Local genetic correlations exist among neurodegenerative and neuropsychiatric diseases

Regina H. Reynolds[1,2,#], Aaron Z. Wagen[1,3,4,#], Frida Lona-Durazo[5], Sonja W. Scholz[6,7], Maryam Shoai[8,10], John Hardy[2,8,10], Sarah A. Gagliano Taliun[5,9], Mina Ryten[1,2,11]

[#] **These authors contributed equally to this work.**

Author affiliations

1. Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London, London, UK
2. Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD 20815, United States
3. Department of Clinical and Movement Neurosciences, Queen Square Institute of Neurology, London, UK
4. Neurodegeneration Biology Laboratory, The Francis Crick Institute, London, UK
5. Montréal Heart Institute, Montréal, Québec, Canada
6. Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA
7. Department of Neurology, Johns Hopkins University Medical Center, Baltimore, MD, USA
8. Department of Neurodegenerative Diseases, Queen Square Institute of Neurology, University College London, London, UK
9. Department of Medicine & Department of Neurosciences, Université de Montréal, Montréal, Québec, Canada
10. UK Dementia Research Institute, University College London, London, UK
11. NIHR Great Ormond Street Hospital Biomedical Research Centre, University College London, London, UK

Correspondence to: Regina H. Reynolds (regina.reynolds.16@ucl.ac.uk) & Mina Ryten (mina.ryten@ucl.ac.uk)

Running title: Local genetic correlations in neurological traits

Abstract

Genetic correlation (r_g) between traits can offer valuable insight into underlying shared biological mechanisms. Neurodegenerative diseases overlap neuropathologically and often manifest comorbid neuropsychiatric symptoms. However, global r_g analyses show minimal r_g among neurodegenerative and neuropsychiatric diseases. Importantly, local r_gs can exist in the absence of global relationships. To investigate this possibility, we applied LAVA, a tool for local r_g analysis, to genome-wide association studies of 3 neurodegenerative diseases (Alzheimer's disease, Lewy body dementia and Parkinson's disease) and 3 neuropsychiatric disorders (bipolar disorder, major depressive disorder and schizophrenia). We identified several local r_gs missed in global analyses, including between (i) all 3 neurodegenerative diseases and schizophrenia and (ii) Alzheimer's and Parkinson's disease. For those local r_gs identified in genomic regions containing disease-implicated genes, such as *SNCA, CLU* and *APOE*, incorporation of expression quantitative trait loci identified genes that may drive genetic overlaps between diseases. Collectively, we demonstrate that complex genetic relationships exist among neurodegenerative and neuropsychiatric diseases, highlighting putative pleiotropic genomic regions and genes. These findings imply sharing of pathogenic processes and the potential existence of common therapeutic targets.

Abbreviations: AD = Alzheimer's disease; BIP = bipolar disorder; bp = base pair; CI = confidence interval; DLB = dementia with Lewy bodies; eQTL = expression quantitative loci; FDR = false discovery rate; GWAS = genome-wide association study; kb = kilobase; LAVA = local analysis of [co]variant annotation; LBD = Lewy body dementia; LD = linkage disequilibrium; LDSC = linkage disequilibrium score regression; MDD = major depressive disorder; PD = Parkinson's disease; SCZ = schizophrenia; SNP = single nucleotide polymorphism; UKBB = UK Biobank; ρ = rho; r_g = genetic correlation

Introduction

Neurodegenerative diseases are a group of syndromically-defined disorders that are characterised by the progressive loss of the structure and function of the central nervous system. They are typically grouped by their predominant neuropathological protein deposit (e.g. synucleinopathies, like Parkinson's disease and Lewy body dementia, by α-synuclein deposition and Alzheimer's disease by deposition of amyloid), but more often than not, they present with co-pathologies, suggesting that they might share common pathogenic pathways[1,2]. This notion is supported by genome-wide association studies (GWASs), which have (i) identified shared risk loci across neurodegenerative diseases, such as *APOE* and *BIN1* in Alzheimer's disease (AD) and Lewy body dementia (LBD), or *GBA*, *SNCA*, *TMEM175* in Parkinson's disease (PD) and LBD and (ii) demonstrated that genetic risk scores derived from one neurodegenerative disease can predict risk of another, as with AD and PD scores predicting risk of LBD[3–5]. The importance of identifying common pathogenic processes cannot be overstated, given the implications for our mechanistic understanding of these diseases as well as identification of common therapeutic targets benefitting a wider range of patients.

From a clinical perspective, neurodegenerative diseases are often also defined in terms of their predominant symptom (e.g. AD by memory impairment or PD by parkinsonism), but in reality, present as highly heterogenous diseases, with symptoms spanning multiple domains including neuropsychiatric symptoms[6,7]. Indeed, a higher prevalence of depression has been observed in individuals with dementia compared to those without dementia[8]. Furthermore, depression and anxiety are more common in individuals with PD compared to the general population, with clinically significant symptoms in 30-35% of patients[9,10]. A similar (albeit reversed) phenomenon has been observed in some neuropsychiatric disorders, with a higher risk of dementia diagnoses observed in individuals with schizophrenia (SCZ) versus individuals without a history of serious mental illness[11,12] and a higher risk of PD in individuals diagnosed with depressive disorder in mid or late life[10,13,14]. Together, these observations suggest the possibility of intersecting pathways between neurodegenerative and neuropsychiatric diseases.

79 Given these clinical and neuropathological overlaps, genetic overlaps would also be expected.
80 However, a study of global genetic correlation between neurological phenotypes demonstrated
81 limited overlap between individual neurodegenerative diseases as well as between
82 neurodegenerative diseases and neuropsychiatric disorders[15,16]. Genetic correlation (r_g) is a
83 frequently used measure of genetic overlap, which is traditionally studied on a genome-wide scale,
84 and thus, represents an average of the shared genetic effects across all causal loci in the genome[17].
85 This global approach may not capture shared genetic effects that are confined to particular regions
86 of the genome (i.e. local r_gs) or local r_gs that have opposing directions across the genome[15,17].
87 Indeed, local r_gs have been observed between neuropsychiatric traits[18] and between AD and PD
88 (specifically in the *HLA*[19] and *MAPT* loci[20]). In addition, previous work using approaches that can
89 detect polygenic overlap (including overlaps where there are mixed patterns of allelic effect
90 directions) have demonstrated a global polygenic overlap between neurodegenerative and
91 neuropsychiatric diseases, such as AD and BIP[21], AD and MDD[22,23], and PD and SCZ[24]. Collectively,
92 these studies indicate that local genetic overlaps likely exist between neurodegenerative and
93 neuropsychiatric diseases.

94 Here, we assess local r_g between 3 neurogenerative diseases (AD, LBD and PD) and 3
95 neuropsychiatric disorders (bipolar disorder, BIP; major depressive disorder, MDD; and SCZ). All 6
96 disease traits represent globally prevalent diseases[25], have reasonably large GWAS cohorts [3,5,26–30],
97 and importantly, have demonstrated evidence of a potential genetic overlap (but have not, to our
98 knowledge, been systematically assessed all together for local r_g). To estimate local r_g from GWAS
99 summary statistics, we used the recently developed tool local analysis of [co]variant annotation
100 (LAVA)[31]. Unlike existing tools, such as rho-HESS[32] and SUPERGNOVA[33], which only permit testing of
101 local r_gs between two traits, LAVA is additionally able to model local genetic relations using more
102 than two traits simultaneously, thus permitting exploration of local conditional genetic relations
103 between multiple traits (a particularly useful feature in the context of neurodegenerative diseases
104 like LBD, which has been hypothesised to lie on a disease continuum between AD and PD[5,34]). In

105 addition, we use data from blood- and brain-derived gene expression traits, in the form of

106 expression quantitative loci (eQTLs), to facilitate functional interpretation of local r_gs between

107 disease traits.

Results

Local analyses reveal genetic correlations among neurodegenerative and neuropsychiatric diseases

We applied LAVA to 3 neurodegenerative diseases (AD, LBD and PD) and 3 neuropsychiatric disorders (BIP, MDD and SCZ) (**Table 1**), all of which represent globally prevalent diseases[25]. Among neurodegenerative diseases, AD and PD are the most common, with a global prevalence of 8.98% and 1.12% in individuals > 70 years of age[6,25,35] and consequently, have large GWAS cohorts (AD, N cases = 71,880; PD, N cases = 33,674)[3,26]. LBD is the second most common dementia subtype after AD, affecting between 4.2-30% of dementia patients[36]. As such, the LBD GWAS cohort is small (N cases = 2,591), but unlike AD and PD neurodegenerative GWASs, 69% of the cohort is pathologically defined[5]. Among neuropsychiatric disorders, MDD is the second most prevalent, with an estimated 185 million people affected globally (equivalent to 2.49% of the general population), while BIP and SCZ have a prevalence of 0.53% and 0.32%, respectively[25]. Accordingly, all 3 disorders have large, well-powered GWASs (BIP, N cases = 41,917; MDD, N cases = 170,756; SCZ, N cases = 40,675)[28–30].

We tested pairwise local genetic correlations (r_gs) across a targeted subset of 300 local autosomal genomic regions that contain genome-wide significant GWAS loci from at least one trait (**Supplementary Figure 1**, **Supplementary Table 1**). These genomic regions, henceforth referred to as linkage disequilibrium (LD) blocks, were filtered from the original 2,495 LD blocks generated by Werme *et al.*[31] using a genome-wide partitioning algorithm that aims to reduce LD between LD blocks.

First, we performed a univariate test for every disease trait at each of the 300 LD blocks to ensure sufficient local genetic signal was present to proceed with bivariate local r_g analyses. Pairs of traits exhibiting a univariate local genetic signal of $p < 0.05/300$ were then carried forward to bivariate tests, resulting in 1,603 bivariate tests across 275 distinct LD blocks. Using a Bonferroni-corrected p-value threshold of $p < 0.05/1,603$, we detected 77 significant bivariate local r_gs across 59 distinct LD

133 blocks, with 25 local r_gs between trait pairs where no significant global r_g was observed (**Figure 1a**,
134 **Figure 1b**, **Supplementary Table 2**, **Supplementary Table 3**). These 25 correlations included: (i) local
135 r_gs between all 3 neurodegenerative diseases and SCZ; (ii) a local r_g between PD and BIP; and (ii) 20
136 local r_gs between AD and PD.For 30 of the 77 local r_gs, the genetic signal of both disease traits may
137 overlap entirely, suggested by the upper limit of the 95% confidence interval (CI) for explained
138 variance (i.e. r^2, the proportion of variance in genetic signal of one disease trait in a pair explained
139 by the other) including 1. Notably, the trait pairs where the upper limit of the 95% CI did not include
140 1 all involved at least one neurodegenerative disease, with the one exception being a local r_g
141 between PD and SCZ, suggesting that the genetic overlap between neurodegenerative diseases is
142 smaller than between neuropsychiatric disorders in the tested LD blocks (**Figure 1c**).

143 We intersected local r_gs with previously reported local genetic associations, which were derived
144 using tools for local r_g estimation (rho-HESS[19] and LAVA[18]) and estimation of global polygenic overlap
145 (the conditional/conjunctional FDR approach[23,24,37]). We found no overlap between local r_gs from
146 our study and local genetic associations reported using rho-HESS (which included an association
147 between AD and PD in the *HLA* locus) or the conditional/conjunctional FDR approach
148 (**Supplementary Note**, **Supplementary Table 4**). We did, however, demonstrate an overlap between
149 local r_gs from our study and local r_gs reported by Gerring *et al*., which also used LAVA, albeit to
150 study 10 psychiatric disorders and 10 substance abuse phenotypes. Between the two studies,
151 bivariate tests for the same trait pairs were performed in 6 LD blocks. We were able to replicate 5 of
152 the 7 overlapping local r_gs (BIP and SCZ in LD block 457; SCZ and BIP or MDD in LD block 951; MDD
153 and SCZ in LD block 952; and BIP and SCZ in LD block 2483; **Supplementary Note; Supplementary**
154 **Figure 2**; **Supplementary Table 4**).

Local analyses associate disease-implicated genomic regions with previously unrelated traits

Across the 77 local r_gs, 22 involved trait pairs where both traits had genome-wide significant single nucleotide polymorphisms (SNPs) overlapping the LD block tested, 35 involved trait pairs where one trait in the pair had genome-wide significant SNPs overlapping the LD block tested and 20 involved trait pairs where neither trait had genome-wide significant SNPs overlapping the LD block tested (**Figure 2a**). Thus, despite the targeted nature of our approach (which biased analyses towards LD blocks that contain genome-wide significant GWAS SNPs), 71% of the detected local r_gs linked genomic regions implicated by one of the six disease traits with seemingly unrelated disease traits.

For example, LD block 1719 (chr11:112,755,447-113,889,019) and 2281 (chr18:52,512,524-53,762,996) both contained genome-wide significant GWAS SNPs from MDD and SCZ, an overlap which was mirrored by a significant local r_g between MDD and SCZ (**Figure 2b**). In addition, both LD blocks implicated disease traits that did not have overlapping genome-wide significant GWAS SNPs in the region, indicating novel disease trait associations. These included (i) LBD in LD block 1719 (chr11:112,755,447-113,889,019), which negatively correlated with SCZ (ρ = -0.65, p = 4.72 x 10^{-6}) and (ii) AD and PD, which were positively correlated in LD block 2281 (chr18:52,512,524-53,762,996, ρ = 0.41, p = 1.24 x 10^{-8}). Notably, both LD blocks contain genes of interest to traits implicated by local r_g analyses, including *DRD2* in LD block 1719 (encodes dopamine receptor D2, a target of drugs used in both PD[7] and SCZ treatment[38]) and *RAB27B* in LD block 2281 (encodes Rab27b, a Rab GTPase recently implicated in α-synuclein clearance[39]).

Local r_g analyses also highlighted relationships between neurodegenerative traits in regions containing well-known, disease-implicated genes, such as: (i) *SNCA* (implicated in monogenic and sporadic forms of PD[3,5]) in LD block 681 (chr4:90,236,972-91,309,863), where a negative local r_g was observed between AD and PD (ρ = -0.41, p = 6.51 x 10^{-13}); (ii) *CLU* (associated with sporadic AD[26,40]) in LD block 1273 (chr8:27,406,512-28,344,176), where a positive local r_g was observed between AD

and PD (ρ = 0.36, p = 8.76 x 10^{-12}); and finally, (iii) *APOE* (E4 alleles associated with increased AD risk[41]) in LD block 2351 (chr19:45,040,933-45,893,307), where r_gs were observed between LBD and both AD and PD (LBD-AD: ρ = 0.59, p = 1.24 x 10^{-139}; LBD-PD: ρ = -0.29, p = 2.75 x 10^{-7}) (**Figure 2c**). We also noted a positive correlation between AD and PD in LD block 2128 (chr16:29,043,178-31,384,210), which contains the AD-associated *KAT8* locus[26] and the PD-associated *SETD1A* locus[3] (of note, rare loss-of-function variants in *SETD1A* are associated with schizophrenia[42]).

