adjusted p-value of local genetic correlation analysis was below 0.05 (after adjusting for all bivariate comparisons performed). Additional analysis was conducted at loci where significant correlations occurred between two trait pairs but not between the final pairwise comparison across the three implicated traits.

Fine-mapping was performed with *susieR* (v0.12.27)[11,29], which implements the 'sum of single effects' (SuSiE) model to represent statistical evidence of causal genetic variation within 'credible sets' and per-SNP posterior inclusion probabilities (PIPs). A 95% credible set indicates 95% certainty that at least one SNP included within the set has a causal association with the phenotype and higher PIPs indicate a greater posterior probability of being a causal variant within a credible set. Multiple credible sets are identified when the data suggest more than one independent causal signal.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Colocalisation analysis was implemented with *coloc* (v5.1.0.1)[12,13,30], which calculates posterior probabilities that a causal variant exists for neither, one, or both of two compared traits, testing also whether evidence for a causal variant in both traits suggests a shared variant (i.e., hypothesis 4 (H4); colocalisation) or independent signals (Hypothesis 3 (H3)). Colocalisation analyses can be performed across all variants sampled in a region, under an assumption of at most one variant implicated per trait. It can also be performed using variants attributed to pairs of credible sets from SuSiE, relaxing the single variant assumption[12]. When evidence of a shared variant is found, the individual SNPs with the highest posterior probability of being that variant can be assessed. With a 95% confidence threshold, these are termed 95% credible SNPs.

Analysis pipeline

We conducted colocalisation and fine-mapping analysis within an open-access pipeline developed for this study using R (v4.2.2)[31]: https://github.com/ThomasPSpargo/COLOC-reporter.

Briefly, in this workflow (see Figure 1), GWAS summary statistics are harmonised across analysed traits for a specified genomic region, including only variants in common between them and available within a reference population. An LD correlation matrix across sampled variants is derived from a reference population using PLINK (v1.90)[32,33].

Quality control is performed per-dataset prior to univariate fine-mapping analysis. Diagnostic tools provided with *susieR* are applied to test for consistency between the LD

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

matrix and Z-scores from the GWAS and identify variants with a potential 'allele flip' (reversed effect estimate encoding) that can impact fine-mapping.

Fine-mapping is performed for each dataset with the *coloc* package *runsusie* function, which wraps around *susie_rss* from *susieR* and is configured to facilitate subsequent colocalisation analysis. Sample size (N_{eff} for binary traits) is specified as the median for SNPs analysed. Colocalisation analysis can be performed with the coloc functions *coloc.abf* and *coloc.susie* when fine-mapping yields at least one credible set for both traits and otherwise using *coloc.abf* only. Genes located near credible sets from fine-mapping and credible SNPs from colocalisation analyses are identified via Ensembl and *biomaRt (v2.54.0)*[34-36].

Analysis parameters can be adjusted by the user in accordance with their needs. Various utilities are included to help interpretation of fine-mapping and colocalisation results, including identification of genes nearby to putatively causal signals, HTML reports to summarise completed analyses, and figures to visualise the results and compare the examined traits.

Current implementation

In this study, LD correlation matrices were derived from the 1KG European cohort. SNPs flagged for potential allele flip issues in either of the compared traits were removed from the analysis. Fine-mapping was performed with the *susie_rss* refine=TRUE option to avoid local maxima during convergence of the algorithm, leaving the other settings to the *runsusie* defaults. Colocalisation analysis was performed using the default priors for *coloc.susie* ($P_1=1x10^{-4}$, $P_2=1x10^{-4}$, $P_{12}=5x10^{-6}$).

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Colocalisation and fine-mapping analyses were performed initially using the genomic blocks defined by LAVA, since these aim to define relatively independent LD partitions across the genome[8]. If a 95% credible set could not be identified in one or both traits, we inspected local Manhattan plots for the region to determine whether potentially relevant signals occurred around the region boundaries. The analysis was repeated with a ±10Kb window around the LAVA-defined genomic region if p-values for SNPs at the edge of the block were $p<1x10^{-4}$ for both traits and the Manhattan plots were suggestive of a 'peak' not represented within the original boundaries.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

3. Results

3.1. Genome-wide analyses

Descriptive information and heritability estimates for the sampled traits and GWAS are presented in Table 1. ALS had nominally significant global genetic correlations with schizophrenia (p = 0.045), PD (p = 0.013), and AD (p = 0.006); no other bivariate genome-wide correlations were statistically significant (see Figure S1).

Table 1. Genome-wide association studies (GWAS) sampled

*Each GWAS is a GWAS meta-analysis of disease risk across people of European ancestry. *Proxy cases from the UK Biobank cohort. †Estimated from cumulative risk after age 45 after correcting for competing risk of mortality and assuming a lifespan of ~85 years. h^2 = heritability*

Trait	Estimated lifetime risk in population	GWAS			Liability scale h^2 (standard error)
		Reference	N Cases	N Controls	
Alzheimer's disease	1/10[37] †	Kunkle *et al.*, 2019[20]	21,982	41,944	0.093 (0.0155)
Amyotrophic lateral sclerosis	1/350[38,39]	van Rheenen *et al.*, 2021[1]	27,205	110,881	0.0277 (0.003)
Frontotemporal dementia	1/742[40]	Ferrari *et al.*, 2014[21]	2,154	4,308	0.0329 (0.0283)
Parkinson's disease	1/37[41]	Nalls *et al.*, 2019[22]	15,056 (+ 18,618 proxies*)	449,056	0.0506 (0.0046)

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Schizophrenia	1/250[42]	Trubetskoy *et al.*, 2022[23]	53,386	77,258	0.1761 (0.0061)

A total of 605 local genetic correlation analyses were performed across all trait pairs in genomic regions where both traits passed the univariate heritability filtering step after restricting to SNPs sampled in both GWAS (see Table 2; Figure 2; Table S1). The number of loci passing to bivariate analysis varied greatly across trait pairs and was congruent with the genome-wide heritability estimates (and their uncertainty) for each trait, reflecting differences in phenotypic variance explained by measured genetic variants and statistical power for each GWAS (see Table 1).

