**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\(^1,5\), Alfredo Iacoangeli\(^1,2,3,5,*\)

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
*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 the presence of a shared causal variant between amyotrophic lateral sclerosis and Alzheimer’s disease in this region. Our study identified candidate loci that might play a role in multiple neuropsychiatric diseases and suggested that disease-implicated variants in these loci often differ between traits. Accordingly, this suggests 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](https://github.com/ThomasPSpargo/COLOC-reporter).

**NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Keywords: statistical genetics, GWAS, fine-mapping, colocalisation, neurodegeneration genetics, psychiatric disease
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. 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. Summary-level results from GWAS are being shared alongside publications with increasing frequency over time, 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. 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. Recent methods allow for a more nuanced analysis, of ‘local’ genetic correlations partitioned across the genome. 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.

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.

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. 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: AD, ALS, FTD, PD, and schizophrenia. Loci highly correlated between trait pairs were further investigated.
with univariate fine-mapping and bivariate colocalisation techniques to examine variants driving these associations.
2. Methods
2.1. Sampled GWAS summary statistics
We leveraged publicly-accessible summary statistics from European ancestry GWAS meta-analyses of risk for AD\textsuperscript{17}, ALS\textsuperscript{1}, FTD\textsuperscript{18}, PD\textsuperscript{19}, and schizophrenia\textsuperscript{20}. 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.

![Image](https://example.com/image.png)

**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/ThomasPSparag/COLOC-reporter](https://github.com/ThomasPSparag/COLOC-reporter).*

2.2.1. Processing of GWAS summary statistics
A standard data cleaning protocol was applied to each set of summary statistics\textsuperscript{21}. 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\textsuperscript{22} (N = 503). They were matched to the 1KG reference panel by GRCh37 chromosomal position using bigsnpr (version 1.11.6)\textsuperscript{23}, harmonising allele order with the reference and assigning SNP IDs.

If not reported, and where possible, effective sample size (N\textsubscript{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\textsubscript{eff}, calculated as a sum of N\textsubscript{eff} values for each cohort within the GWAS meta-analysis\textsuperscript{24}.
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)\textsuperscript{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 overlap from the bivariate intercept.

These analyses were performed using the HapMap3\textsuperscript{25} 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)\textsuperscript{8} was applied to obtain local genetic correlation estimates across 2495 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.

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} (0.05/2495) 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)\textsuperscript{10,26}, 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.

Colocalisation analysis was implemented with coloc (v5.1.0.1)\textsuperscript{11,12,27}, 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\textsuperscript{11}. 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)\textsuperscript{28}: \url{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)\textsuperscript{29,30}.

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 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 \texttt{runsusie} function, which wraps around \texttt{susie_rss} from susieR and is configured to facilitate subsequent colocalisation analysis. Sample size ($N_{\text{eh}}$ 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)\textsuperscript{31-33}.

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 ($P_1=1 \times 10^{-4}, P_2=1 \times 10^{-4}, P_{12}=5 \times 10^{-6}$).

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. 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 $p<1 \times 10^{-4}$ for both traits and the Manhattan plots were suggestive of a ‘peak’ not represented within the original boundaries.
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

<table>
<thead>
<tr>
<th>Trait</th>
<th>Estimated lifetime risk in population</th>
<th>Reference</th>
<th>GWAS</th>
<th>Liability scale $h^2$ (standard error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer's disease</td>
<td>1/10$^{3,4}$†</td>
<td>Kunkle et al., 2019$^7$</td>
<td>21,982</td>
<td>41,944</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis</td>
<td>1/350$^{35,36}$</td>
<td>van Rheenen et al., 2021$^1$</td>
<td>27,205</td>
<td>110,881</td>
</tr>
<tr>
<td>Frontotemporal dementia</td>
<td>1/742$^{37}$</td>
<td>Ferrari et al., 2014$^{18}$</td>
<td>2,154</td>
<td>4,308</td>
</tr>
<tr>
<td>Parkinson's disease</td>
<td>1/37$^{38}$</td>
<td>Nalls et al., 2019$^{19}$</td>
<td>15,056 (+18,618 proxies*)</td>
<td>449,056</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>1/250$^{39}$</td>
<td>Trubetskoy et al., 2022$^{20}$</td>
<td>53,386</td>
<td>77,258</td>
</tr>
</tbody>
</table>

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).
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 co-localisation 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 ≥ 0.05) |
| 0 | 1 | 17 | 77 | 394 |
| 1 | 1 | 4 | 18 | 80 |
| 2 | 0 | 3 | 2 | 8 |

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; while 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 = fronto temporal 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 ($p < 8.26 \times 10^{-5}$; $0.05/605$). While some regions included genome-wide significant SNPs ($p < 5 \times 10^{-8}$) for one or both traits, 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; 6p22.1-21.34), 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).