Sensitivity analysis indicates that by-proxy cases do not drive spurious local correlations among neurodegenerative diseases

Given concerns that UK Biobank (UKBB) by-proxy cases could potentially be mislabelled (i.e. parents of by-proxy case suffered from another type of dementia) and lead to spurious r_gs between neurodegenerative traits, we performed sensitivity analyses using GWASs for AD and PD that excluded UKBB by-proxy cases[27]. Of the 21 LD blocks where significant local r_gs were observed between the 3 neurodegenerative traits using AD and PD GWASs with by-proxy cases, only LD block 1273 (chr8:27,406,512-28,344,176) and LD block 2351 (chr19:45,040,933-45,893,307) had sufficient local genetic signal for both AD and PD without by-proxy cases. This likely reflects the decrease in cohort numbers when UKBB by-proxy cases are excluded from AD and PD GWASs (**Table 1**). We were able to replicate 2 of the 3 significant local r_gs observed in LD block 1273 (chr8:27,406,512-28,344,176) and 2351 (chr19:45,040,933-45,893,307), including the positive r_g between AD and PD in LD block 1273 and the positive r_g between AD and LBD in LD block 2351 (**Supplementary Figure 3**, **Supplementary Table 5**). Further, while the local r_g between LBD and PD in LD block 2351 was non-significant when using the PD GWAS without by-proxy cases, the correlation was in the same direction in the complementary analysis using the PD GWAS with by-proxy cases (no by-proxy: ρ = -0.201, ρ CI = -0.443 to 0.007, p = 0.061; by-proxy: ρ = -0.293, ρ CI = -0.405 to -0.184, p = 2.75 x 10^{-7}) (**Supplementary Figure 3**, **Supplementary Table 5**).

Local heritability of Lewy body dementia in an *APOE*-containing LD block is only partly explained by Alzheimer's disease and Parkinson's disease

Eleven LD blocks were associated with > 1 trait pair, of which 8 LD blocks had a trait in common across multiple trait pairs. In other words, the genetic component of one disease trait (the outcome trait) could be modelled using the genetic components of multiple predictor disease traits. This included 3 LD blocks (LD block 758, chr4:175,959,698-177,129,678; LD block 951, chr6:26,396,201-27,261,035; and LD block 952, chr6:27,261,036-28,666,364) where all 3 neuropsychiatric disorders were significantly correlated with one another, and thus could arguably be the outcome trait. In these situations, each neuropsychiatric disorder was separately modelled as the outcome trait, resulting in 3 independent models within each of these 3 LD blocks. The remaining 5 LD blocks only had 1 trait in common across correlated trait pairs, therefore only one model was constructed for each. A total of 14 multivariate models were run across all 8 LD blocks, of which 6 models were found to contain a predictor trait that significantly contributed to the local heritability of an outcome trait (**Figure 3a**, **Supplementary Table 6**).

We noted that all models with a neuropsychiatric outcome trait and a significant neuropsychiatric predictor trait had a high multivariate r^2 (range: 0.53-1), with upper confidence intervals including 1 (**Figure 3b**), suggesting that the genetic signal of the neuropsychiatric outcome trait could be entirely explained by its predictor traits in these LD blocks. In contrast, in the *APOE*-containing LD block 2351 (chr19:45,040,933-45,893,307), which was modelled with LBD as the outcome and AD and PD as predictors, the multivariate r^2 was 0.43 (95% CI: 0.38 to 0.5), a result that held using GWASs for AD and PD that excluded by-proxy cases (r^2 = 0.49, 95% CI: 0.44 to 0.57; **Supplementary Figure 3**). Thus, while AD and PD jointly explained approximately 40% of the local heritability of LBD, a proportion of the local heritability for LBD was independent of AD and PD.

Incorporation of gene expression traits to facilitate functional interpretation of disease trait correlations

To dissect whether regulation of gene expression might underlie local r_gs between disease traits, we performed local r_g analyses using expression quantitative trait loci (eQTLs) from eQTLGen[43] and PsychENCODE[44], which represent large human blood and brain expression datasets, respectively (**Table 1**). We used LAVA to study relationships between gene expression and disease traits on account of its ability to model the uncertainty in eQTL effect estimates (unlike the commonly used TWAS framework, which has been shown to have an increased type 1 error rate[45], as a result of it not accounting for the uncertainty in the estimated genetic component of gene expression). In addition, where three-way relationships were observed between 2 disease traits and an eQTL, we computed partial correlations to determine whether correlations between disease traits could be explained by the eQTL.

We restricted analyses to the 5 LD blocks highlighted in **Figure 2** (LD block 681, chr4:90,236,972-91,309,863; LD block 1273, chr8:27,406,512-28,344,176; LD block 1719, chr11:112,755,447-113,889,019; LD block 2281, chr18:52,512,524-53,762,996; LD block 2351, chr19:45,040,933-45,893,307), which contained genes of interest to at least one of the disease traits implicated by local r_g analyses. From these LD blocks of interest, we defined genic regions (gene start and end coordinates ± 100 kb) for all overlapping protein-coding, antisense or lincRNA genes (n = 92).

We detected a total of 135 significant bivariate local r_gs across 47 distinct genic regions (FDR < 0.05), with 43 local r_gs across 27 distinct genic regions between trait pairs involving a disease trait and a gene expression trait (**Supplementary Figure 4**, **Supplementary Table 7**). We noted that the explained variance (r^2) between trait pairs involving a disease trait and a gene expression trait tended to be lower than between trait pairs involving two disease traits (**Supplementary Figure 5**), an observation that aligns with a recent study that found only 11% of trait heritability to be mediated by bulk-tissue gene expression[46].

254 With the exception of the *SNCA*-containing LD block 681 (chr4:90,236,972-91,309,863), where eQTLs
255 for only 1 out of 5 genes tested in the block were correlated with a disease trait (negative r_g
256 between blood-derived *SNCA* eQTLs and PD), the expression of multiple genes was associated with
257 disease traits across the remaining LD blocks (**Figure 4a**). In addition, the expression of several genes
258 was associated with more than one disease trait (**Figure 4b**). For example, blood- and brain-derived
259 *ANKK1* eQTLs (*DRD2*-containing LD block 1719, chr11:112,755,447-113,889,019) were negatively
260 correlated with both MDD and SCZ, which themselves were positively correlated (**Figure 4c**). A SNP
261 residing in the coding region of *ANKK1* (rs1800497, commonly known as TaqIA SNP) has been
262 previously associated with alcoholism, schizophrenia and eating disorders, although it is unclear
263 whether this SNP exerts its effect via *DRD2* or *ANKK1*[47]. As *DRD2* is not expressed in blood, and
264 brain-derived *DRD2* eQTLs did not pass the univariate test for sufficient local genetic signal, we were
265 unable to test for local r_gs between *DRD2* eQTLs and any neuropsychiatric disorder. The data
266 available would therefore suggest that the shared risk of MDD and SCZ in the overlapping *ANKK1*
267 and *DRD2* genic regions may be partly driven by *ANKK1* gene expression. Indeed, the local r_g
268 between MDD and SCZ was weakened when conditioned on *ANKK1* eQTLs (eQTLGen: MDD~SCZ, r_g =
269 0.72, p = 0.000132; MDD~SCZ|*ANKK1*, r_g = 0.60, p = 0.0203; PsychENCODE: MDD~SCZ, r_g = 0.67, p =
270 0.000271; MDD~SCZ|*ANKK1*, r_g = 0.61, p = 0.00441; **Supplementary Table 9**).

271 A high degree of eQTL sharing across disease traits was observed in the *CLU*-containing LD block
272 1273 (chr8:27,406,512-28,344,176), with blood-derived eQTLs from 5 out of the 6 genes implicated
273 in local r_gs found to correlate with both AD and PD (**Figure 4b,d**). This included situations where
274 eQTL-disease trait correlations had (i) the same direction of effect across both disease traits (as
275 observed with *PBK, PNOC* and *SCARA5*) or (ii) opposing directions of effect across both disease traits
276 (as observed with *CLU* and *ESCO2*) (**Figure 4d**). Notably, while a significant positive local r_g was
277 observed between AD and PD in the *PBK* and *SCARA5* genic regions (reflecting the positive local r_g
278 observed between AD and PD across the entire LD block), no local r_g was observed between AD and

279 PD in the *CLU* genic region, suggesting that the shared risk of AD and PD in LD block 1273 may be
280 driven by the expression of genes other than the AD-associated *CLU* (**Figure 4e**). As a ferritin
281 receptor involved in ferritin internalisation, *SCARA5* could plausibly drive shared AD and PD risk,
282 given that cellular iron overload and iron-induced oxidative stress have been implicated in several
283 neurodegenerative diseases such as AD and PD[48,49]. However, conditioning the local r_g between AD
284 and PD on *SCARA5* eQTLs had little effect on the strength and significance of the correlation (AD~PD,
285 r_g = 0.16, p = 0.000135; AD~PD |*SCARA5*, r_g = 0.15, p = 0.000487; **Supplementary Table 9**),
286 suggesting that the relationship of *SCARA5* eQTLs with AD and PD operates independently across
287 each disease. In contrast, the local r_g between AD and PD was weakened and no longer significant
288 after conditioning on *PBK* eQTLs (AD~PD, r_g = 0.14, p = 0.0187; AD~PD |*PBK*, r_g = 0.07, p = 0.259;
289 **Supplementary Table 9**), indicating the *PBK* eQTLs may partly explain the local r_g between AD and
290 PD. *PBK* encodes a serine-threonine kinase involved in regulation of cellular proliferation and cell-
291 cycle progression[50], which has been shown to be overexpressed in proliferative cells, including
292 neural precursors cells in the subventricular zone of the adult brain[50,51].

293 Compared to LD block 1273, the degree of eQTL sharing across disease traits was lower in the *APOE*-
294 containing LD block 2351 (chr19:45,040,933-45,893,307), with eQTLs from 4 out of 16 genes
295 implicated in local r_gs found to correlate with AD and one of PD or LBD (**Figure 4b, f**). Shared eQTL
296 genes were only observed in blood and included *BCL3*, *CLPTM1, PVRL2* and *TOMM40*, with
297 expression of *BCL3* and *CLPTM1* positively correlating with AD and PD and expression of *PVRL2* and
298 *TOMM40* positively correlating with AD and LBD. As the exception, *PVR* eQTLs were negatively
299 associated with both AD and PD albeit in different tissues: AD in brain and PD in blood. Expression of
300 the remaining 11 genes was exclusively associated with either AD (n = 8) or PD (n = 3). No significant
301 local r_g was observed between *APOE* eQTLs and AD (FDR < 0.05), although a nominal positive r_g was
302 observed in blood (ρ = 0.178, ρ CI = 0.007 to 0.352, p = 0.039; **Supplementary Figure 4e**,
303 **Supplementary Table 7**). Overall, these results indicate that risk of neurodegenerative diseases (in
304 particular, AD) is associated with expression of multiple genes in the *APOE*-containing LD block.

305 Further, they add to a growing body of evidence suggesting that in parallel with the well-studied

306 *APOE-ε4* risk allele, there are additional *APOE*-independent risk factors in the region (such as *BCL3*[52]

307 and *PVRL2*[53]) that may contribute to AD risk.

308 For a complete overview of all genic regions tested across the 5 LD blocks of interest, see

309 **Supplementary Figure 4** and **Supplementary Table 7**.

310

Discussion

Despite clinical and neuropathological overlaps between neurodegenerative diseases, global analyses of genetic correlation (r_g) show minimal r_g among neurodegenerative diseases or across neurodegenerative and neuropsychiatric diseases. However, local r_gs can deviate from the genome-wide average estimated by global analyses and may even exist in the absence of a genome-wide r_g, thus motivating the use of tools to model local genetic relations.

Here, we applied LAVA to 3 neurodegenerative diseases and 3 neuropsychiatric disorders to determine whether local r_gs exist in a subset of 300 LD blocks that contain genome-wide significant GWAS loci from at least one of six investigated disease traits. We identified 77 significant bivariate local r_gs across 59 distinct LD blocks, with 25 local r_gs between trait pairs where no significant global r_g was observed, including between (i) all 3 neurodegenerative diseases and SCZ and (ii) AD and PD. Local r_gs highlighted expected associations (e.g. AD and LBD in the *APOE*-containing LD block 2351[5], chr19:45,040,933-45,893,307) and putative new associations (e.g. AD and PD in the *CLU*-containing LD block 1273, chr8:27,406,512-28,344,176) in genomic regions containing well-known, disease-implicated genes. Likewise, incorporation of eQTLs confirmed known relationships between diseases and genes, such as the association of AD with *CLU* expression[40] and PD with *SNCA* expression in blood[54], and revealed putative new disease-gene relationships. Together, these results indicate that more complex aetiological relationships exist between neurodegenerative and neuropsychiatric diseases than those revealed by global r_gs. Further, they highlight potential gene expression intermediaries that may account for local r_gs between disease traits.

These findings have important implications for our understanding of neurodegenerative diseases and the extent to which they overlap. An overlap between the synucleinopathies and AD is often acknowledged in the context of LBD, which has been hypothesised to lie on a disease continuum between AD and PD[5,34]. In support of this continuum, LBD was found to associate with both AD and PD in the *APOE*-containing LD block 2351 (chr19:45,040,933-45,893,307). Multiple regression

336 analyses confirmed that AD and PD were significant predictors of LBD heritability in LD block 2351.
337 Importantly, when AD and PD were modelled together, they explained only ~ 40% of the local
338 heritability of LBD in LD block 2351, implying that LBD represents more than the union of AD and PD.
339 Further, the associations of AD and PD with LBD had opposing regression coefficients, suggesting
340 that the contribution of AD and PD to LBD in the *APOE* locus may not be synergistic. This mirrors the
341 observation that genome-wide genetic risk scores of AD and PD do not interact in LBD,[5] and may
342 indicate that different biological pathways underlie the association between LBD and AD/PD. Indeed,
343 only blood-derived *PVRL2* and *TOMM40* eQTLs were found to correlate with both AD and LBD, while
344 no shared eQTL genes were detected between PD and LBD.

345 Less acknowledged is the genetic overlap between AD and PD, with no global r_g reported between
346 the two diseases[16,55] and no significant evidence for the presence of loci that increase the risk of
347 both diseases[56]. As the exception, genetic overlaps have been reported between AD and PD in the
348 *HLA*[19] and *MAPT* loci[20], hinting that pleiotropy may exist locally. In support of local pleiotropy, we
349 observed 20 local r_gs between AD and PD in genomic regions containing disease-implicated genes,
350 such as *SNCA* (LD block 681, chr4:90,236,972-91,309,863) and *CLU* (LD block 1273, chr8:27,406,512-
351 28,344,176). In the case of the *CLU*-containing LD block 1273 (chr8:27,406,512-28,344,176),
352 incorporation of eQTLs demonstrated an association of AD and PD with the expression of 5 genes,
353 although partial correlations suggested that only *PBK* expression could explain the correlation
354 between AD and PD, with the remaining associations between eQTLs and AD or PD appearing to
355 operate independently across diseases. In contrast, only blood-derived *SNCA* eQTLs were associated
356 with PD in LD block 681 (chr4:90,236,972-91,309,863), suggesting that the association between AD
357 and PD at the *SNCA* locus could be driven by tissue- or context-dependent gene expression or
358 alternatively other molecular phenotypes.