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 2. Comparison of genome-wide SNP significance against local genetic correlation significance thresholds in all trait pairs and loci analysed

All loci analysed showed sufficient local univariate heritability across compared traits to allow bivariate correlation analysis. Subsequent fine-mapping and colocalisation analyses were performed in this study for regions with at least a false discovery rate (FDR) adjusted significance for the local genetic correlation. SNP = single nucleotide polymorphism.

Number of traits in pair with genome-wide significant ($p < 5x10^{-8}$) SNP in locus	Smallest significance threshold for local genetic correlation			
	Bonferroni ($p < 8.26x10^{-5}$; 0.05/605)	FDR ($p_{fdr} < 0.05$)	Nominal ($p < 0.05$)	Non-significant ($p \geq 0.05$)
0	1	17	77	394
1	1	4	18	80
2	0	3	2	8

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

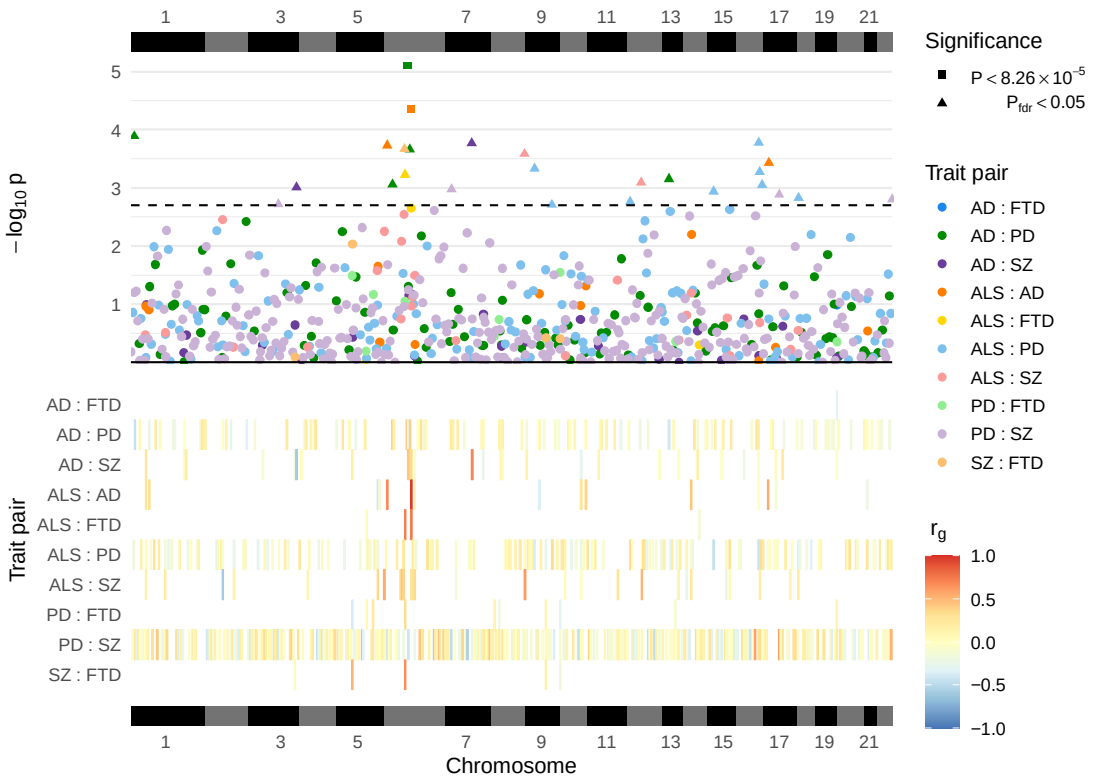

Figure 2. Local genetic correlation analyses between trait pairs

The lower panel displays a heatmap of genetic correlations (r_g) across genomic regions where any bivariate analyses were performed; white colouring indicates that the region was not analysed for a given trait pair owing to insufficient univariate heritability in one or both traits. The upper panel shows a Manhattan plot of p-values from each correlation analysis, denoting trait pairs by colour and comparisons passing defined significance thresholds by shape (square for a strict Bonferroni threshold and triangle for a false discovery rate (FDR) adjusted threshold); the hatched line indicates the threshold p-value above which P_{fdr}<0.05. The panels are both ordered by relative genomic position, with bars above and below indicating each chromosome. AD = Alzheimer's disease, ALS = amyotrophic lateral sclerosis, FTD = frontotemporal dementia, PD = Parkinson's disease, SZ = schizophrenia. Table S1 provides a complete summary of local genetic correlation analyses performed.

