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 brain traits

Abstract

Neurodegenerative diseases overlap neuropathologically and clinically, frequently manifesting with co-pathologies and comorbid symptoms that span multiple domains, including the neuropsychiatric. Genetically, shared risk loci have been identified across neurodegenerative diseases; however, global analyses of genetic correlation (r_g) show minimal r_g among neurodegenerative diseases or across neurodegenerative and neuropsychiatric diseases. Importantly, local r_gs can exist in the absence of global relationships and can offer valuable insight into underlying shared biological mechanisms. To determine whether regions of local r_g exist among common neurodegenerative and neuropsychiatric diseases, we applied local analysis of [co]variant annotation (LAVA) 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). In addition, we used data from blood- and brain-derived expression quantitative trait loci (eQTLs) to facilitate functional interpretation of local r_gs between disease traits. 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 well-known, disease-implicated genes, such as *SNCA, CLU* and *APOE*, incorporation of eQTLs suggested that genetic overlaps between diseases may be driven by more than one gene. Collectively, our data demonstrate that complex genetic relationships exist among neurodegenerative and neuropsychiatric diseases and highlight putative pleiotropic genomic regions and genes, findings that imply sharing of pathogenic processes and the potential existence of common therapeutic targets.

Keywords: local genetic correlation, neurodegenerative disease, neuropsychiatric disorders

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

57 annotation; LBD = Lewy body dementia; LD = linkage disequilibrium; LDSC = linkage disequilibrium

58 score regression; MDD = major depressive disorder; PD = Parkinson's disease; SCZ = schizophrenia;

59 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 by α-synuclein deposition), 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.

85 Importantly, clinical and neuropathological overlaps are not reflected in global genetic correlations
86 (r_g), with a recent study of global r_g between neurological phenotypes demonstrating limited
87 overlap between individual neurodegenerative diseases as well as between neurodegenerative
88 diseases and neuropsychiatric disorders.[15,16] One explanation for the lack of global genetic
89 correlation is that global studies only consider the average r_g across the entire genome. A genome-
90 wide average of r_g may fail to detect strong local r_gs confined to specific genomic regions or local r_gs
91 that have opposing directions across the genome.[15,17] This limitation can be addressed with recently-
92 developed bioinformatics tools such as local analysis of [co]variant annotation (LAVA), which is able
93 to evaluate local heritability over multiple traits of interest (using summary statistics) and detect
94 local regions of shared genetic association.[18] Here, we apply LAVA to GWASs derived from 3
95 neurogenerative diseases (AD, LBD and PD)[3,5,19,20] and 3 neuropsychiatric disorders (bipolar disorder,
96 BIP; major depressive disorder, MDD; and SCZ)[21–23]. In addition, we use data from blood- and brain-
97 derived gene expression traits, in the form of expression quantitative loci (eQTLs), to facilitate
98 functional interpretation of local r_gs between 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.[24] 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[6,24,25] and consequently, have large GWAS cohorts (AD, N cases = 71,880; PD, N cases = 33,674).[3,19] LBD is the second most common dementia subtype after AD, affecting between 4.2-30% of dementia patients.[26] 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.[24] Accordingly, all 3 disorders have large, well-powered GWASs (BIP, N cases = 41,917; MDD, N cases = 170,756; SCZ, N cases = 40,675).[21–23]

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.*[18] 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

blocks, with 25 local r_gs between trait pairs where no significant global r_g was observed (**Figure 1a**, **Figure 1b**, **Supplementary Table 2**, **Supplementary Table 3**). These 25 correlations included: (i) local r_gs between all 3 neurodegenerative diseases and SCZ; (ii) a local r_g between PD and BIP; and (ii) 20 local r_gs between AD and PD.

For 30 of the 77 local r_gs, the genetic signal of both disease traits may overlap entirely, suggested by the upper limit of the 95% confidence interval (CI) for explained variance (i.e. r^2, the proportion of variance in genetic signal of one disease trait in a pair explained by the other) including 1. Notably, the trait pairs where the upper limit of the 95% CI did not include 1 all involved at least one neurodegenerative disease, with the one exception being a local r_g between PD and SCZ, suggesting that the genetic overlap between neurodegenerative diseases is smaller than between neuropsychiatric disorders in the tested LD blocks (**Figure 1c**).

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 and 2281 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 trait associations. These included (i) LBD in LD

block 1719, 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 (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 and SCZ treatment)[7,27] and *RAB27B* in LD block 2281 (encodes Rab27b, a Rab GTPase recently implicated in α-synuclein clearance).[28]

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, 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)[19,29] in LD block 1273 , 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)[30] in LD block 2351, 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, which contains the AD-associated *KAT8* locus and the PD-associated *SETD1A* locus.[3,19] Importantly, while a genetic overlap between AD and PD has been previously reported at the *MAPT* locus (rs393152 shown to increase risk of both AD and PD),[31] we were unable to replicate this association due to insufficient univariate signal for AD in the LD block containing *MAPT* (LD block 2207, chr17: 43,460,501-44,865,832).