359 A few studies have demonstrated genetic overlaps between neurodegenerative and
360 neuropsychiatric diseases, such as AD and BIP[21], AD and MDD[22,23], and PD and SCZ[24], while others

361 have demonstrated no overlap[16,57], with divergences in outcomes ascribed to differences in
362 methodology and cohort[22]. Here, we observed a local r_g between BIP and PD, in addition to local r_gs
363 between schizophrenia and all 3 neurodegenerative diseases, which in the case of LBD was observed
364 in an LD block containing the gene *DRD2* (LD block 1719, chr11:112,755,447-113,889,019). Notably,
365 parkinsonism in dementia with Lewy bodies (DLB, one of the two LBDs), is often less responsive to
366 dopaminergic treatments than in PD[58]. Furthermore, methylation of the *DRD2* promoter in
367 leukocytes has been shown to differ between DLB and PD[59], while D2 receptor density has been
368 shown to be significantly reduced in the temporal cortex of DLB patients, but not AD[60], suggesting
369 that the *DRD2* locus may harbour markers that could distinguish between these neurodegenerative
370 diseases. Our study adds to the body of evidence in favour of a shared genetic basis between
371 neurodegenerative and neuropsychiatric diseases, although further work will be required to
372 determine whether this genetic overlap underlies the clinical and epidemiological links observed
373 between these two disease groups.

374 This study is not without its limitations, with several limitations related to the input data. These
375 limitations include: (i) the variability in cohort size (sample size is a key determinant of the power to
376 detect the association of a variant with a trait), which in the case of the smallest GWAS, LBD, may
377 explain the limited number of local r_gs observed involving this trait; (ii) the risk of misdiagnosis
378 (particularly in GWASs that include broader definitions of a disorder, such as the MDD GWAS, which
379 includes the UK Biobank broad definition of depression as well as clinically-derived phenotypes for
380 MDD); and (iii) the lack of genetic diversity (i.e. all traits used were derived from individuals of
381 European ancestry). Given that population-specific genetic risk factors exist, such as the lack of
382 *MAPT* GWAS signal in the largest GWAS of Asian patients with PD[61], and that transethnic global r_gs
383 between traits such as gene expression are significantly less than 1[62], it is imperative that studies of
384 local r_g are expanded to include diverse populations.

385 Among methodological limitations, both LDSC and LAVA only consider autosomal chromosomes,
386 leaving out chromosome X, which is not only longer than chromosome 8-22, but also encodes 858
387 and 689 protein-coding and non-coding genes, respectively (Ensembl v106)[63]. Furthermore, as
388 mentioned by the developers of LAVA[31], local r_gs could potentially be confounded by association
389 signals from adjacent genomic regions, a limitation which is particularly pertinent in our analysis of
390 gene expression traits where LD blocks were divided into smaller (often overlapping) genic regions.
391 Additional fine-mapping (both computational and biological) could be helpful in narrowing down the
392 set of potentially causal variants and consequently the genomic regions of interest[64].

393 Importantly, as with any genetic correlation analysis, an observed r_g does not guarantee the
394 presence of true pleiotropy. Spurious r_gs can occur due to LD or misclassification[17]. Here, we
395 attempted to address the potential misclassification of by-proxy cases via sensitivity analyses using
396 GWASs for AD and PD that excluded UKBB by-proxy cases. We replicated 2 of the 3 significant local
397 r_gs observed in 2 LD blocks when using GWASs with by-proxy cases. However, we were unable to
398 test for local r_gs across the remaining 19 LD blocks due to insufficient univariate signal, which could
399 reflect (i) a genuine contribution of by-proxy cases to trait h^2 in the region or (ii) a lack of statistical
400 power to detect a genetic signal. Given the substantial decrease in cohort numbers when UKBB by-
401 proxy cases are excluded from AD and PD GWASs (**Table 1**), a lack of statistical power seems the
402 more likely explanation, warranting a revisit of this analysis as clinically-diagnosed and/or
403 pathologically-defined cohorts increase in size.

404 Finally, even where observed $r_g s$ potentially represent true pleiotropy, LAVA cannot discriminate
405 between vertical and horizontal pleiotropy (also known as mediated and biological pleiotropy,
406 respectively[17,31]). Thus, while incorporation of gene expression can provide testable hypotheses
407 regarding the underlying genes and biological pathways that drive relationships between
408 neurodegenerative and neuropsychiatric diseases, experimental validation is required to establish
409 the extent to which these genes represent genuine intermediary phenotypes.

410 In summary, our results have important implications for our understanding of the genetic
411 architecture of neurodegenerative and neuropsychiatric diseases, including the demonstration of
412 local pleiotropy particularly between neurodegenerative diseases. Not only do these findings suggest
413 that neurodegenerative diseases may share common pathogenic processes, highlighting putative
414 gene expression intermediaries which may underlie relationships between diseases, but they also
415 infer the existence of common therapeutic targets across neurodegenerative diseases that could be
416 leveraged for the benefit of broader patient groups.

Materials and methods

Trait pre-processing

Summary statistics from a total of 8 distinct traits were used, including 6 disease traits and 2 gene expression traits. Disease traits included 3 neurodegenerative diseases (Alzheimer's disease, AD; Lewy body dementia, LBD; and Parkinson's disease, PD) and 3 neuropsychiatric disorders (bipolar disorder, BIP; major depressive disorder, MDD; and schizophrenia, SCZ)[3,5,26–30]. Gene expression traits were used to facilitate functional interpretation of local genetic correlations (r_g) between disease traits. Gene expression traits included expression quantitative trait loci (eQTLs) from eQTLGen[43] and PsychENCODE[44], which represent large human blood and brain expression datasets, respectively. All traits used were derived from individuals of European ancestry. Details of all summary statistics used can be found in **Table 1**.

Where necessary, SNP genomic coordinates were mapped to Reference SNP cluster IDs (rsIDs) using the SNPlocs.Hsapiens.dbSNP144.GRCh37 package[65]. In the case of the PD GWAS without UK Biobank (UKBB) data (summary statistics were kindly provided by the International Parkinson Disease Genomics Consortium), additional quality control filtering was applied, including removal of SNPs (i) with MAF < 1%, (ii) displaying an I^2 heterogeneity value of ≥ 80 and (iii) where the SNP was not present in at least 9 out of the 13 cohorts included in the meta-analysis.

Global genetic correlation analysis and estimation of sample overlaps

Across disease trait pairs, LD score regression (LDSC) was used to (i) estimate the observed-scale SNP heritability of each trait (which assumes a continuous liability, and thus may differ from liability-scale estimates of SNP heritability), (ii) determine the global r_g and (iii) estimate sample overlap[66,67]. All disease traits had significant SNP-based heritability (*Z*-score > 2) and met with the criteria suggested for reliable estimates of genetic correlation, which include: (i) heritability *Z*-score > 1.5 (optimal > 4), (ii) mean Chi square of test statistics > 1.02, and (iii) intercept estimated from SNP heritability

analysis is between 0.9-1.1[68] (**Supplementary Table 2**). We note that the heritability *Z*-score of LBD was 2.27, which is below the optimal suggested, and as such, can be expected to produce larger standard errors around estimates of global r_g.

Summary statistics for each trait were pre-processed using LDSC's `munge_sumstats.py` (https://github.com/bulik/ldsc/blob/master/munge_sumstats.py) together with HapMap Project Phase 3 SNPs[69]. For the LD reference panel, 1000 Genomes Project Phase 3 European population SNPs were used[70]. Both HapMap Project Phase 3 SNPs and European LD Scores from the 1000 Genomes Project are made available by the developers of LDSC[66,67] from the following repository: https://alkesgroup.broadinstitute.org/LDSCORE/ (see **Key resources** for details).

The estimated sample overlap was used as an input for LAVA, given that potential sample overlap between GWASs could impact estimated local r_gs[31]. Any shared variance due to sample overlap was modelled as a residual genetic covariance. As performed by Werme *et al.*[31], a symmetric matrix was constructed, with off-diagonal elements populated by the intercepts for genetic covariance derived from cross-trait LDSC and diagonal elements populated by comparisons of each trait with itself. This symmetric matrix was then converted to a correlation matrix. Importantly, it is not possible to estimate sample overlap with eQTL summary statistics, but given that the cohorts used in the GWASs were different from the cohorts included in the eQTL datasets, we assumed sample overlap between GWASs and eQTL datasets to be negligible. Thus, they were set to 0 in the correlation matrix. However, given the inclusion of GTEx samples in both eQTL datasets and our inability to estimate this overlap, downstream LAVA analyses were performed separately for each eQTL dataset.

Defining genomic regions for local genetic correlation analysis

Between disease traits

Genome-wide significant loci ($p < 5 \times 10^{-8}$) were derived from publicly available AD, BIP, LBD, MDD, PD and SCZ GWASs. Genome-wide significant loci were overlapped with linkage disequilibrium (LD)

467 blocks generated by Werme *et al.*[31] using a genome-wide partitioning algorithm. Briefly, each
468 chromosome was recursively split into blocks using (i) a break point to minimise LD between the
469 resulting blocks and (ii) a minimum size requirement. The resulting LD blocks represent
470 approximately equal-sized, semi-independent blocks of SNPs, with a minimum size requirement of
471 2,500 SNPs (resulting in an average block size of around 1Mb). Only those LD blocks containing
472 genome-wide significant GWAS loci from at least one trait were carried forward in downstream
473 analyses, resulting in a total of 300 autosomal LD blocks. Of the 22 possible autosomes, 21 contained
474 LD blocks with overlapping loci, with the highest number of LD blocks located in chromosome 1 and
475 6 (**Supplementary Figure 1**). LD block locations were in reference to build GRCh37 and are presented
476 in the format: LD block identifier, chromosome:start-end.

477 **Between disease and gene expression traits**

478 A total of 5 LD blocks, as highlighted by bivariate local r_g analysis of disease traits, were used in this
479 analysis (LD block 681, chr4:90,236,972-91,309,863; LD block 1273, chr8:27,406,512-28,344,176; LD
480 block 1719, chr11:112,755,447-113,889,019; LD block 2281, chr18:52,512,524-53,762,996; LD block
481 2351, chr19:45,040,933-45,893,307). From these LD blocks of interest, we defined genic regions for
482 all protein-coding, antisense or lincRNA genes that overlapped an LD block of interest. Genic regions
483 were defined as the start and end coordinates of a gene (Ensembl v87, GRCh37) with an additional
484 100 kb upstream and 100 kb downstream of gene start/end coordinates. We included a 100-kb
485 window as most lead *cis*-eQTL SNPs (i.e. the SNP with the most significant p-value in a SNP-gene
486 association) lie outside the gene start and end coordinates and are located within 100 kb of the gene
487 (in eQTLGen, 55% of lead-eQTL SNPs were outside the gene body and 92% were within 100 kb from
488 the gene[43]). These genic regions (n = 92) were carried forward in downstream analyses. Within these
489 genic regions, eQTL summary statistics were filtered to include all eQTLs (including both significant
490 and non-significant, with significance defined as FDR < 0.05) for the relevant gene (e.g. only SNP-
491 gene pairs that relate to *CLU* in the *CLU* genic region). This was to avoid SNP duplication, which can
492 occur where an eQTL SNP regulates more than one gene.

Estimating bivariate local genetic correlations

Between disease traits

The detection of valid and interpretable local r_g requires the presence of sufficient local genetic signal. For this reason, a univariate test was performed as a filtering step for bivariate local r_g analyses. Bivariate local r_g analyses were only performed for pairs of disease traits which both exhibited a significant univariate local genetic signal ($p < 0.05/300$, where the denominator represents the total number of tested LD blocks). This step resulted in a total of 1,603 bivariate tests spanning 275 distinct LD blocks. Bivariate results were considered significant when $p < 0.05/1603$.

Between disease and gene expression traits

For each genic region, only those disease traits that were found to have significant local r_g in the associated LD block were carried forward to univariate and bivariate analyses with eQTL summary statistics. As previously described, a univariate test was performed as a filtering step for bivariate local r_g analyses. Thus, bivariate local r_g analyses were only performed (i) if the gene expression trait (i.e. eQTL genes) exhibited a significant univariate local genetic signal and (ii) for pairs of traits (disease and gene expression) which both exhibited a significant univariate local genetic signal. A cut-off of $p < 0.05/92$ (the denominator represents the total number of tested genic regions) was used to determine univariate significance. A 100-kb window resulted in a total of 354 bivariate tests spanning 55 distinct genic regions. Bivariate results were corrected for multiple testing using two strategies: (i) a more lenient FDR correction and (ii) a more stringent Bonferroni correction ($p <$ 0.05/n_tests, where the denominator represents the total number of bivariate tests). We discuss results passing FDR < 0.05, but we make the results of both correction strategies available (**Supplementary Table 7, Supplementary Table 8**).

We evaluated the effect of window size on bivariate correlations by re-running all analyses using a 50-kb window. Following filtering for significant univariate local genetic signal (as described above), a total of 267 bivariate tests were run spanning 50 distinct genic regions. We detected 110

significant bivariate local r_gs (FDR < 0.05), 83 of which were also significant when using a 100-kb window (**Supplementary Figure 6**). We observed strong positive Pearson correlations in local r_g coefficient and p-value estimates across the two window sizes, indicating that our results are robust to the choice of window size (**Supplementary Figure 6**). Of note, p-value estimates between disease and gene expression traits tended to be lower when using the 50-kb window, as compared to the 100-kb window, as evidenced by the fitted line falling below the equivalent of y = x. This observation may be a reflection of stronger *cis*-eQTLs tending to have a smaller distance between SNP and gene[43]. In contrast, p-value estimates between two disease traits were comparable across the two window sizes.

Partial correlations were computed where three-way relationships were observed between 2 disease traits and an eQTL. The partial correlation reflects the correlation between 2 traits (e.g. disease X and Y) that can be explained by a third trait (e.g. eQTL, Z). Thus, a partial correlation approaching 0 suggests that trait Z captures an increasing proportion of the correlation between traits X and Y. Due to the three-way nature of the relationships, 3 possible conformations were possible (i.e. X~Y|Z, X~Z|Y and Y~Z|X); partial correlations were computed for all 3.

Local multiple regression

For LD blocks with significant bivariate local r_g between one disease trait and ≥ 2 disease traits, multiple regression was used to determine the extent to which the genetic component of the outcome trait could be explained by the genetic components of multiple predictor traits. These analyses permitted exploration of the independent effects of predictor traits on the outcome trait. A predictor trait was considered significant when p < 0.05.