Twenty-six bivariate comparisons were significant following FDR adjustment (p_{fdr} <0.05), two of which also passed the stringent Bonferroni threshold (p <8.26x10^{-5}; 0.05/605). While some regions included genome-wide significant SNPs (p <5x10^{-8}) for one or both traits,

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

others occurred in regions where GWAS associations were weaker (see Table 2). Five of these associations occurred at loci within the human leukocyte antigen (HLA) region (GRCh37: Chr6:28.48-33.45Mb; 6p22.1-21.3[43]), and all five traits were implicated in at least one of these.

3.2. Targeted genetic analyses

Univariate fine-mapping and bivariate colocalisation analyses were subsequently performed to test for variants jointly implicated between trait pairs in regions with local genetic correlation P_{fdr} <0.05. The ALS and schizophrenia trait pair was additionally examined at Chr6:32.22-32.45Mb because significant genetic correlations were found between ALS and FTD and between schizophrenia and FTD at this locus. The correlation between ALS and schizophrenia at this locus had not been analysed owing to insufficient univariate heritability for ALS after restricting to SNPs in common with the schizophrenia GWAS.

Fine-mapping identified at least one 95% credible set for each of the compared traits for 7 of the 27 comparisons performed (see Table 3), and for one trait only in a further 5 (see Table S2; Table S3). This analysis suggested two credible sets for schizophrenia in the Chr12:56.99-58.75Mb locus, for AD in Chr6:32.45-32.54Mb, and (only when harmonised to SNPs in common with the ALS GWAS) for FTD in Chr6:32.22-32.45Mb (see Table S3). Although both positive and negative local genetic correlations passed the FDR-adjusted significance threshold, we observed only positive local genetic correlations in loci where fine-mapping credible sets were identified for both traits in the pair. This reflects that the absolute correlation coefficients and variant associations from the analysed GWAS studies were generally stronger in the positively correlated loci (see Figure S2).

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Colocalisation analyses performed across fine-mapping credible sets and across all SNPs in a region generally gave support to the equivalent hypothesis (Table 3; Table S2). Moreover, comparisons suggesting a signal was present in one trait only were largely concordant with the identification of fine-mapping credible sets in only that trait (Table S2). Figure S3 compares per-SNP p-values across trait pairs for comparisons with evidence of a relevant signal in both traits. Figure S4 shows patterns of LD across SNPs assigned to credible sets for these analyses.

Strong evidence was found for a shared variant between ALS and AD within the *HLA* region (Posterior probability of shared variant = 0.9; see Figure 3). The 95% credible SNPs for this association were distributed around the *MTCO3P1* pseudogene and rs9275477, the lead genome-wide significant SNP from the ALS GWAS in this region, had the highest posterior probability of being implicated in both traits. Figure S5 presents sensitivity analysis showing that the result is robust to a range of values for the shared variant hypothesis prior probability.

The other comparisons that found fine-mapping credible sets in both traits suggested that overlaps from the correlation analysis were driven by distinct causal variants (see Table 3; Table S2).

Univariate fine-mapping of PD and schizophrenia at Chr17:43.46-44.87Mb found large credible sets spanning many genes, including *MAPT*[44-47] and *CRHR1*[48,49] which have been previously implicated in the traits we have analysed. These expansive credible sets reflect

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the strong LD in the region and indicate a signal that is difficult to localise (see Figure S4(F); Table S3). The colocalisation analysis suggested independent variants for each trait despite many SNPs overlapping across their respective credible sets (see Figure S4). Sensitivity analysis showed robust support for the two independent variants hypothesis across shared-variant hypothesis priors (Figure S5). However, the colocalisation analysis will increasingly favour the two independent variants hypothesis as the number of analysed variants increases[50]. Hence, the wide-spanning LD of this region may have obstructed identification of variants and mechanisms shared between the traits.

Table 3. Colocalisation analysis conducted across 95% credible sets identified during univariate fine-mapping of trait pairs

*N SNPs refers to the number of SNPs present for both traits and the 1000 genomes reference panel in the region within colocalisation and fine-mapping analysis. *Indicates comparisons with genetic correlation analysis p <8.26x10^{-5} (0.05/605). [Δ] Denotes locus extended by ±10kb for fine-mapping and colocalisation analysis. [†]Variant identified in colocalisation as having the highest posterior probability of being shared variant assuming hypothesis 4 is true (see Figure 3). [§]Differences in fine-mapping solutions across trait pairs in the Chr6:32.21-32.45Mb locus reflect differences in the SNPs retained after restricting to those in common between the compared GWAS [ø] H0 = no causal variant for either trait, H1 = variant causal for trait 1, H2 = variant causal for trait 2, H3 = distinct causal variants for each trait, H4 = a shared causal variant between traits. PIP = posterior inclusion probability. AD = Alzheimer's disease, ALS = amyotrophic lateral sclerosis, FTD = frontotemporal dementia, PD = Parkinson's disease, SZ = schizophrenia.*