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.[20] 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 2 (LD block 1273 and 2351) 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 and 2351, 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 2**, **Supplementary Table 4**). 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 2**, **Supplementary Table 4**).

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 (758, 951 and 952) 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 5**).

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, 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 2**). 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 and PsychENCODE, which represent large human blood and brain expression datasets, respectively (**Table 1**).[32,33] We restricted analyses to the 5 LD blocks highlighted in **Figure 2** (LD blocks: 681, 1273, 1719, 2281, 2351), 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 3**, **Supplementary Table 6**). We noted that the explained variance (r^2) between trait pairs involving a disease trait and a gene expression trait

223 tended to be lower than between trait pairs involving two disease traits (**Supplementary Figure 4**),
224 an observation that aligns with a recent study that found only 11% of trait heritability to be
225 mediated by bulk-tissue gene expression.[34]

226 With the exception of the *SNCA*-containing LD block 681, where eQTLs for only 1 out of 5 genes
227 tested in the block were correlated with a disease trait (negative r_g between blood-derived *SNCA*
228 eQTLs and PD), the expression of multiple genes was associated with disease traits across the
229 remaining LD blocks (**Figure 4a**). In addition, the expression of several genes was associated with
230 more than one disease trait (**Figure 4b**). For example, blood- and brain-derived *ANKK1* eQTLs (*DRD2*-
231 containing LD block 1719) were negatively correlated with both MDD and SCZ, which themselves
232 were positively correlated (**Figure 4c**). A SNP residing in the coding region of *ANKK1* (rs1800497,
233 commonly known as TaqIA SNP) has been previously associated with alcoholism, schizophrenia and
234 eating disorders, although it is unclear whether this SNP exerts its effect via *DRD2* or *ANKK1*.[35] As
235 *DRD2* is not expressed in blood, and brain-derived *DRD2* eQTLs did not pass the univariate test for
236 sufficient local genetic signal, we were unable to test for local r_gs between *DRD2* eQTLs and any
237 neuropsychiatric disorder. The data available would therefore suggest that the shared risk of MDD
238 and SCZ in the overlapping *ANKK1* and *DRD2* genic regions may be partly driven by *ANKK1* gene
239 expression.

240 A high degree of eQTL sharing across disease traits was observed in the *CLU*-containing LD block
241 1273, with blood-derived eQTLs from 5 out of the 6 genes implicated in local r_gs found to correlate
242 with both AD and PD (**Figure 4b,d**). This included situations where eQTL-disease trait correlations
243 had (i) the same direction of effect across both disease traits (as observed with *PBK, PNOC* and
244 *SCARA5*) or (ii) opposing directions of effect across both disease traits (as observed with *CLU* and
245 *ESCO2*) (**Figure 4d**). Notably, while a significant positive local r_g was observed between AD and PD in
246 the *SCARA5* genic region (reflecting the positive local r_g observed between AD and PD across the
247 entire LD block), no local r_g was observed between AD and PD in the *CLU* genic region, suggesting

248 that the shared risk of AD and PD in LD block 1273 may be driven by the expression of genes other
249 than the AD-associated *CLU* (**Figure 4e**). As a ferritin receptor involved in ferritin internalisation,
250 *SCARA5* could plausibly drive shared AD and PD risk, given that cellular iron overload and iron-
251 induced oxidative stress have been implicated in several neurodegenerative diseases such as AD and
252 PD.[36,37]

253 Compared to LD block 1273, the degree of eQTL sharing across disease traits was lower in the *APOE*-
254 containing LD block 2351, with eQTLs from 4 out of 16 genes implicated in local r_gs found to
255 correlate with AD and one of PD or LBD (**Figure 4b, f**). Shared eQTL genes were only observed in
256 blood and included *BCL3*, *CLPTM1, PVRL2* and *TOMM40*, with expression of *BCL3* and *CLPTM1*
257 positively correlating with AD and PD and expression of *PVRL2* and *TOMM40* positively correlating
258 with AD and LBD. As the exception, *PVR* eQTLs were negatively associated with both AD and PD
259 albeit in different tissues: AD in brain and PD in blood. Expression of the remaining 11 genes was
260 exclusively associated with either AD (n = 8) or PD (n = 3). No significant local r_g was observed
261 between *APOE* eQTLs and AD (FDR < 0.05), although a nominal positive r_g was observed in blood (ρ =
262 0.178, ρ CI = 0.007 to 0.352, p = 0.039; **Supplementary Figure 3e**, **Supplementary Table 6**). Overall,
263 these results indicate that risk of neurodegenerative diseases (in particular, AD) is associated with
264 expression of multiple genes in the *APOE*-containing LD block. Further, they add to a growing body
265 of evidence suggesting that in parallel with the well-studied *APOE-ε4* risk allele, there are additional
266 *APOE*-independent risk factors in the region (such as *BCL3*[38] and *PVRL2*[39]) that may contribute to AD
267 risk.

268 For a complete overview of all genic regions tested across the 5 LD blocks of interest, see
269 **Supplementary Figure 3** and **Supplementary Table 6**.