Sensitivity analysis using by-proxy cases

As UK Biobank (UKBB) by-proxy cases could potentially be mislabelled (i.e. parent of by-proxy case suffered from another type of dementia) and lead to spurious r_gs between neurodegenerative traits,

we performed replication analyses using GWASs for AD[27] and PD that excluded UKBB by-proxy cases. LD blocks were filtered to include only those where significant bivariate local r_gs were observed between LBD and either by-proxy AD or by-proxy PD GWASs, in addition to between by-proxy AD and by-proxy PD GWASs. These criteria limited the number of LD blocks to 21. Bivariate local correlations were only performed for pairs of traits which both exhibited a significant univariate local genetic signal ($p < 0.05/21$, where the denominator represents the total number of tested loci), which resulted in a total of 10 bivariate tests spanning 6 distinct loci. We additionally performed multiple regression in LD block 2351 using LBD as the outcome and AD and PD (both excluding UKBB by-proxy cases) as predictors. A predictor trait was considered significant when $p < 0.05$.

R packages

All analyses were performed in R (v 4.0.5)[71]. As indicated in the accompanying GitHub repository (https://github.com/RHReynolds/neurodegen-psych-local-corr), all relevant packages were sourced from CRAN, Bioconductor (via BiocManager[72]) or directly from GitHub. Figures were produced using *circlize*, *ggplot2* and *ggraph*[73–75]. All open-source software used in this paper is listed in **Key resources**.

Code availability

Code used to pre-process GWASs, run genetic correlation analyses and to generate figures for the manuscript are available at: https://github.com/RHReynolds/neurodegen-psych-local-corr (doi:10.5281/zenodo.6587707). All other open-source software used in this paper is listed in **Key resources**.

Data availability

Analyses in this study relied on publicly available data, all of which are listed in **Key resources**. In the case of the PD GWAS without UK Biobank (UKBB) data, summary statistics were kindly provided by the International Parkinson Disease Genomics Consortium: https://pdgenetics.org/.

Key resources

Resource	Source/Reference	Identifier/URL
Deposited Data		
Ensembl GRCh37 v87	Ensembl genome browser[63]	http://ftp.ensembl.org/pub/grch37/current/gtf/homo_sapiens/
eQTLGen eQTLs	Vosã *et al.*, 2021	https://www.eqtlgen.org/cis-eqtls.html
GWAS, Alzheimer's disease (clinically diagnosed + UK Biobank proxy cases and controls)	Jansen *et al.*, 2019[26]	https://ctg.cncr.nl/software/summary_statistics
GWAS, Alzheimer's disease (clinically diagnosed)	Kunkle *et al.*, 2019[27]	https://www.niagads.org/igap-rv-summary-stats-kunkle-p-value-data
GWAS, Bipolar disease	Mullins *et al.*, 2021[28]	https://www.med.unc.edu/pgc/download-results/
GWAS, Lewy body dementia	Chia *et al.*, 2021[5]	https://www.ebi.ac.uk/gwas/studies/GCST90001390
GWAS, Parkinson's disease excluding 23andMe	Nalls *et al.*, 2019[3]	https://pdgenetics.org/resources
GWAS, Major depressive disorder	Howard *et al.*, 2019[29]	https://www.med.unc.edu/pgc/download-results/
GWAS, Schizophrenia	Pardiñas *et al.*, 2018[30]	https://www.med.unc.edu/pgc/download-results/
LAVA LD blocks	Werme *et al.*, 2021[31]	https://github.com/cadeleeuw/lava-partitioning
LDSC: HapMap Project Phase 3 SNPs	International HapMap 3 Consortium[69]	https://alkesgroup.broadinstitute.org/LDSCORE/; file name: `w_hm3_snplist`
LDSC: 1000 Genomes European LD Scores	1000 Genomes Project Consortium[70]	https://alkesgroup.broadinstitute.org/LDSCORE/; file name: `eur_w_ld_chr.tar.bz2`
PsychENCODE eQTLs	Wang *et al.*, 2018[44]	http://resource.psychencode.org/
Software		
Bioconductor		http://www.bioconductor.org; RRID:SCR_006442
BiocManager (v 1.30.16)	Morgan, 2021[72]	https://CRAN.R-project.org/package=BiocM

		anager
CRAN		http://cran.r-project.org/; RRID:SCR_003005
circlize (v 0.4.13)	Gu *et al.*, 2014[73]	https://github.com/jokergoo/circlize; RRID:SCR_002141
cowplot (v 1.1.1)	Wilke, 2020[76]	https://CRAN.R-project.org/package=cowplot; RRID:SCR_018081
data.table (v 1.14.2)	Dowle and Srinivasan, 2021[77]	https://CRAN.R-project.org/package=data.table
doSNOW (v 1.0.19)	Microsoft and Weston, 2020[78]	https://CRAN.R-project.org/package=doSNOW
foreach (v 1.5.1)	Microsoft and Weston, 2020[79]	https://CRAN.R-project.org/package=foreach
GenomicRanges (v 1.42.0)	Lawrence *et al.*, 2013[80]	https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html; RRID:SCR_000025
ggbeeswarm (v 0.6.0)	Clarke and Sherrill-Mix, 2017[81]	https://CRAN.R-project.org/package=ggbeeswarm
ggplot2 (v 3.3.5)	Wickham, 2016[74]	https://ggplot2.tidyverse.org; RRID:SCR_014601
ggpubr (v 0.4.0)	Kassambara, 2020[82]	https://CRAN.R-project.org/package=ggpubr; RRID:SCR_021139
ggraph (v 2.0.5)	Pedersen, 2021[75]	https://CRAN.R-project.org/package=ggraph; RRID:SCR_021239
gtools (v 3.9.2)	Warnes *et al.*, 2021[83]	https://CRAN.R-project.org/package=gtools
here (v 1.0.1)	Müller, 2020[84]	https://CRAN.R-project.org/package=here
janitor (v 2.1.0)	Firke, 2021[85]	https://CRAN.R-project.org/package=janitor
LAVA (v 0.0.6; commit #7be342)	Werme *et al.*, 2021[31]	https://github.com/josefin-werme/LAVA
LDSC (v 1.0.1)	Bulik-Sullivan *et al.*, 2015[86]	https://github.com/bulik/ldsc
openxlsx (v 4.2.4)	Schauberger and Walker, 2021[87]	https://CRAN.R-project.org/package=openxlsx; RRID:SCR_019185
qdapTools (v 1.3.5)	Rinker, 2015[88]	http://github.com/trinker/qdapTools
readxl (v 1.3.1)	Wickham and Bryan, 2019[89]	https://CRAN.R-project.org/package=readxl; RRID:SCR_018083
R (v 4.0.5)	R Core Team[71]	http://www.r-project.org/;

rtracklayer (v 1.50.0)	Lawrence *et al.*, 2009[90]	RRID:SCR_001905 https://bioconductor.org/packages/release/bioc/html/rtracklayer.html;
SNPlocs.Hsapiens.dbSNP144.GRCh37 (v 0.99.20)	Pagès, 2017[65]	RRID:SCR_021325 https://bioconductor.org/packages/release/data/annotation/html/SNPlocs.Hsapiens.dbSNP144.GRCh37.html
tidyverse (v 1.3.1)	Wickham et al., 2019[91]	https://www.tidyverse.org/; RRID:SCR_019186

573

Acknowledgements

We are grateful to Dr Cornelis Blauwendraat for feedback throughout this project. We would also like to thank the scientific community behind all the GWAS datasets, as well as data scientists developing the above-mentioned analysis tools, for making them publicly available and thus enabling the completion of this study.

Funding

This research was funded in whole or in part by Aligning Science Across Parkinson's [Grant numbers: ASAP-000478 and ASAP-000509] through the Michael J. Fox Foundation for Parkinson's Research (MJFF). For the purpose of open access, the author has applied a CC BY public copyright license to all Author Accepted Manuscripts arising from this submission.

AZW was supported through the award of a Clinical Research Fellowship funded by Eisai, Ltd and the Wolfson Foundation. SWS was supported in part by the Intramural Research Program of the U.S. National Institutes of Health (National Institute of Neurological Disorders and Stroke; project number: 1ZIANS003154). SAGT was supported by a Fonds de Recherche du Québec – Santé Junior 1 Award and by operational funds from the Institut de valorisation des données (IVADO). MR was supported through the award of a UKRI Medical Research Council Clinician Scientist Fellowship (MRC Grant Code: MR/N008324/1). JH was supported through the UKRI Medical Research Council (MRC Grant Code: MR/N026004/), the UK Dementia Research Institute, the Dolby Family Fund, and the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

Author contributions

RHR, SAGT and MR conceived and designed the study. RHR and AZW analysed data and drafted the figures. FLD, MS and SAGT consulted on the statistical analysis. SWS and MR provided clinical insight to data interpretation. RHR wrote the initial manuscript. All authors contributed to the critical analysis and revision of the manuscript.

Competing interests

AZW served as a medical monitor for Neuroscience Trials Australia, receiving no personal compensation. SWS serves on the scientific advisory council of the Lewy Body Dementia Association and receives grant support from Cerevel Therapeutics.

References

1. Robinson, J. L. *et al.* Neurodegenerative disease concomitant proteinopathies are prevalent, age-related and APOE4-associated. *Brain J. Neurol.* **141**, 2181–2193 (2018).
2. De Jager, P. L., Yang, H.-S. & Bennett, D. A. Deconstructing and targeting the genomic architecture of human neurodegeneration. *Nat. Neurosci.* **21**, 1310–1317 (2018).
3. Nalls, M. A. *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.* **18**, 1091–1102 (2019).
4. Guerreiro, R. *et al.* Investigating the genetic architecture of dementia with Lewy bodies: a two-stage genome-wide association study. *Lancet Neurol.* **17**, 64–74 (2018).
5. Chia, R. *et al.* Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nat. Genet.* **53**, 294–303 (2021).
6. Masters, C. L. *et al.* Alzheimer's disease. *Nat. Rev. Dis. Primer* **1**, 15056 (2015).
7. Poewe, W. *et al.* Parkinson disease. *Nat. Rev. Dis. Primer* **3**, 17013 (2017).
8. Kuring, J. K., Mathias, J. L. & Ward, L. Prevalence of Depression, Anxiety and PTSD in People with Dementia: a Systematic Review and Meta-Analysis. *Neuropsychol. Rev.* **28**, 393–416 (2018).
9. Reijnders, J. S. A. M., Ehrt, U., Weber, W. E. J., Aarsland, D. & Leentjens, A. F. G. A systematic review of prevalence studies of depression in Parkinson's disease. *Mov. Disord. Off. J. Mov. Disord. Soc.* **23**, 183–9; quiz 313 (2008).
10. Weintraub, D. *et al.* The neuropsychiatry of Parkinson's disease: advances and challenges. *Lancet Neurol.* **21**, 89–102 (2022).
11. Ribe, A. R. *et al.* Long-term risk of dementia in persons with schizophrenia: A danish population-based cohort study. *JAMA Psychiatry* **72**, 1095–1101 (2015).
12. Stroup, T. S. *et al.* Age-Specific Prevalence and Incidence of Dementia Diagnoses among Older US Adults with Schizophrenia. *JAMA Psychiatry* **78**, 632–641 (2021).

632 13. Gustafsson, H., Nordström, A. & Nordström, P. Depression and subsequent risk of Parkinson
633 disease A nationwide cohort study. *Neurology* **84**, 2422–2429 (2015).
634 14. Kazmi, H. *et al.* Late onset depression: Dopaminergic deficit and clinical features of prodromal
635 Parkinson's disease: A cross-sectional study. *J. Neurol. Neurosurg. Psychiatry* **92**, 158–164
636 (2021).
637 15. Bellou, E., Stevenson-Hoare, J. & Escott-Price, V. Polygenic risk and pleiotropy in
638 neurodegenerative diseases. *Neurobiol. Dis.* **142**, 104953 (2020).
639 16. Brainstorm Consortium *et al.* Analysis of shared heritability in common disorders of the brain.
640 *Science* **360**, eaap8757 (2018).
641 17. van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H. & Wray, N. R. Genetic correlations of
642 polygenic disease traits: from theory to practice. *Nat. Rev. Genet.* **20**, 567–581 (2019).
643 18. Gerring, Z. F., Thorp, J. G., Gamazon, E. R. & Derks, E. M. A Local Genetic Correlation Analysis
644 Provides Biological Insights Into the Shared Genetic Architecture of Psychiatric and Substance
645 Use Phenotypes. *Biol. Psychiatry* **92**, 583–591 (2022).
646 19. Stolp Andersen, M., Tan, M., Holtman, I. R., Hardy, J. & Pihlstrøm, L. Dissecting the limited
647 genetic overlap of Parkinson's and Alzheimer's disease. *Ann. Clin. Transl. Neurol.* 1–7 (2022)
648 doi:10.1002/acn3.51606.
649 20. Desikan, R. S. *et al.* Genetic overlap between Alzheimer's disease and Parkinson's disease at the
650 MAPT locus. *Mol. Psychiatry* **20**, 1588–1595 (2015).
651 21. Drange, O. K. *et al.* Genetic overlap between alzheimer's disease and bipolar disorder implicates
652 the MARK2 and VAC14 genes. *Front. Neurosci.* **13**, 1–11 (2019).
653 22. Lutz, M. W., Sprague, D., Barrera, J. & Chiba-Falek, O. Shared genetic etiology underlying
654 Alzheimer's disease and major depressive disorder. *Transl. Psychiatry* **10**, (2020).
655 23. Monereo-Sánchez, J. *et al.* Genetic Overlap Between Alzheimer's Disease and Depression
656 Mapped Onto the Brain. *Front. Neurosci.* **15**, 653130 (2021).

657 24. Smeland, O. B. *et al.* Genome-wide Association Analysis of Parkinson's Disease and
658 Schizophrenia Reveals Shared Genetic Architecture and Identifies Novel Risk Loci. *Biol.*
659 *Psychiatry* **89**, 227–235 (2021).
660 25. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204
661 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease
662 Study 2019. *Lancet Lond. Engl.* **396**, 1204–1222 (2020).
663 26. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways
664 influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
665 27. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk
666 loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
667 28. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases
668 provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
669 29. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent
670 variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–
671 352 (2019).
672 30. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes
673 and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
674 31. Werme, J., van der Sluis, S., Posthuma, D. & de Leeuw, C. A. An integrated framework for local
675 genetic correlation analysis. *Nat. Genet.* **54**, 274–282 (2022).
676 32. Shi, H., Mancuso, N., Spendlove, S. & Pasaniuc, B. Local Genetic Correlation Gives Insights into
677 the Shared Genetic Architecture of Complex Traits. *Am. J. Hum. Genet.* **101**, 737–751 (2017).
678 33. Zhang, Y. *et al.* SUPERGNOVA: local genetic correlation analysis reveals heterogeneous etiologic
679 sharing of complex traits. *Genome Biol.* **22**, 262 (2021).
680 34. Jellinger, K. A. Dementia with Lewy bodies and Parkinson's disease-dementia: current concepts
681 and controversies. *J. Neural Transm.* **125**, 615–650 (2018).
682 35. Poewe, W. *et al.* Parkinson disease. *Nat. Rev. Dis. Primer* **3**, 17013 (2017).