Trait		Genomic position (GRCh37)	Local genetic correlation estimate (95% Confidence Interval)	Fine-mapping Credible set for trait		N SNPs	SNP with highest PIP for fine-mapping credible set (nearest gene; sense-strand base pair distance)		Posterior probability for hypothesis[ø]				
1	2			1	2		Trait 1	Trait 2	H0	H1	H2	H3	H4
AD	PD	Chr6:32576785-32639239 [Δ]	0.406 (0.197, 0.648)	1	1	958	rs9271247 (*HLA-DQA1*; +15,844)	rs3129751 (*HLA-DQA1*; +13,767)	<0.01	<0.01	<0.01	**0.95**	0.05
ALS	AD	Chr6:32629240-32682213 [*]	0.974 (0.717, 1.000)	1	1	475	rs9275477[†] (*MTCO3P1;* +1,260)	rs9275207 (*MTCO3P1*; +16,191)	<0.01	<0.01	<0.01	0.10	**0.90**
ALS	FTD	Chr6:32208902-32454577 [§]	0.723 (0.370, 1.000)	1	1	1709	rs9268833 (*HLA-DRB9*; 0)	rs1980493 (*BTNL2*; 0)	<0.01	<0.01	0.01	**0.99**	<0.01
					2			rs9767620 (*HLA-DRB9*; +1,498)	<0.01	<0.01	0.01	**0.99**	<0.01
ALS	SZ	Chr6:32208902-32454577 [§]	-	1	1	1711	rs9268833 (*HLA-DRB9*; 0)	rs9268219 (*C6orf10*; 0)	<0.01	<0.01	<0.01	**0.98**	<0.01
		Chr12:56987106-58748139	0.506 (0.218, 0.807)	1	1	2260	rs113247976 (*KIF5A;* 0)	rs12814239 (*LRP1*; 0)	<0.01	<0.01	<0.01	**1.00**	<0.01
					2			rs324017 (*NAB2*; 0)	<0.01	<0.01	<0.01	**1.00**	<0.01
PD	SZ	Chr17:43460501-44865832	0.595 (0.266, 0.950)	1	1	2453	rs58879558 (*MAPT*; 0)	rs62062288 (*MAPT*; 0)	<0.01	<0.01	<0.01	**0.81**	0.19
SZ	FTD	Chr6:32208902-32454577 [§]	0.669 (0.379, 0.990)	1	1	1657	rs9268219 (*C6orf10*; 0)	rs9268877 (*HLA-DRB9*; 0)	<0.01	<0.01	<0.01	**1.00**	<0.01

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

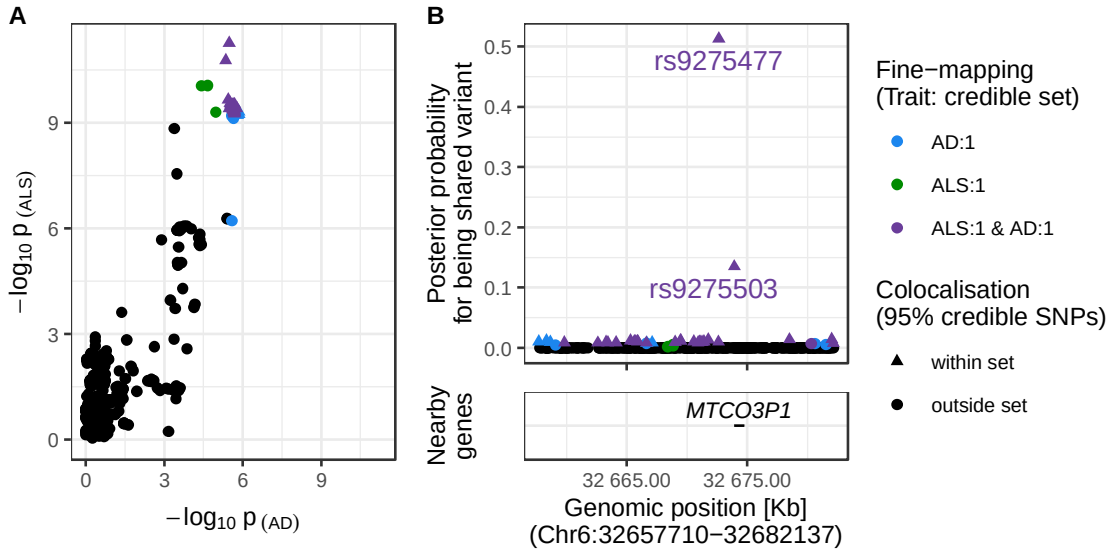

Figure 3. Evidence for colocalisation between amyotrophic lateral sclerosis (ALS) and Alzheimer's disease (AD) in the Chr6:32.63-32.68Mb region

Panel A: *SNP-wise p-value distribution between ALS and AD across Chr6:32.63-32.68Mb, in which colocalisation analysis found 0.90 posterior probability of the shared variant hypothesis (see Table 3).* ***Panel B:*** *(upper) Per-SNP posterior probabilities for being a shared variant between ALS and AD, (lower) positions of HGNC gene symbols nearby to the 95% credible SNPs. Posterior probabilities for being a shared variant sum to 1 across all SNPs analysed and are predicated on the assumption that a shared variant exists; 95% credible SNPs are those spanned by the top 0.95 of posterior probabilities. The x-axis for Panel B is truncated by the base pair range of the credible SNPs and genomic positions are based on GRCh37.*

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

4. Discussion

We examined genetic overlaps between the neuropsychiatric conditions Alzheimer's disease, amyotrophic lateral sclerosis, frontotemporal dementia, Parkinson's disease, and schizophrenia. Through genetic correlation analysis, we replicated genome-wide correlations previously described between the studied traits[1,2,18,19]. Leveraging a more recent local genetic correlation approach, we identified specific genomic loci jointly implicated between pairs of traits which were further investigated using statistical fine-mapping and colocalisation techniques.