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 S2 compares per-SNP p-values across trait pairs for comparisons with evidence of a relevant signal in both traits. Figure S3 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 S4 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$^{41-44}$ and CRHR1$^{45,46}$ which have been previously implicated in the traits we have analysed. These expansive credible sets reflect the strong LD in the region and indicate a signal that is difficult to localise (see Figure S3(F);
Table S3). The colocalisation analysis suggested independent variants for each trait despite many SNPs overlapping across their respective credible sets (see Figure S3). Sensitivity analysis showed robust support for the two independent variants hypothesis across shared-variant hypothesis priors (Figure S4).
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^{-6} (0.05/605). \( ^\Delta \) 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). \( ^\ddagger \) 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.

\[ H_0 = \text{no causal variant for either trait}, \ H_1 = \text{variant causal for trait } 1, \ H_2 = \text{variant causal for trait } 2, \ H_3 = \text{distinct causal variants for each trait}, \ H_4 = \text{a shared causal variant} \]

\[ P_1 = \text{posterior inclusion probability}. \ AD = \text{Alzheimer’s disease}, \ ALS = \text{amyotrophic lateral sclerosis}, \ FTD = \text{frontotemporal dementia}, \ PD = \text{Parkinson’s disease}, \ SZ = \text{schizophrenia}. \]

<table>
<thead>
<tr>
<th>Trait</th>
<th>Genomic position (GRCh37)</th>
<th>Local genetic correlation estimate (95% Confidence Interval)</th>
<th>Fine-mapping Credible set for trait</th>
<th>N SNPs</th>
<th>SNP with highest PIP for fine-mapping credible set (nearest gene; sense-strand base pair distance)</th>
<th>Posterior probability for hypothesis*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H0</td>
</tr>
<tr>
<td>AD</td>
<td>PD</td>
<td>Chr6:3257678-5-32639239 ( ^\Delta )</td>
<td>0.406 (0.197, 0.648)</td>
<td>1 1</td>
<td>rs9271247 (HLA-DQA1; +15,844)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>ALS</td>
<td>AD</td>
<td>Chr6:3262924-0-32682213 ( ^\ddagger )</td>
<td>0.974 (0.717, 1.000)</td>
<td>1 1</td>
<td>rs9275477 ( ^\ddagger ) (MTOC3P1; +1,260)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>ALS</td>
<td>FTD</td>
<td>Chr6:3220890-2-32454577 ( ^\ddagger )</td>
<td>0.723 (0.370, 1.000)</td>
<td>1 1</td>
<td>rs9268833 (HLA-DRB9; 0)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>ALS</td>
<td>SZ</td>
<td>Chr6:3220890-2-32454577 ( ^\ddagger )</td>
<td>-</td>
<td>1 1</td>
<td>rs9268833 (HLA-DRB9; 0)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>ALS</td>
<td>SZ</td>
<td>Chr12:569871-06-58748139</td>
<td>0.506 (0.218, 0.807)</td>
<td>1 1</td>
<td>rs113247976 (KIF5A; 0)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PD</td>
<td>SZ</td>
<td>Chr17:434605-01-44865832</td>
<td>0.595 (0.266, 0.950)</td>
<td>1 1</td>
<td>rs58879558 (MAPT; 0)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SZ</td>
<td>FTD</td>
<td>Chr6:3220890-2-32454577 ( ^\ddagger )</td>
<td>0.669 (0.379, 0.990)</td>
<td>1 1</td>
<td>rs9268219 (C6orf10; 0)</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>
Figure 3. Evidence for colocalisation between amyotrophic lateral sclerosis (ALS) and Alzheimer’s disease (AD) in the Chr6:32.63-32.68 Mb region

Panel A: SNP-wise p-value distribution between ALS and AD across Chr6:32.63-32.68 Mb, 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 predicted 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.
4. Discussion
We examined genetic overlaps between the neuropsychiatric conditions Alzheimer’s disease, amyotrophic lateral sclerosis, frontotemporal dementia, Parkinson’s disease, and schizophrenia. Associated genomic regions between pairs of traits were identified with local genetic correlation analysis and further analysed with statistical fine-mapping and colocalisation techniques.

Significant 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\textsuperscript{5,47}. HLA is particularly known for its role in immune response and it is implicated in various types of disease\textsuperscript{48,49}. 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\textsuperscript{48-57}. 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\textsuperscript{58}. Notably, 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\textsuperscript{59,60}. 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\textsuperscript{61}. 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\textsuperscript{57,62}.

Our analysis of this region gave stronger support for colocalisation between the ALS and AD GWAS than a previous study. The previous study defined a 100Kb window 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\textsuperscript{1}. The difference between these studies reflects differences in the processing of GWAS data; in this study all summary statistics underwent quality control to ensure only high-quality variants were retained.

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\textsuperscript{47}. 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 uncommon, and diagnostic tools available within the susieR package allow testing for inconsistencies between the reference and GWAS samples10. 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 discovery58,63.

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.
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/RO24804/1) 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–ESCRIAL. 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 was also provided by the King's College London DRIVE-Health Centre for Doctoral Training and the Perrin 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. 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.
References


30. Purcell S. PLINK. 1.9.0. Available from: http://pngu.mgh.harvard.edu/purcell/plink/
57. 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/10.1344/50006297918090122


64. 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.