683 36. Erkkinen, M. G., Kim, M. & Geschwind, M. D. Clinical Neurology and Epidemiology of the Major
684 Neurodegenerative Diseases. *Cold Spring Harb. Perspect. Biol.* **10**, (2018).
685 37. Drange, O. K. *et al.* Genetic overlap between alzheimer's disease and bipolar disorder implicates
686 the MARK2 and VAC14 genes. *Front. Neurosci.* **13**, 1–11 (2019).
687 38. Urs, N. M., Peterson, S. M. & Caron, M. G. New Concepts in Dopamine D2Receptor Biased
688 Signaling and Implications for Schizophrenia Therapy. *Biol. Psychiatry* **81**, 78–85 (2017).
689 39. Underwood, R. *et al.* The GTPase Rab27b regulates the release, autophagic clearance, and
690 toxicity of α-synuclein. *J. Biol. Chem.* **295**, 8005–8016 (2020).
691 40. Foster, E. M., Dangla-Valls, A., Lovestone, S., Ribe, E. M. & Buckley, N. J. Clusterin in Alzheimer's
692 Disease: Mechanisms, Genetics, and Lessons From Other Pathologies. *Front. Neurosci.* **13**, 164
693 (2019).
694 41. Bertram, L., McQueen, M. B., Mullin, K., Blacker, D. & Tanzi, R. E. Systematic meta-analyses of
695 Alzheimer disease genetic association studies: The AlzGene database. *Nat. Genet.* **39**, 17–23
696 (2007).
697 42. Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and
698 developmental disorders. *Nat. Neurosci.* **19**, 571–577 (2016).
699 43. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and
700 polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
701 44. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the
702 human brain. *Science* **362**, (2018).
703 45. Leeuw, C. de, Werme, J., Savage, J. E., Peyrot, W. J. & Posthuma, D. Reconsidering the validity of
704 transcriptome-wide association studies. 2021.08.15.456414 Preprint at
705 https://doi.org/10.1101/2021.08.15.456414 (2022).
706 46. Yao, D. W., O'Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease
707 mediated by assayed gene expression levels. *Nat. Genet.* **52**, 626–633 (2020).

708 47. Ponce, G. *et al.* The ANKK1 kinase gene and psychiatric disorders. *Neurotox. Res.* **16**, 50–59
709 (2009).
710 48. Li, J. Y. *et al.* Scara5 Is a Ferritin Receptor Mediating Non-Transferrin Iron Delivery. *Dev. Cell* **16**,
711 35–46 (2009).
712 49. Crielaard, B. J., Lammers, T. & Rivella, S. Targeting iron metabolism in drug discovery and
713 delivery. *Nat. Rev. Drug Discov.* **16**, 400–423 (2017).
714 50. Han, Z., Li, L., Huang, Y., Zhao, H. & Luo, Y. PBK/TOPK: A Therapeutic Target Worthy of Attention.
715 *Cells* **10**, 371 (2021).
716 51. Dougherty, J. D. *et al.* PBK/TOPK, a Proliferating Neural Progenitor-Specific Mitogen-Activated
717 Protein Kinase Kinase. *J. Neurosci.* **25**, 10773–10785 (2005).
718 52. Nho, K. *et al.* Association analysis of rare variants near the APOE region with CSF and
719 neuroimaging biomarkers of Alzheimer's disease. *BMC Med. Genomics* **10**, (2017).
720 53. Zhou, X. *et al.* Non-coding variability at the APOE locus contributes to the Alzheimer's risk. *Nat.*
721 *Commun.* **10**, 1–16 (2019).
722 54. Locascio, J. J. *et al.* Association between α-synuclein blood transcripts and early, neuroimaging-
723 supported Parkinson's disease. *Brain* **138**, 2659–2671 (2015).
724 55. Guerreiro, R. *et al.* Genome-wide analysis of genetic correlation in dementia with Lewy bodies,
725 Parkinson's and Alzheimer's diseases. *Neurobiol. Aging* **38**, 214.e7-214.e10 (2016).
726 56. Moskvina, V. *et al.* Analysis of genome-wide association studies of alzheimer disease and of
727 parkinson disease to determine if these 2 diseases share a common genetic risk. *JAMA Neurol.*
728 **70**, 1268–1276 (2013).
729 57. Gibson, J. *et al.* Assessing the presence of shared genetic architecture between Alzheimer's
730 disease and major depressive disorder using genome-wide association data. *Transl. Psychiatry* **7**,
731 (2017).
732 58. McKeith, I. G. *et al.* Diagnosis and management of dementia with Lewy bodies: Fourth consensus
733 report of the DLB Consortium. *Neurology* **89**, 88–100 (2017).

734 59. Ozaki, Y. *et al.* DRD2 methylation to differentiate dementia with Lewy bodies from Parkinson's
735 disease. *Acta Neurol. Scand.* **141**, 177–182 (2020).
736 60. Piggott, M. A. *et al.* Selective loss of dopamine D2 receptors in temporal cortex in dementia with
737 Lewy bodies, association with cognitive decline. *Synapse* **61**, 903–911 (2007).
738 61. Foo, J. N. *et al.* Identification of Risk Loci for Parkinson Disease in Asians and Comparison of Risk
739 Between Asians and Europeans: A Genome-Wide Association Study. *JAMA Neurol.* **77**, 746–754
740 (2020).
741 62. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price,
742 A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am. J.*
743 *Hum. Genet.* **99**, 76–88 (2016).
744 63. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
745 64. Broekema, R. V., Bakker, O. B. & Jonkers, I. H. A practical view of fine-mapping and gene
746 prioritization in the post-genome-wide association era. *Open Biol.* **10**, 190221.
747 65. Pagès, H. SNPlocs.Hsapiens.dbSNP144.GRCh37: SNP locations for Homo sapiens (dbSNP Build
748 144). *R Package Version 09920* (2017).
749 66. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in
750 genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
751 67. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat.*
752 *Genet.* **47**, 1236–1241 (2015).
753 68. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression
754 that maximizes the potential of summary level GWAS data for SNP heritability and genetic
755 correlation analysis. *Bioinforma. Oxf. Engl.* **33**, 272–279 (2017).
756 69. International HapMap 3 Consortium *et al.* Integrating common and rare genetic variation in
757 diverse human populations. *Nature* **467**, 52–8 (2010).
758 70. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature*
759 **526**, 68–74 (2015).

760 71. R Core Team. R: A language and environment for statistical computing. Preprint at (2021).
761 72. Morgan, M. BiocManager: Access the Bioconductor Project Package Repository. Preprint at
762 (2021).
763 73. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular
764 visualization in R. *Bioinforma. Oxf. Engl.* **30**, 2811–2 (2014).
765 74. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.* (Springer-Verlag New York, 2016).
766 75. Pedersen, T. L. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks.
767 Preprint at (2021).
768 76. Wilke, C. O. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. Preprint at
769 (2020).
770 77. Dowle, M. & Srinivasan, A. data.table: Extension of `data.frame`. Preprint at (2021).
771 78. Corporation, M. & Weston, S. doSNOW: Foreach Parallel Adaptor for the 'snow' Package.
772 Preprint at (2020).
773 79. Microsoft & Weston, S. foreach: Provides Foreach Looping Construct. Preprint at (2020).
774 80. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.*
775 **9**, e1003118 (2013).
776 81. Clarke, E. & Sherrill-Mix, S. ggbeeswarm: Categorical Scatter (Violin Point) Plots. Preprint at
777 (2017).
778 82. Kassambara, A. ggpubr: 'ggplot2' Based Publication Ready Plots. Preprint at (2020).
779 83. Warnes, G. R., Bolker, B. & Lumley, T. gtools: Various R Programming Tools. Preprint at (2021).
780 84. Müller, K. here: A Simpler Way to Find Your Files. Preprint at (2020).
781 85. Firke, S. janitor: Simple Tools for Examining and Cleaning Dirty Data. Preprint at (2021).
782 86. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat.*
783 *Genet.* **47**, 1236–1241 (2015).
784 87. Schauberger, P. & Walker, A. openxlsx: Read, Write and Edit xlsx Files. Preprint at (2021).
785 88. Rinker, T. W. {qdapTools}: Tools to Accompany the qdap Package. Preprint at (2015).

786 89. Wickham, H. & Bryan, J. readxl: Read Excel Files. Preprint at (2019).

787 90. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: An R package for interfacing with genome

788 browsers. *Bioinformatics* **25**, 1841–1842 (2009).

789 91. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

790

Figures

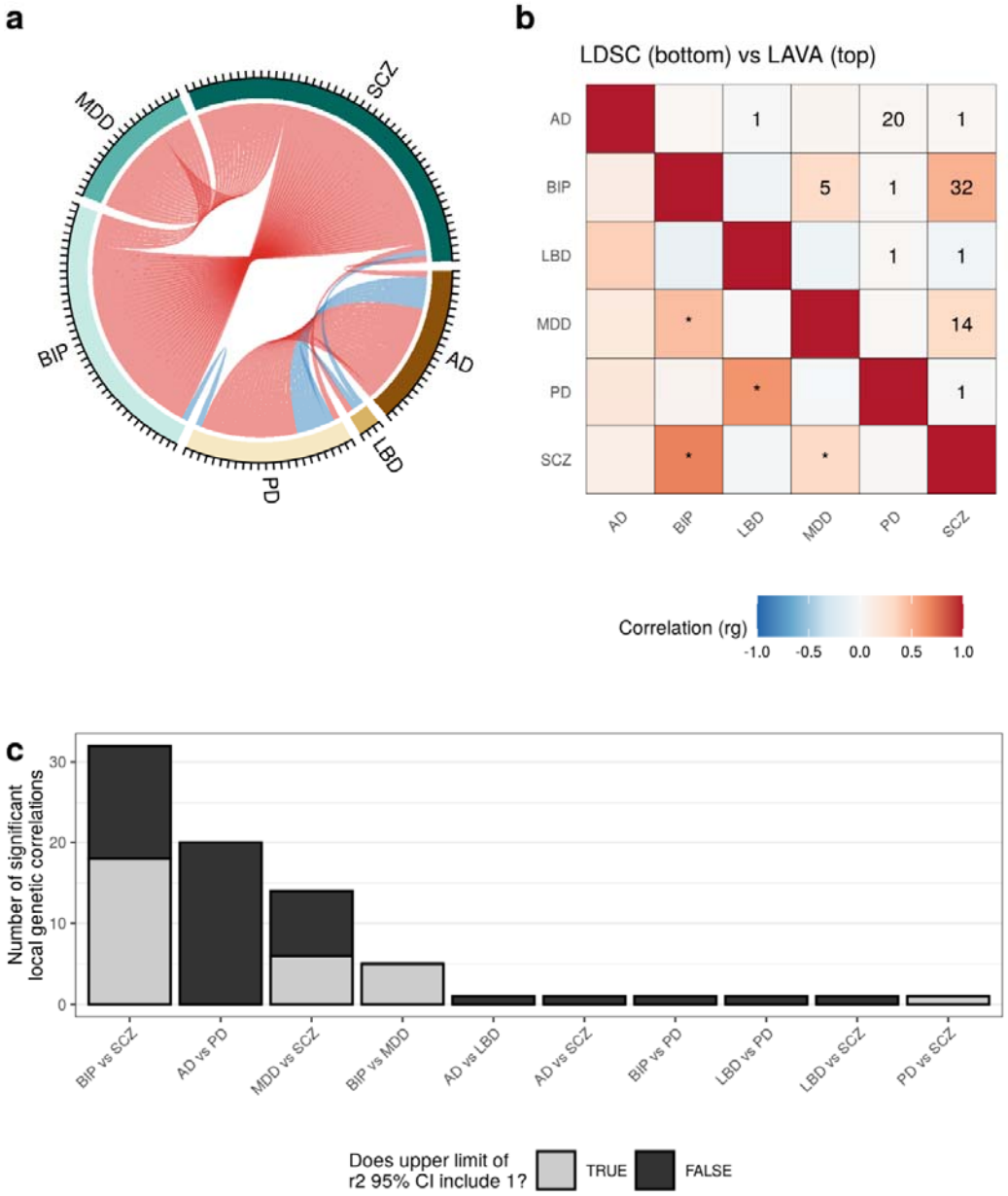


Figure 1 Overview of local and global genetic correlations between neurodegenerative diseases and neuropsychiatric disorders.
(a) Chord diagram showing the number of significant bivariate local s ($p < 0.05/1603$) between each of the disease traits across all LD blocks. Positive and negative correlations are coloured red and blue, respectively. **(b)** Comparison between the global s estimated by LDSC (bottom) and the mean local from LAVA (top) across all tested LD blocks. Significant global s ($p < 0.05/15$) are indicated with *. The number of significant local s is indicated by a number in each tile. **(c)** Bar plot showing the number of significant local s between disease trait pairs. The fill of the bars indicates the number of significant LD blocks for which the upper limit of the 95% confidence interval (CI) included 1.

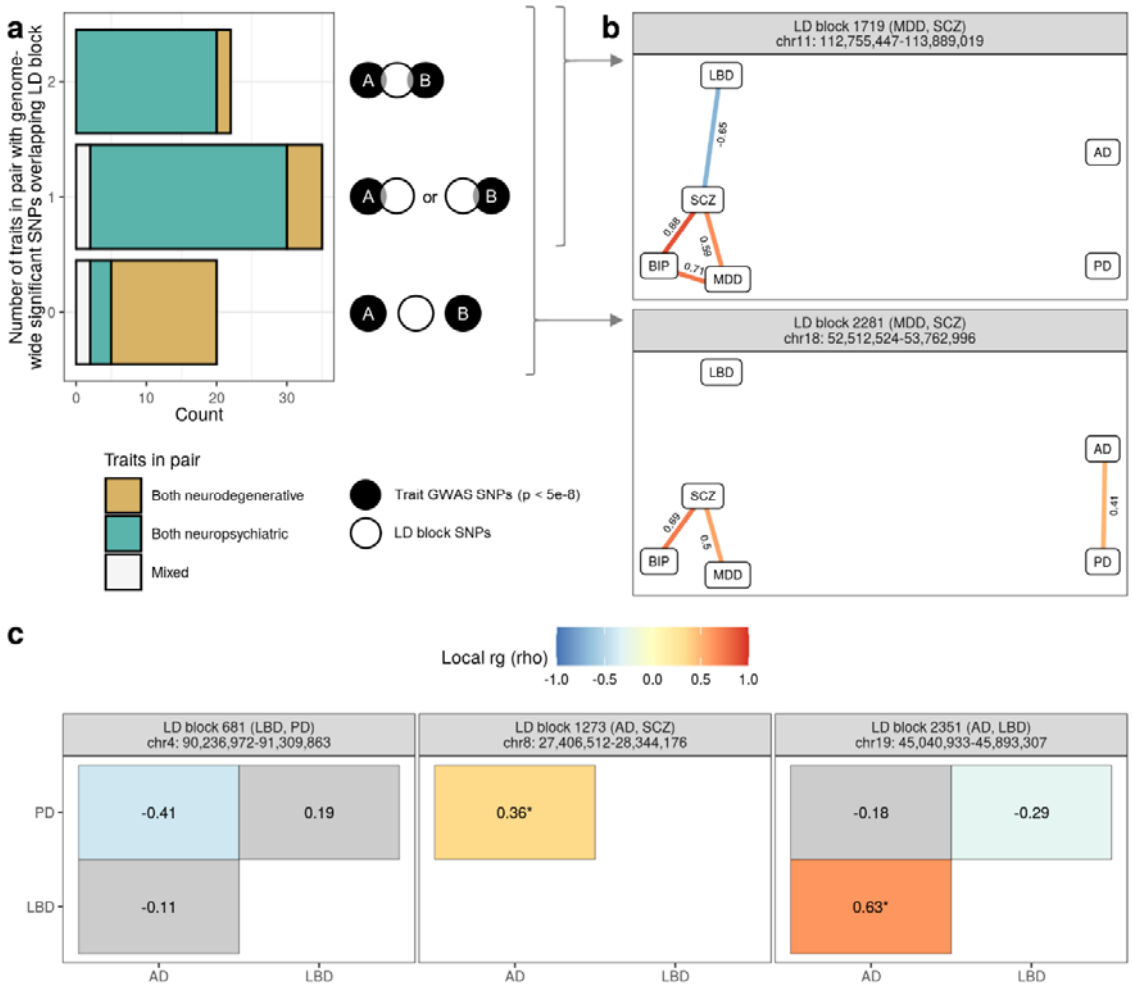


Figure 2 Local analyses associate disease-implicated genomic regions with previously unrelated traits.