Significant local genetic correlations were most frequent across genomic blocks within the *HLA* region, implicating each of the studied traits in at least one comparison. Several associated regions contained genes with known relevance for the traits studied, such as *KIF5A, MAPT, and CRHR1*. Colocalisation analysis found strong evidence for a shared genetic variant between ALS and AD in the Chr6:32.62-32.68Mb locus within *HLA*, while the other colocalisation analyses suggested causal signals distinct across traits, for one trait only, or for neither trait.

The tendency for association between traits around the *HLA* region is reasonable, since this is a known hotspot for pleiotropy[8,51]. HLA is particularly known for its role in immune response and it is implicated in various types of disease[52,53]. Mounting evidence has linked *HLA* and associated genetic variation to the traits we have analysed, and mechanisms underlying these associations are beginning to be understood[52-61]. For instance, AD is associated with variants around the *HLA-DQA1* and *HLA-DRB1* genes and several SNPs in the non-coding region between them have been shown to modulate their expression[62]. Notably,

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

one of the SNPs with a demonstrated regulatory role, rs9271247, had the highest probability of being causal for AD across the 95% credible set identified in the fine-mapping of the region.

Variants showing evidence for colocalisation between AD and ALS were distributed around the *MTCO3P1* pseudogene in the *HLA* class II non-coding region between *HLA-DQB1* and *HLA-DQB2*. *MTCO3P1* has been previously identified as one of the most pleiotropic genes in the GWAS catalog[63,64]. Previous studies have suggested the relevance of this region in both traits. *HLA-DQB1* and *HLA-DQB2* are both upregulated in the spinal cord of people with ALS, alongside other genes implicated in various immunological processes for antigen processing and inflammatory response[65]. *HLA* class II complexes, and their subcomponents, have been identified as upregulated in multiple brain regions of people with AD, using both gene and protein expression techniques[61,66].

Our analysis of this region gave stronger support for colocalisation between the ALS and AD GWAS than a previous study[1]. The previous study analysed a 200Kb window of over 2,000 SNPs around the lead genome-wide significant SNP from the ALS GWAS, rs9275477, and found ~0.50 posterior probability for each of the shared and two independent variant(s) hypotheses. The current analysis used 475 SNPs occurring within a semi-independent LD block of ~50kb in this locus. Since the posterior probability of the two independent variants hypothesis (H3) increases exponentially with the number of variants in the region whilst the shared variant hypothesis (H4) scales linearly, it is expected that our analysis would give stronger support for the latter[50]. Given that the previous study defined regions for analysis based on an arbitrary window of ±100kb around each lead genome-wide significant SNP

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

from the ALS GWAS and we defined each analysis region based on patterns of LD in European ancestry populations, it is reasonable to favour the current finding.

More broadly, our analyses suggest that regions with a strong genetic correlation between the five traits studied often result from adjacent but trait-specific signals, likely reflecting overlaps between LD blocks[51]. Correlations also occurred in regions with weaker overall GWAS associations (see Table 2), where fine-mapping and colocalisation analyses did not suggest causal associations in one or either trait. Such patterns likely reflect a shared polygenic trend across the region, rather than associations attributable to discrete variants. Accordingly, other approaches may be better suited for identifying regions containing genetic variation jointly causal across diseases, including the traditional approach of testing regions around overlapping genome-wide significant variants.

This study has used gold-standard statistical tools to examine genetic relationships between traits. The local genetic correlation analysis approach enabled targeted investigation of genomic regions which appear to overlap between traits. The application of colocalisation analysis alongside a prior univariate fine-mapping step allowed for associations to be tested without conflating independent but nearby signals under the single-variant assumption of colocalisation analysis across all variants sampled in a region.

The study is not without limitation. We necessarily used the 1KG European reference population to estimate LD between SNPs. Fine-mapping is ideally performed with an LD matrix from the GWAS sample and is sensitive to misspecification when inconsistencies in LD occur between the reference and GWAS cohorts. Use of a reference population is not

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

uncommon, and diagnostic tools available within the *susieR* package allow testing for inconsistencies between the reference and GWAS samples[11]. We accordingly implemented these tools centrally into our workflow and determined that the LD matrices from the 1KG reference were suitable for the data (estimates of Z-score and LD consistency are available in Table S3). Nevertheless, repeating this study in under-represented populations would be an important future step to validate our findings.

We employed statistical methods to identify and analyse genomic regions containing variants which might be jointly implicated across traits. These approaches provide useful associations between traits identified from large-scale genomic datasets. However, they alone are not sufficient for translation into clinical practice. Future studies should aim to extend any associations found by integrating functional and multi-omics datasets to gain mechanistic insights into observed trends and facilitate treatment discovery[62,67].

The fine-mapping and colocalisation analysis pipeline we have used is available as an open-access resource on GitHub to facilitate the application of these methods in future studies: https://github.com/ThomasPSpargo/COLOC-reporter. Specified genomic regions can be readily analysed by providing GWAS summary statistics for binary or quantitative traits of interest and a population-appropriate reference dataset for estimation of LD. The pipeline returns resources including detailed reports that overview the analyses performed.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Supplementary materials

Supplementary tables and figures can be found at https://github.com/KHP-Informatics/COLOC-reporter_supplementary_materials.

Funding

This project was part funded by the MND Association and the Wellcome Trust. This is an EU Joint Programme-Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organizations under the aegis of JPND–http://www.neurodegenerationresearch.eu/ [United Kingdom, Medical Research Council (MR/L501529/1 and MR/R024804/1) and Economic and Social Research Council (ES/L008238/1)]. AAC is a NIHR Senior Investigator. AAC received salary support from the National Institute for Health Research (NIHR) Dementia Biomedical Research Unit at South London and Maudsley NHS Foundation Trust and King's College London. The work leading up to this publication was funded by the European Community's Health Seventh Framework Program (FP7/2007–2013; grant agreement number 259867) and Horizon 2020 Program (H2020-PHC-2014-two-stage; grant agreement number 633413). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 772376–EScORIAL. This study represents independent research part funded by the NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, King's College London, or the Department of Health and Social Care. Funding

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

was also provided by the King's College London DRIVE-Health Centre for Doctoral Training and the Perron Institute for Neurological and Translational Science. AI is funded by South London and Maudsley NHS Foundation Trust, MND Scotland, Motor Neurone Disease Association, National Institute for Health and Care Research, Spastic Paraplegia Foundation, Rosetrees Trust, Darby Rimmer MND Foundation, the Medical Research Council (UKRI) and Alzheimer's Research UK. OP is supported by a Sir Henry Wellcome Postdoctoral Fellowship [222811/Z/21/Z]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

This research was funded in whole or in part by the Wellcome Trust [222811/Z/21/Z]. For the purpose of open access, the author has applied a CC-BY public copyright licence to any author accepted manuscript version arising from this submission.

Acknowledgements

The authors acknowledge the use of the CREATE research computing facility at King's College London[68]. We also acknowledge Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (United Kingdom), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

1. van Rheenen W, van der Spek RAA, Bakker MK, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. Nat Genet. 2021; 53(12):1636-48. doi:https://doi.org/10.1038/s41588-021-00973-1
2. Li C, Yang T, Ou R, Shang H. Overlapping Genetic Architecture Between Schizophrenia and Neurodegenerative Disorders. Front Cell Dev Biol. 2021; 9. doi:https://doi.org/10.3389/fcell.2021.797072
3. Ranganathan R, Haque S, Coley K, Shepheard S, Cooper-Knock J, Kirby J. Multifaceted Genes in Amyotrophic Lateral Sclerosis-Frontotemporal Dementia. Front Neurosci. 2020; 14. doi:https://doi.org/10.3389/fnins.2020.00684
4. Abdellaoui A, Yengo L, Verweij KJH, Visscher PM. 15 years of GWAS discovery: Realizing the promise. Am J Hum Genet. 2023; 110(2):179-94. doi:https://doi.org/10.1016/j.ajhg.2022.12.011
5. Reales G, Wallace C. Sharing GWAS summary statistics results in more citations. Commun Biol. 2023; 6(1):116. doi:https://doi.org/10.1038/s42003-023-04497-8
6. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015; 47(3):291-5. doi:https://doi.org/10.1038/ng.3211
7. Bulik-Sullivan BK, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015; 47(11):1236-41. doi:https://doi.org/10.1038/ng.3406
8. Werme J, van der Sluis S, Posthuma D, de Leeuw CA. An integrated framework for local genetic correlation analysis. Nat Genet. 2022; 54(3):274-82. doi:https://doi.org/10.1038/s41588-022-01017-y
9. Zhang Y, Lu Q, Ye Y, et al. SUPERGNOVA: local genetic correlation analysis reveals heterogeneous etiologic sharing of complex traits. Genome Biol. 2021; 22(1):262. doi:https://doi.org/10.1186/s13059-021-02478-w
10. Shi H, Mancuso N, Spendlove S, Pasaniuc B. Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. Am J Hum Genet. 2017; 101(5):737-51. doi:https://10.1016/j.ajhg.2017.09.022
11. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the “Sum of Single Effects” model. PLoS Genet. 2022; 18(7):e1010299. doi:https://doi.org/10.1371/journal.pgen.1010299
12. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet. 2021; 17(9):e1009440. doi:https://doi.org/10.1371/journal.pgen.1009440
13. Giambartolomei C, Zhenli Liu J, Zhang W, et al. A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics. 2018; 34(15):2538-45. doi:https://doi.org/10.1093/bioinformatics/bty147
14. Foley CN, Staley JR, Breen PG, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. Nat Commun. 2021; 12(1):764. doi:https://doi.org/10.1038/s41467-020-20885-8
15. Weintraub D, Mamikonyan E. The Neuropsychiatry of Parkinson Disease: A Perfect Storm. Am J Geriatr Psychiatry. 2019; 27(9):998-1018. doi:https://doi.org/10.1016/j.jagp.2019.03.002