(a) Bar plot (left) showing the number of traits within trait pairs demonstrating significant local s that had genome-wide significant SNPs overlapping the tested LD block (as illustrated by the schematic on the right). **(b)** Two LD blocks illustrating the situations depicted in **(a)**. Edge diagrams for each LD block show the standardised coefficient for (rho, ρ) for each significant bivariate local . Significant negative and positive s are indicated by blue and red colour, respectively. **(c)** Heatmaps show the rho for each bivariate local within the LD block. Asterisks (*) indicate s that were replicated when using AD and PD GWASs that excluded UK Biobank by-proxy cases. Significant negative and positive s are indicated by blue and red fill, respectively. Non-significant s have a grey fill. In both **(b)** and **(c)** panels are labelled by the LD block identifier, the traits with genome-wide significant SNPs overlapping the LD block (indicated in the brackets) and the genomic coordinates of the LD block (in the format chromosome:start-end, GRCh37).

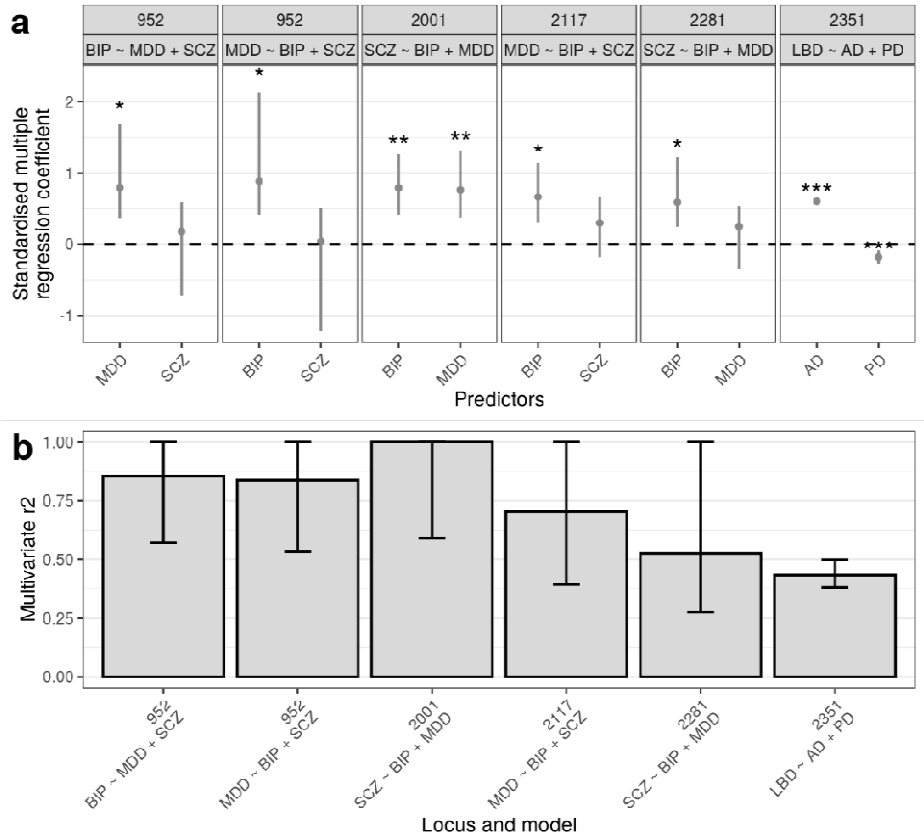
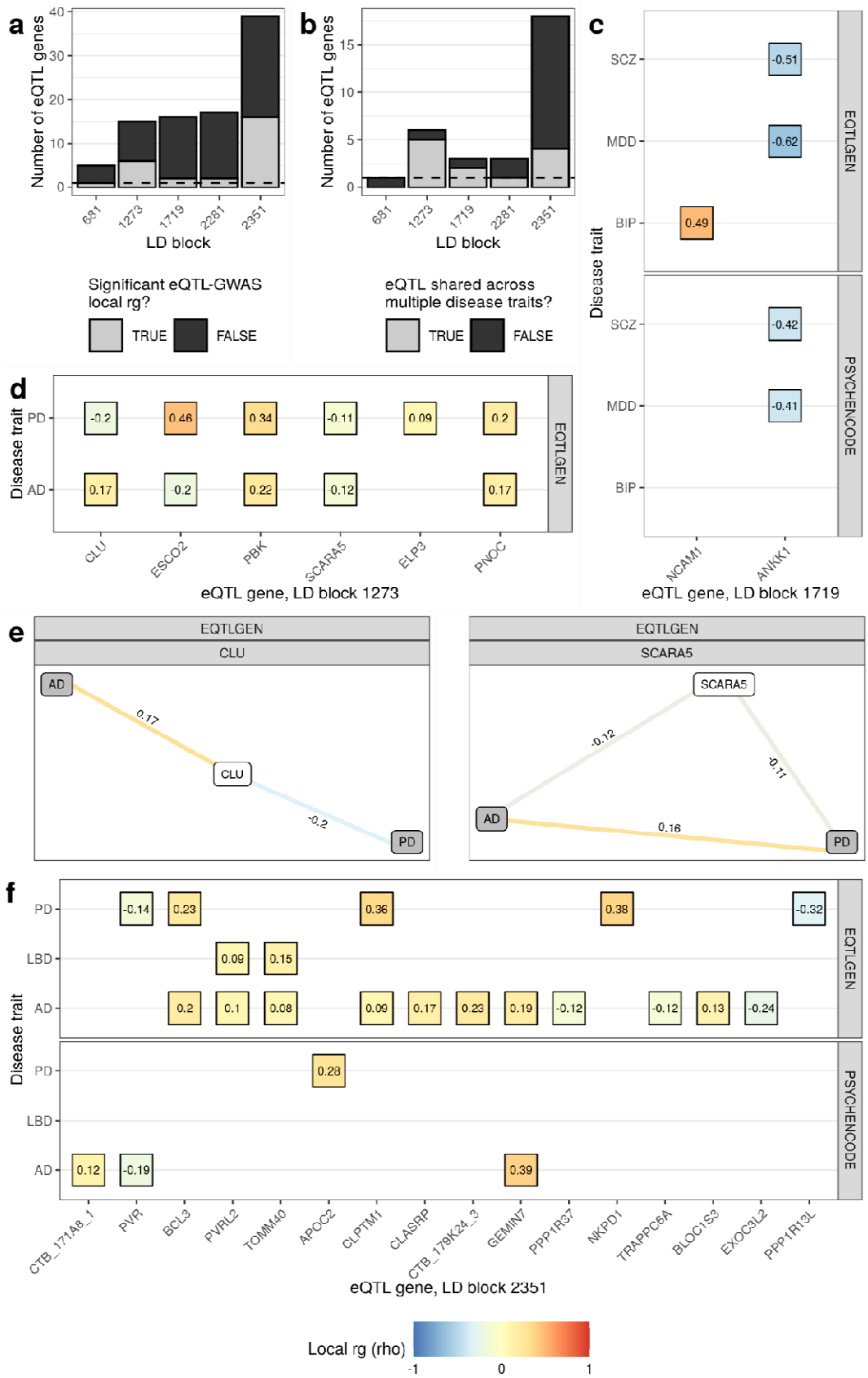


Figure 3 Multiple regression across LD blocks with multiple trait pair correlations.

For both plots, only those multiple regression models with at least one significant predictor ($p < 0.05$) are shown. **(a)** Plots of standardised coefficients for each predictor in multiple regression models across each LD block, with whiskers spanning the 95% confidence interval for the coefficients. Panels are labelled by the LD block identifier and the regression model. **(b)** Multivariate for each LD block and model, where multivariate represents the proportion of variance in genetic signal for the outcome trait explained by all predictor traits simultaneously. Whiskers span the 95% confidence interval for the . ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$. Coordinates for LD blocks (in the format chromosome:start-end, GRCh37): 952, chr6:27,261,036-28,666,364; 2001, chr14:71,140,427-72,665,319; 2117, chr16:12,793,150-13,893,407; 2281, chr18:52,512,524-53,762,996; 2351, chr19:45,040,933-45,893,307.


a
Number of eQTL genes
LD block
Significant eQTL-GWAS local rg?
TRUE
FALSE
b
eQTL shared across multiple disease traits?
c
Disease trait
SCZ
MDD
BIP
EQTLGEN
PSYCHENCODE
eQTL gene, LD block 1719
d
PD
AD
CLU
ESCO2
PBK
SCARA5
ELP3
PNOC
eQTL gene, LD block 1273
e
f
LBD
CTB_171A8_1
PVR
BCL3
PVRL2
TOMM40
APOC2
CLPTM1
CLASRP
CTB_179K24_3
GEMIN7
PPP1R37
NKPD1
TRAPPC6A
BLOC1S3
EXOC3L2
PPP1R13L
eQTL gene, LD block 2351
Local rg (rho)

827

Figure 4 Incorporation of gene expression traits to facilitate functional interpretation of disease trait correlations.

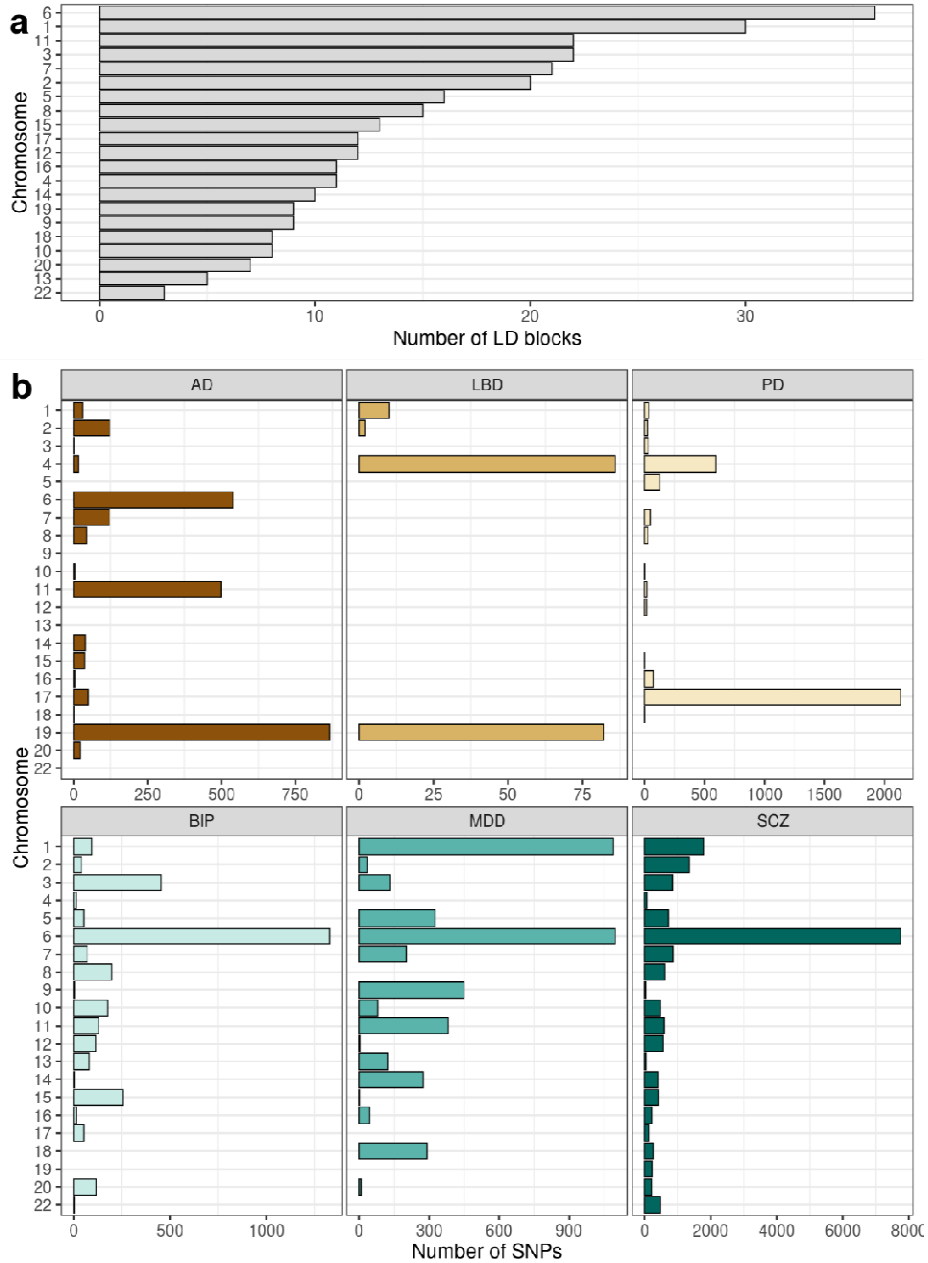
(a) Bar plot of the number of eQTL genes (as defined by their genic regions) tested in each LD block. The fill of the bars indicates whether eQTL genes were significantly correlated with at least one disease trait. **(b)** Bar plot of the number of eQTL genes that were significantly correlated with at least one disease trait. The fill of the bars indicates whether eQTL genes in local r_gs were correlated with one or more disease traits. **(c, d, f)** Heatmaps of the standardised coefficient for r_g (rho) for each significant gene expression-disease trait correlation (FDR < 0.05) within LD block **(c)** 1719, **(d)** 1273 and **(f)** 2351. Genes are ordered left to right on the x-axis by the genomic coordinate of their gene start. Panels are labelled by the eQTL dataset from which eQTL genes were derived (either PsychENCODE's analysis of adult brain tissue from 1387 individuals or the eQTLGen meta-analysis of 31,684 blood samples from 37 cohorts). **(e)** Edge diagrams for representative genic regions show the rho for each significant bivariate local r_g (FDR < 0.05). GWAS and eQTL nodes are indicated by grey and white fill, respectively. Panels are labelled by the gene tested and the eQTL dataset from which eQTL genes were derived. In panels **c-f** significant negative and positive r_gs are indicated by blue and red colour, respectively. Coordinates for LD blocks (in the format chromosome:start-end, GRCh37): 681, chr4:90,236,972-91,309,863; 1273, chr8:27,406,512-28,344,176; 1719, chr11:112,755,447-113,889,019; 2281, chr18:52,512,524-53,762,996; 2351, chr19:45,040,933-45,893,307.