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

16. Ferrari R, Wang Y, Vandrovcova J, et al. Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. J Neurol Neurosurg Psychiatry. 2017; 88(2):152-64. doi:https://doi.org/10.1136/jnnp-2016-314411
17. Beck J, Poulter M, Hensman D, et al. Large *C9orf72* hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. Am J Hum Genet. 2013; 92(3):345-53. doi:https://doi.org/10.1016/j.ajhg.2013.01.011
18. Wainberg M, Andrews SJ, Tripathy SJ. Shared genetic risk loci between Alzheimer's disease and related dementias, Parkinson's disease, and amyotrophic lateral sclerosis. Alzheimers Res Ther. 2023; 15(1):113. doi:https://10.1186/s13195-023-01244-3
19. McLaughlin RL, Schijven D, van Rheenen W, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. Nat Commun. 2017; 8(1):14774. doi:https://doi.org/10.1038/ncomms14774
20. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019; 51(3):414-30. doi:https://doi.org/10.1038/s41588-019-0358-2
21. Ferrari R, Hernandez DG, Nalls MA, et al. Frontotemporal dementia and its subtypes: a genome-wide association study. Lancet Neurol. 2014; 13(7):686-99. doi:https://doi.org/10.1016/S1474-4422(14)70065-1
22. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019; 18(12):1091-102. doi:https://doi.org/10.1016/S1474-4422(19)30320-5
23. Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature. 2022; 604(7906):502-8. doi:https://doi.org/10.1038/s41586-022-04434-5
24. Pain O, Glanville KP, Hagenaars SP, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. PLoS Genet. 2021; 17(5):e1009021. doi:https://doi.org/10.1371/journal.pgen.1009021
25. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. Nature. 2015; 526(7571):68-74. doi:https://doi.org/10.1038/nature15393
26. Privé F, Aschard H, Ziyatdinov A, Blum MGB. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. Bioinformatics. 2018; 34(16):2781-7. doi:https://doi.org/10.1093/bioinformatics/bty185
27. Grotzinger AD, Fuente Jdl, Privé F, Nivard MG, Tucker-Drob EM. Pervasive Downward Bias in Estimates of Liability-Scale Heritability in Genome-wide Association Study Meta-analysis: A Simple Solution. Biol Psychiatry. 2023. doi:https://doi.org/10.1016/j.biopsych.2022.05.029
28. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010; 467(7311):52-8. doi:https://doi.org/10.1038/nature09298
29. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2020; 82(5):1273-300. doi:https://doi.org/10.1111/rssb.12388

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

30. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. 2014; 10(5):e1004383. doi:https://doi.org/10.1371/journal.pgen.1004383
31. R Core Team. R: A language and environment for statistical computing. 2021. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/
32. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007; 81(3):559-75. doi:https://doi.org/10.1086/519795
33. Purcell S. PLINK. 1.9.0. Available from: http://pngu.mgh.harvard.edu/purcell/plink/
34. Durinck S, Moreau Y, Kasprzyk A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005; 21(16):3439-40. doi:https://doi.org/10.1093/bioinformatics/bti525
35. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009; 4(8):1184-91. doi:https://doi.org/10.1038/nprot.2009.97
36. Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. Nucleic Acids Res. 2022; 50(D1):D988-D95. doi:https://doi.org/10.1093/nar/gkab1049
37. Chêne G, Beiser A, Au R, et al. Gender and incidence of dementia in the Framingham Heart Study from mid-adult life. Alzheimers Dement. 2015; 11(3):310-20. doi:https://doi.org/10.1016/j.jalz.2013.10.005
38. Alonso A, Logroscino G, Jick SS, Hernán MA. Incidence and lifetime risk of motor neuron disease in the United Kingdom: a population-based study. European Journal of Neurology. 2009; 16(6):745-51. doi:https://doi.org/10.1111/j.1468-1331.2009.02586.x
39. Johnston CA, Stanton BR, Turner MR, et al. Amyotrophic lateral sclerosis in an urban setting: A population based study of inner city London. J Neurol. 2006; 253(12):1642-3. doi:https://doi.org/10.1007/s00415-006-0195-y
40. Coyle-Gilchrist ITS, Dick KM, Patterson K, et al. Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes. Neurology. 2016; 86(18):1736. doi:https://doi.org/10.1212/WNL.0000000000002638
41. Parkinson's UK. The Incidence and Prevalence of Parkinson's in the UK: Results from the Clinical Practice Research Datalink Reference Report. 2017. Available from: https://www.parkinsons.org.uk/professionals/resources/incidence-and-prevalence-parkinsons-uk-report
42. Saha S, Chant D, Welham J, McGrath J. A Systematic Review of the Prevalence of Schizophrenia. PLoS Med. 2005; 2(5):e141. doi:https://doi.org/10.1371/journal.pmed.0020141
43. Genome Reference Consortium. Human Genome Region MHC [cited 2023 7th March]. Available from: https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC?asm=GRCh37.
44. Allen M, Kachadoorian M, Quicksall Z, et al. Association of MAPT haplotypes with Alzheimer's disease risk and MAPT brain gene expression levels. Alzheimers Res Ther. 2014; 6(4):39. doi:https://doi.org/10.1186/alzrt268
45. Snowden JS, Adams J, Harris J, et al. Distinct clinical and pathological phenotypes in frontotemporal dementia associated with MAPT, PGRN and C9orf72 mutations. Amyotroph Lateral Scler Frontotemporal Degener. 2015; 16(7-8):497-505. doi:https://doi.org/10.3109/21678421.2015.1074700

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

46. Origone P, Geroldi A, Lamp M, et al. Role of *MAPT* in Pure Motor Neuron Disease: Report of a Recurrent Mutation in Italian Patients. Neurodegener Dis. 2018; 18(5-6):310-4. doi:https://doi.org/10.1159/000497820
47. Nakayama S, Shimonaka S, Elahi M, et al. Tau aggregation and seeding analyses of two novel MAPT variants found in patients with motor neuron disease and progressive parkinsonism. Neurobiol Aging. 2019. doi:https://doi.org/10.1016/j.neurobiolaging.2019.02.016
48. Cheng WW, Zhu Q, Zhang HY. Identifying Risk Genes and Interpreting Pathogenesis for Parkinson's Disease by a Multiomics Analysis. Genes (Basel). 2020; 11(9). doi:https://doi.org/10.3390/genes11091100
49. Bigdeli TB, Fanous AH, Li Y, et al. Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans. Schizophr Bull. 2020; 47(2):517-29. doi:https://doi.org/10.1093/schbul/sbaa133
50. Wallace C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. PLoS Genet. 2020; 16(4):e1008720. doi:http://10.1371/journal.pgen.1008720
51. Watanabe K, Stringer S, Frei O, et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet. 2019; 51(9):1339-48. doi:https://doi.org/10.1038/s41588-019-0481-0
52. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. Nature Reviews Immunology. 2018; 18(5):325-39. doi:https://doi.org/10.1038/nri.2017.143
53. Trowsdale J, Knight JC. Major Histocompatibility Complex Genomics and Human Disease. Annu Rev Genom Hum. 2013; 14(1):301-23. doi:https://doi.org/10.1146/annurev-genom-091212-153455
54. Wang Z-X, Wan Q, Xing A. HLA in Alzheimer's Disease: Genetic Association and Possible Pathogenic Roles. Neuromolecular Med. 2020; 22(4):464-73. doi:https://doi.org/10.1007/s12017-020-08612-4
55. Song S, Miranda CJ, Braun L, et al. Major histocompatibility complex class I molecules protect motor neurons from astrocyte-induced toxicity in amyotrophic lateral sclerosis. Nat Med. 2016; 22(4):397-403. doi:https://doi.org/10.1038/nm.4052
56. Yu E, Ambati A, Andersen MS, et al. Fine mapping of the HLA locus in Parkinson's disease in Europeans. NPJ Parkinson's Dis. 2021; 7(1):84. doi:https://doi.org/10.1038/s41531-021-00231-5
57. Broce I, Karch CM, Wen N, et al. Immune-related genetic enrichment in frontotemporal dementia: An analysis of genome-wide association studies. PLoS Med. 2018; 15(1):e1002487. doi:https://doi.org/10.1371/journal.pmed.1002487
58. Ferrari R, Forabosco P, Vandrovcova J, et al. Frontotemporal dementia: insights into the biological underpinnings of disease through gene co-expression network analysis. Mol Neurodegener. 2016; 11(1):21. doi:https://doi.org/10.1186/s13024-016-0085-4
59. Al-Diwani AAJ, Pollak TA, Irani SR, Lennox BR. Psychosis: an autoimmune disease? Immunology. 2017; 152(3):388-401. doi:https://doi.org/10.1111/imm.12795
60. Mokhtari R, Lachman HM. The Major Histocompatibility Complex (MHC) in Schizophrenia: A Review. J Clin Cell Immunol. 2016; 7(6). doi:https://doi.org/10.4172/2155-9899.1000479
61. Aliseychik MP, Andreeva TV, Rogaev EI. Immunogenetic Factors of Neurodegenerative Diseases: The Role of HLA Class II. Biochemistry (Moscow). 2018; 83(9):1104-16. doi:https://doi.org/10.1134/S0006297918090122

medRxiv preprint doi: https://doi.org/10.1101/2023.03.30.23287950; this version posted June 26, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

62. Zhang X, Zou M, Wu Y, et al. Regulation of the Late Onset Alzheimer's Disease Associated *HLA-DQA1/DRB1* Expression. Am J Alzheimers Dis Other Demen. 2022; 37:15333175221085066. doi:https://doi.org/10.1177/15333175221085066
63. Chesmore K, Bartlett J, Williams SM. The ubiquity of pleiotropy in human disease. Hum Genet. 2018; 137(1):39-44. doi:https://doi.org/10.1007/s00439-017-1854-z
64. Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2022; 51(D1):D977-D85. doi:https://doi.org/10.1093/nar/gkac1010
65. Andrés-Benito P, Moreno J, Aso E, Povedano M, Ferrer I. Amyotrophic lateral sclerosis, gene deregulation in the anterior horn of the spinal cord and frontal cortex area 8: implications in frontotemporal lobar degeneration. Aging. 2017; 9(3):823-51. doi:https://doi.org/10.18632/aging.101195
66. Hopperton KE, Mohammad D, Trépanier MO, Giuliano V, Bazinet RP. Markers of microglia in post-mortem brain samples from patients with Alzheimer's disease: a systematic review. Mol Psychiatry. 2018; 23(2):177-98. doi:https://doi.org/10.1038/mp.2017.246
67. Pain O, Jones A, Khleifat AA, et al. Harnessing Transcriptomic Signals for Amyotrophic Lateral Sclerosis to Identify Novel Drugs and Enhance Risk Prediction. medRxiv. 2023:2023.01.18.23284589. doi:https://doi.org/10.1101/2023.01.18.23284589
68. King’s College London. King's Computational Research, Engineering and Technology Environment (CREATE) 2022 [October 12 2022]. Available from: https://doi.org/10.18742/rnvf-m076.