Tables

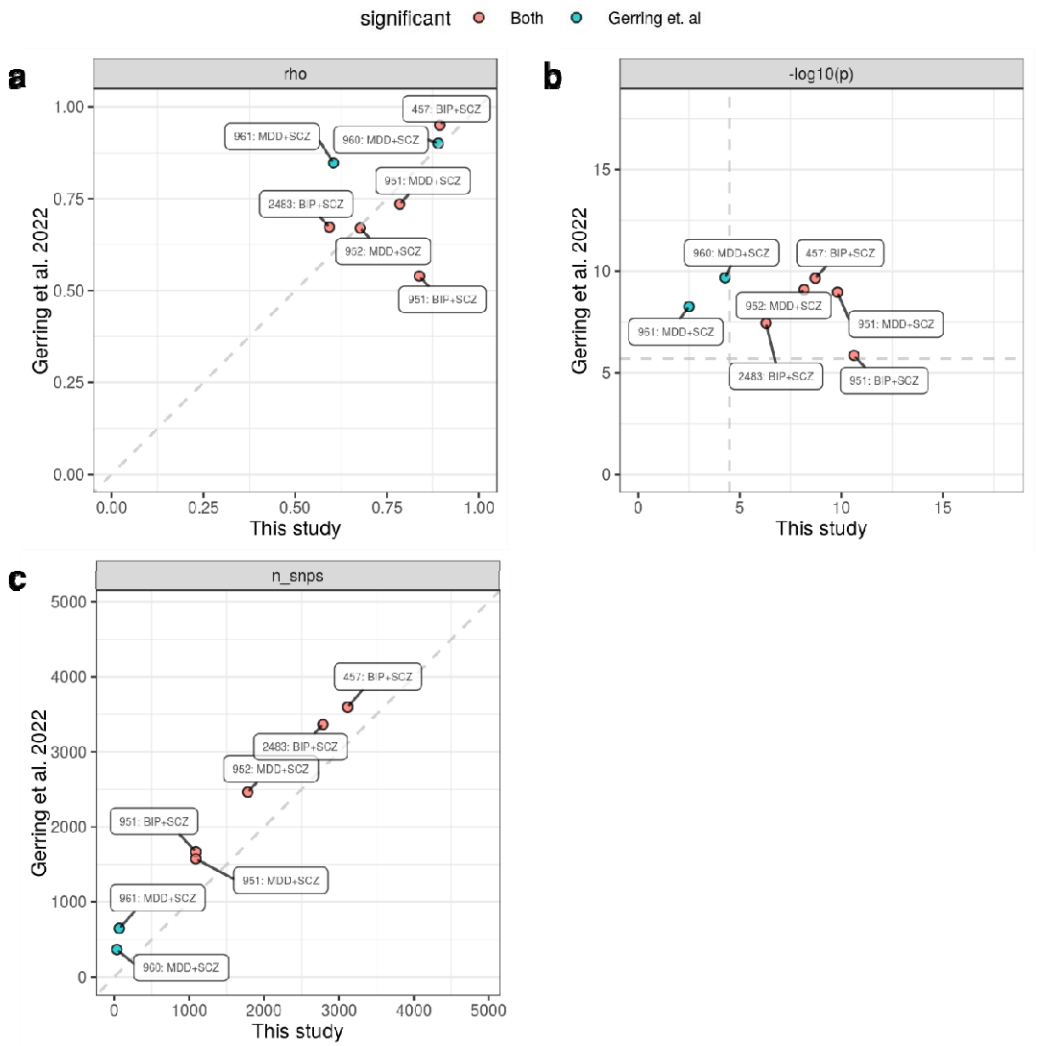
Table 1 Overview of traits included in this study. Global observed-scale SNP heritability (h^2) for each trait computed using LDSC[86] assuming a continuous liability (which may differ from liability scale h^2 estimates). The h^2 Z-score was calculated by dividing the global h^2 estimate by its standard error (SE).

Trait type	Trait	Abbreviation	N	N cases	N controls	Global h^2 (SE)	h^2 *Z*-score	Original study
Disease	Alzheimer's disease Clinically diagnosed + UK Biobank proxy cases and controls	AD	455,258	71,880 (46,613 proxy)	383,378 (318,246 proxy)	1.5% (0.2)	7.14	Jansen *et al.*, 2019[26]
Disease	Alzheimer's disease Clinically diagnosed	AD (no proxy)	63,926	21,982	41,944	7.1% (1.1)	6.25	Kunkle *et al.*, 2019[27]
Disease	Bipolar disorder	BIP	413,466	41,917	371,549	7.1% (0.3)	26.2	Mullins *et al.*, 2021[28]
Disease	Lewy body dementia Autopsy-confirmed + clinically diagnosed	LBD	6,618	2,591 (1,789 autopsy-confirmed)	4,027	17.1% (7.6)	2.27	Chia *et al.*, 2021[5]
Disease	Major depressive disorder excluding 23andMe Clinically diagnosed (33 PGC cohorts) + UK Biobank broad depression phenotype	MDD	500,199	170,756	329,443	6% (0.2)	26	Howard *et al.*, 2019[29]
Disease	Parkinson's disease excluding 23andMe Clinically diagnosed + UK Biobank proxy cases and controls	PD	482,730	33,674 (18,618 proxy)	449,056 (436,419 proxy)	1.9% (0.2)	10.9	Nalls *et al.*, 2019[3]
Disease	Parkinson's disease excluding 23andMe Clinically diagnosed	PD (no proxy)	27,693	15,056	12,637	30.6% (2.8)	11.1	Nalls *et al.*, 2019[3]
Disease	Schizophrenia	SCZ	105,318	40,675	64,643	41% (1.4)	29.7	Pardiñas *et al.*, 2018[30]
Gene expression	eQTLGen Blood-derived eQTLs	eQTLGEN	31,684	-	-	-		Vosã *et al.*, 2021[43]
Gene expression	PsychENCODE Brain-derived eQTLs	PSYCHENCODE	1,387	-	-	-		Wang *et al.*, 2018[44]

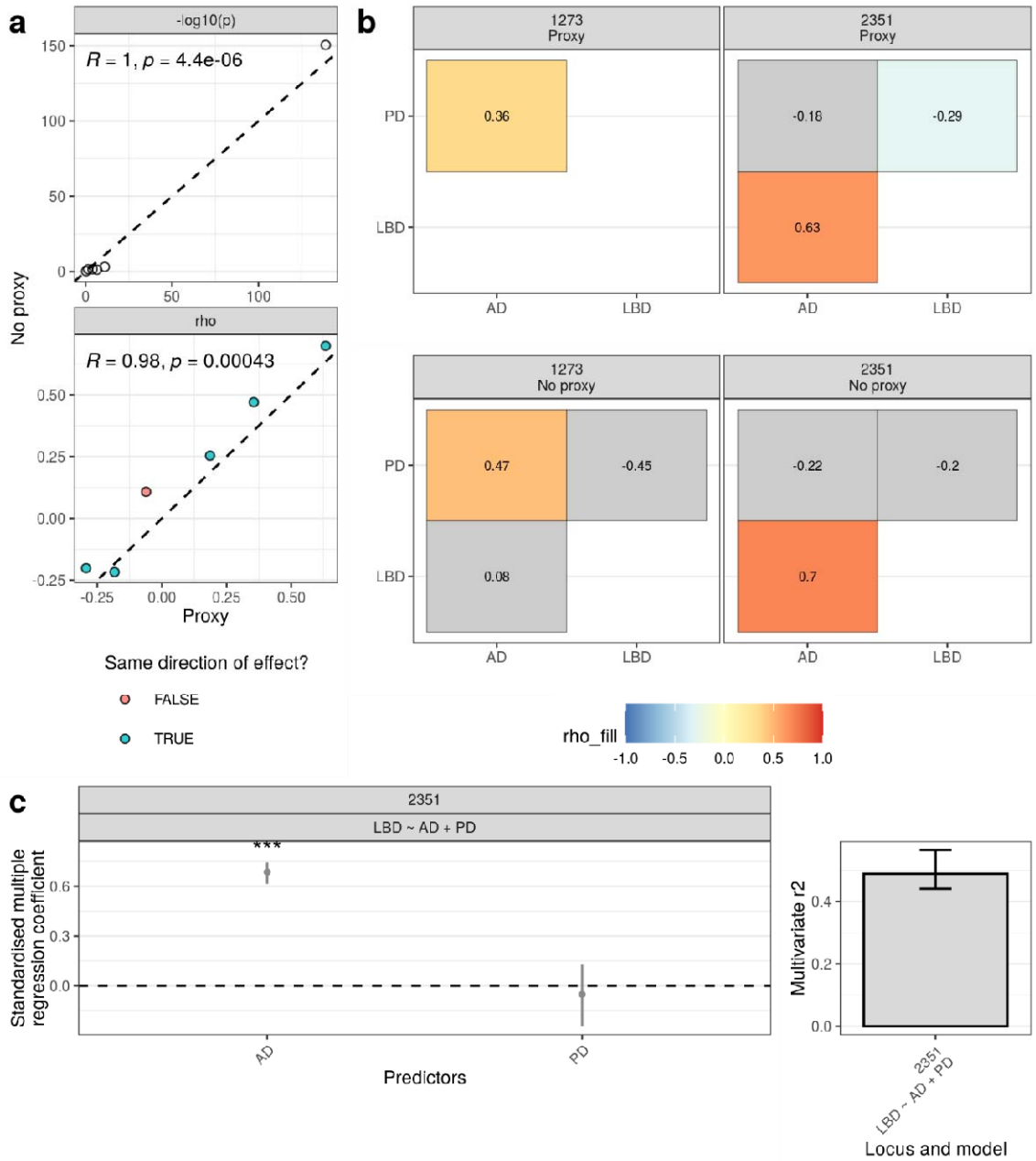
Supplementary Figures



Supplementary Figure 1 (a) Number of LD blocks containing genome-wide significant loci per chromosome. Chromosomes have been ordered by the total number of LD blocks in each chromosome. **(b)** Number of genome-wide significant AD, BIP, LBD, MDD, PD and SCZ SNPs per autosome.



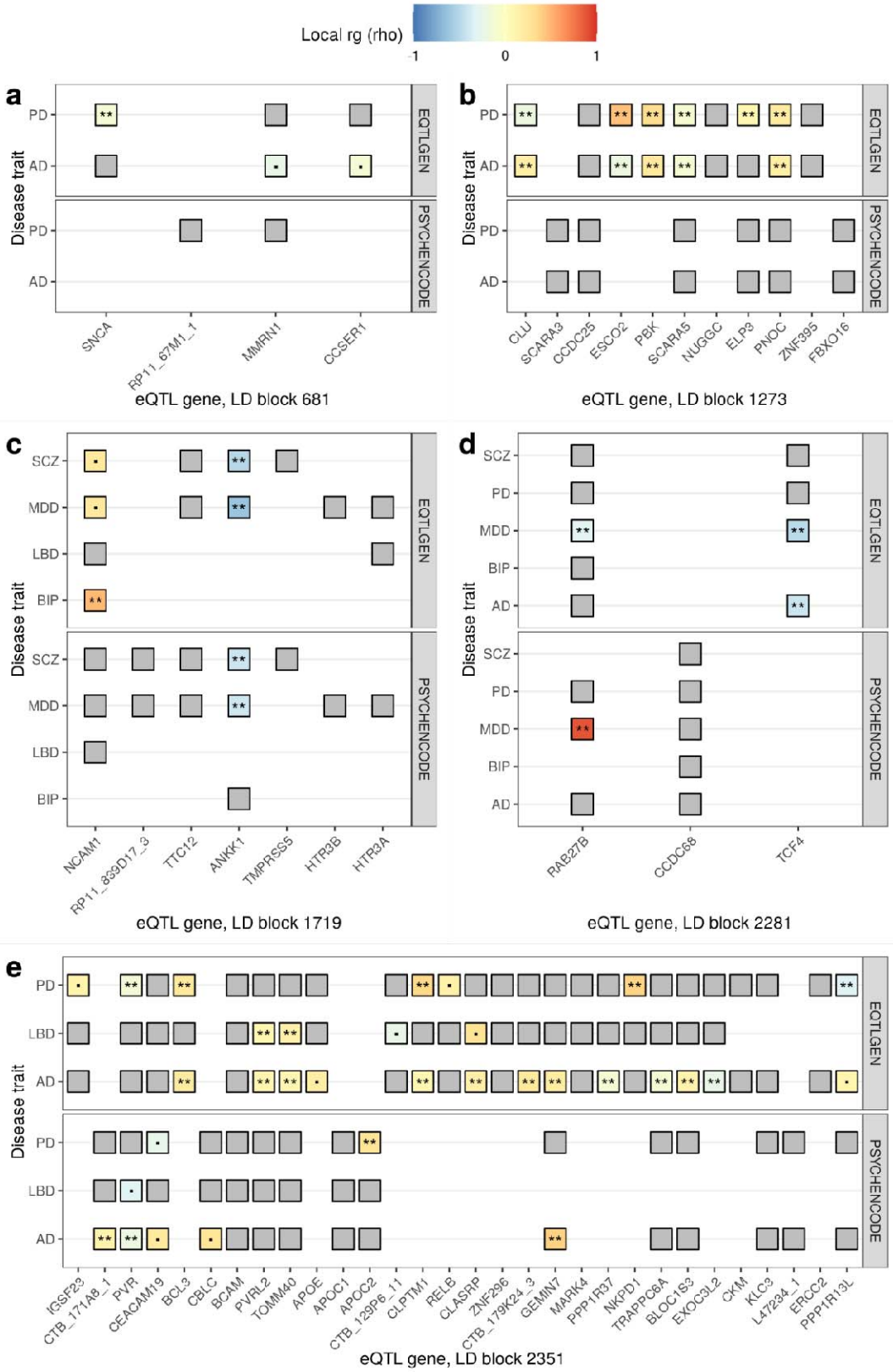
Supplementary Figure 2 Scatter plot of **(a)** the standardised coefficient for (rho, ρ), **(b)** -log10(p-value) and **(c)** number of overlapping SNPs for each pair of traits that overlapped between our study and the study by Gerring *et al.*[18] In **(a)** the dashed lines indicate significance thresholds in each study (Gerring *et al.*, $p < 0.05/24{,}054$; this study, $p < 0.05/1603$) and in **(b, c)** the dashed line represents the line y = x. Points in both plots are coloured by whether they passed significance thresholds in both studies or only the Gerring *et al.* study. Points are labelled by LD block and trait pair. Coordinates for LD blocks (in the format chromosome:start-end, GRCh37): 457, chr3:36,840,137-38,729,767; 951, chr6:26,396,201-27,261,035; 952, chr6:27,261,036-28,666,364; 960, chr6: 31,320,269-31,427,209; 961, chr6:31,427,210-32,208,901; 2483, chr22:38,718,590-40,378,783 .



Supplementary Figure 3 Impact of excluding UK Biobank by-proxy cases on local genetic correlations and multiple regression.

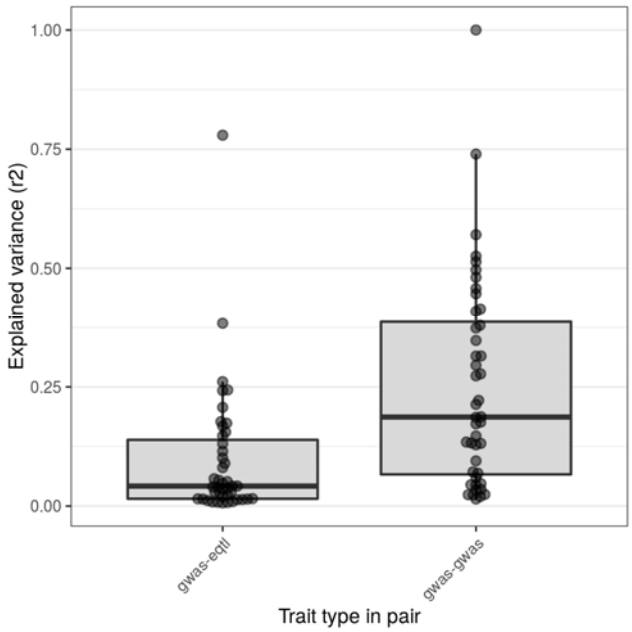
(a) Scatter plot of -log10(p-value) and the standardised coefficient for (rho, ρ) for each pair of traits with sufficient univariate signal to carry out a bivariate test using AD/PD GWASs with or without by-proxy cases. In each panel, Pearson's coefficient (R) and associated p-value (p) are displayed. The black dashed line represents the line y = x. Points are coloured, where applicable, by whether they share the same direction of effect. **(b)** Significant bivariate local genetic correlations using AD/PD GWASs with or without by-proxy cases (as indicated in panel headers). Heatmaps show the rho for all tested associations within the LD block, with significant negative and positive correlations indicated by blue and red fill, respectively. Non-significant correlations have a grey fill. **(c)** Results of multiple regression model across LD block 2351. Plot (left) of standardised coefficients for each predictor in multiple regression model in LD block 2351, with whiskers spanning the 95% confidence interval for the coefficients. Plot (right) of multivariate for LD block 2351, where multivariate represents the proportion of variance in genetic signal for LBD explained by AD and PD simultaneously. Whiskers span the 95% confidence interval for the multivariate . ***, p < 0.001. Coordinates for

LD blocks (in the format chromosome:start-end, GRCh37): 1273, chr8:27,406,512-28,344,176; 2351, chr19:45,040,933-45,893,307.


Local rg (rho)
-1
0
1
a
b
c
d
e
Disease trait
PD
AD
SCZ
MDD
LBD
BIP
EQTLGEN
PSYCHENCODE
SNCA
RP11_67M1_1
MMRN1
CCSER1
eQTL gene, LD block 681
CLU
SCARA3
CCDC25
ESCO2
PBK
SCARA5
NUGGC
ELP3
PNOC
ZNF395
FBXO16
eQTL gene, LD block 1273
NCAM1
RP11_339D17_3
TTC12
ANKK1
TMPRSS5
HTR3B
HTR3A
eQTL gene, LD block 1719
RAB27B
CCDC68
TCF4
eQTL gene, LD block 2281
IGSF23
CTB_171A8_1
PVR
CEACAM19
BCL3
CBLC
BCAM
PVRL2
TOMM40
APOE
APOC1
APOC2
CTB_129P6_11
CLPTM1
RELB
CLASRP
ZNF296
CTB_179K24_3
GEMIN7
MARK4
PPP1R37
NKPD1
TRAPPC6A
BLOC1S3
EXOC3L2
CKM
KLC3
L47234_1
ERCC2
PPP1R13L
eQTL gene, LD block 2351

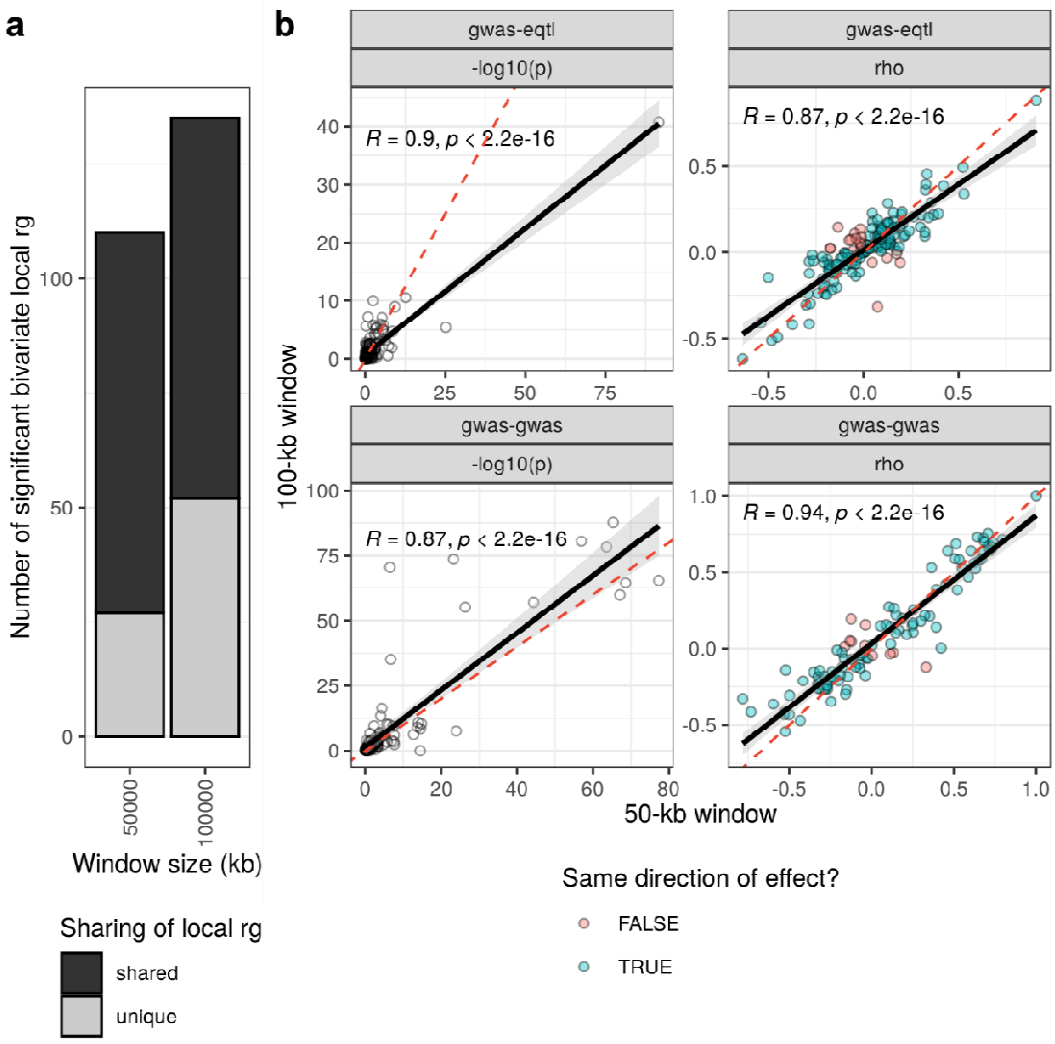
Supplementary Figure 4 Local gene expression and disease trait correlations across 5 LD blocks of interest.

Heatmaps of the standardised coefficient for r_g (rho) for all tested gene expression-disease trait correlation within LD block **(a)** 681, **(b)** 1273, **(c)** 1719, **(d)** 2281 and **(e)** 2351. All negative and positive r_gs with p < 0.05 are indicated by blue and red colour, respectively, while the remainder have a grey fill. Significant local r_gs (FDR < 0.05) are indicated by two asterisks (**), while nominally significant local r_gs (p < 0.05) are indicated with a black square (□). Genes are ordered left to right on the x-axis by the genomic coordinate of their gene start. Coordinates for LD blocks (in the format chromosome:start-end, GRCh37): 681, chr4:90,236,972-91,309,863; 1273, chr8:27,406,512-28,344,176; 1719, chr11:112,755,447-113,889,019; 2281, chr18:52,512,524-53,762,996; 2351, chr19:45,040,933-45,893,307.



Supplementary Figure 5 Explained variance in trait pairs with different trait types.

Boxplot of explained variance (r^2, the proportion of variance in genetic signal of one disease trait in a pair explained by the other) in trait pairs involving a disease and gene expression trait (gwas-eqtl) or two disease traits (gwas-gwas). Only local r_gs that passed significance are plotted (FDR < 0.05; N, local r_gs = 87).



Supplementary Figure 6 Effect of window size on local genetic correlations.
(a) Number of significant bivariate local 's across window sizes. Bars are coloured by whether 's are significant across both window sizes (shared) or only one (unique). (b) Scatter plot of -log10(p-value) and the standardised coefficient for (rho, ρ) for each pair of traitss that could be tested across genic regions with a 50-kb or 100-kb window. Panels indicate whether the pair of traits included a disease and gene expression trait (gwas-eqtl) or two disease traits (gwas-gwas). Points are coloured by whether they share the same direction of effect. The black line represents a linear model fitted to the data, with the 99% confidence interval indicated with a grey fill. Further, Pearson's coefficient (R) and associated p-value (p) are displayed. The red dashed line represents the line y = x.

Supplementary Tables

Supplementary Table 1 LD blocks, their associated disease traits (as determined by overlap of genome-wide significant SNPs) and overlapping genes.

Supplementary Table 2 Results of LDSC using the six disease traits.

Supplementary Table 3 Results of LAVA using the six disease traits.

Supplementary Table 4 Overlaps between results of LAVA and results from Drange *et al*[37], Smeland *et al.*[24] and Gerring *et al.*[18]

Supplementary Table 5 Results of LAVA using GWASs for AD and PD that exclude UK Biobank by-proxy cases.

Supplementary Table 6 Results of multiple regression analyses.

Supplementary Table 7 Results of LAVA using disease and gene expression traits (100-kb window). Sheets containing bivariate results for each LD block also contain (a) locus plot of genic regions (including 100-kb window). Significant bivariate local genetic correlations between a disease and gene expression trait are highlighted in blue (FDR < 0.05). (b) Edge diagrams for genic regions where a significant bivariate local genetic correlation was observed between a disease and gene expression trait (FDR < 0.05). Edges display the standardised coefficient for genetic correlation (rho) for significant bivariate local genetic correlations, with negative and positive correlations indicated by blue and red colour, respectively. GWAS and eQTL nodes are indicated by grey and white fill, respectively.

Supplementary Table 8 Results of LAVA using disease and gene expression traits (50-kb window). Sheets containing bivariate results for each LD block also contain (a) locus plot of genic regions (including 50-kb window). Significant bivariate local genetic correlations between a disease and gene expression trait are highlighted in blue (FDR < 0.05). (b) Edge diagrams for genic regions where a significant bivariate local genetic correlation was observed between a disease and gene expression trait (FDR < 0.05). Edges display the standardised coefficient for genetic correlation (rho) for significant bivariate local genetic correlations, with negative and positive correlations indicated by blue and red colour, respectively. GWAS and eQTL nodes are indicated by grey and white fill, respectively.

Supplementary Table 9 Results of partial correlations using disease and gene expression traits (100-kb window).

Supplementary Note

Comparison of local r_gs to existing reports of local genetic relations

Methods

We compared local r_gs to existing results from studies of: (i) AD and PD using rho-HESS[19]; (ii) AD and BIP[21], AD and MDD[23], PD and SCZ[24], all of which used a conditional/conjunctional FDR approach (conditional FDR is an extension of the standard FDR method, and re-ranks the test statistics of a primary phenotypes based on the strength of the association with a secondary phenotype, while conjunctional FDR is used post-hoc to identify shared genetic loci); and (iii) 10 psychiatric disorders and 10 substance abuse phenotypes using LAVA[18]. For all comparisons, genomic coordinates were used to overlap either SNPs[21,24] or genomic regions[18,19] with LD blocks. SNPs were converted from rsIDs to their GRCh37 genomic coordinates using the SNPlocs.Hsapiens.dbSNP144.GRCh37 package[65]. In all comparisons, local r_gs from this study were filtered to include only relevant disease traits.

Results

We were unable to replicate the genetic overlaps between AD and PD reported in the *HLA* (specifically in chr6: 31,571,218-32,682,664[19]) and *MAPT* (rs393152 shown to increase risk of both AD and PD[20]) loci. In the *MAPT* locus, this was due to insufficient univariate signal for AD in the LD block containing *MAPT* (LD block 2207, chr17: 43,460,501-44,865,832), while in the *HLA* locus, several of the overlapping LD blocks (LD block 961-6, ranging across chr6: 31,427,210-32,682,213) had too few overlapping SNPs between the 6 disease traits.

We were unable to replicate any of the associations reported using a conditional/conjunctional FDR approach. For 4 associations (rs11649476 in AD and BIP[21]; rs5011436 in AD and MDD[23]; rs2979160 and rs4921739 in PD and SCZ[24]) this was due to a lack of overlap with the 300 LD blocks tested. For 1 association (rs62333164 in PD and SCZ[24]), this was due to insufficient univariate signal for SCZ in the overlapping LD block (LD block 752, chr4:169555115-170682809). Of the 9 loci that were jointly associated with PD and SCZ[24], 6 intersected with LD blocks where we performed a bivariate test for

PD and SCZ, but none of the bivariate tests passed the threshold for significance ($p < 0.05/1{,}603$; **Supplementary Table 4**, **Supplementary Table 5**).

Of the 14 LD blocks found to contain significant local r_gs between any of BIP, MDD and SCZ in the study by Gerring *et al.,* 6 LD blocks intersected with LD blocks where we performed a bivariate test using the same phenotype pairs. We were able to replicate 5 of the 7 overlapping local r_gs (BIP and SCZ in LD block 457; SCZ and BIP or MDD in LD block 951; MDD and SCZ in LD block 952; and BIP and SCZ in LD block 2483; **Supplementary Figure 2**; **Supplementary Table 4**). Our inability to replicate 2 local r_gs could be due to: (i) the lower number of overlapping SNPs present in the tested LD blocks in our study compared to Gerring *et al.* (**Supplementary Figure 2c**); and/or (ii) differences in GWASs used (while both studies used the same MDD and SCZ GWASs, our study did not include 23andMe data from the MDD GWAS and used a more recent BIP GWAS).