270

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] and putative new associations (e.g. AD and PD in the *CLU*-containing LD block 1273) 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 and PD with *SNCA* expression in blood,[29,40] 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.[5,41] In support of this continuum, LBD was found to associate with both AD and PD in the *APOE*-containing LD block 2351. Multiple regression analyses confirmed that AD and PD

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

305 Less acknowledged is the genetic overlap between AD and PD, with no global r_g reported between
306 the two diseases[16,42] and no significant evidence for the presence of loci that increase the risk of
307 both diseases.[43] As the exception, a genetic overlap has been reported between AD and PD at the
308 *MAPT* locus,[31] hinting that pleiotropy may exist locally. In support of local pleiotropy, we observed
309 20 local r_gs between AD and PD in genomic regions containing disease-implicated genes, such as
310 *SNCA* (LD block 681) and *CLU* (LD block 1273). In the case of the *CLU*-containing LD block 1273,
311 incorporation of eQTLs indicated that the association of AD and PD may be driven by the expression
312 of several genes, including the ferritin receptor *SCARA5*. In contrast, only blood-derived *SNCA* eQTLs
313 were associated with PD in LD block 681, suggesting that the association between AD and PD at the
314 *SNCA* locus could be driven by tissue- or context-dependent gene expression or alternatively other
315 molecular phenotypes.

316 A few studies have demonstrated genetic overlaps between neurodegenerative and
317 neuropsychiatric diseases, such as AD and BIP as well as AD and MDD,[44,45] while others have
318 demonstrated no overlap,[16,46] with divergences in outcomes ascribed to differences in methodology
319 and cohort.[45] Here, we observed a local r_g between BIP and PD, in addition to local r_gs between
320 schizophrenia and all 3 neurodegenerative diseases, which in the case of LBD was observed in an LD

321 block containing the *DRD2*. Notably, parkinsonism in dementia with Lewy bodies (DLB, one of the
322 two LBDs), is often less responsive to dopaminergic treatments than in PD.[47] Furthermore,
323 methylation of the *DRD2* promoter in leukocytes has been shown to differ between DLB and PD,[48]
324 while D2 receptor density has been shown to be significantly reduced in the temporal cortex of DLB
325 patients, but not AD,[49] suggesting that the *DRD2* locus may harbour markers that could distinguish
326 between these neurodegenerative diseases. Our study adds to the body of evidence in favour of a
327 shared genetic basis between neurodegenerative and neuropsychiatric diseases, although further
328 work will be required to determine whether this genetic overlap underlies the clinical and
329 epidemiological links observed between these two disease groups.

330 This study is not without its limitations, with several limitations related to the input data. These
331 limitations include: (i) the variability in cohort size, which in the case of the smallest GWAS, LBD, may
332 explain the limited number of local r_gs observed involving this trait; (ii) the risk of misdiagnosis; and
333 (iii) the lack of genetic diversity (i.e. all traits used were derived from individuals of European
334 ancestry). Given that population-specific genetic risk factors exist, such as the lack of *MAPT* GWAS
335 signal in the largest GWAS of Asian patients with PD, and that transethnic global r_gs between traits
336 such as gene expression are significantly less than 1, it is imperative that studies of local r_g are
337 expanded to include diverse populations.[50,51]

338 Among methodological limitations, both LDSC and LAVA only consider autosomal chromosomes,
339 leaving out chromosome X, which is not only longer than chromosome 8-22, but also encodes 858
340 and 689 protein-coding and non-coding genes, respectively (Ensembl v106).[52] Furthermore, as
341 mentioned by the developers of LAVA, local r_gs could potentially be confounded by association
342 signals from adjacent genomic regions, a limitation which is particularly pertinent in our analysis of
343 gene expression traits where LD blocks were divided into smaller (often overlapping) genic regions.[18]

344 Importantly, as with any genetic correlation analysis, an observed r_g does not guarantee the
345 presence of true pleiotropy. Spurious r_gs can occur due to LD or misclassification.[17] Here, we

346 attempted to address the potential misclassification of by-proxy cases via sensitivity analyses using
347 GWASs for AD and PD that excluded UKBB by-proxy cases. We replicated 2 of the 3 significant local
348 r_gs observed in 2 LD blocks when using GWASs with by-proxy cases. However, we were unable to
349 test for local r_gs across the remaining 19 LD blocks due to insufficient univariate signal, which could
350 reflect (i) a genuine contribution of by-proxy cases to trait h^2 in the region or (ii) a lack of statistical
351 power to detect a genetic signal. Given the substantial decrease in cohort numbers when UKBB by-
352 proxy cases are excluded from AD and PD GWASs (**Table 1**), a lack of statistical power seems the
353 more likely explanation, warranting a revisit of this analysis as clinically-diagnosed and/or
354 pathologically-defined cohorts increase in size.

355 Finally, even where observed $r_g s$ potentially represent true pleiotropy, LAVA cannot discriminate
356 between vertical and horizontal pleiotropy (also known as mediated and biological pleiotropy,
357 respectively).[17,18] Thus, while incorporation of gene expression can provide testable hypotheses
358 regarding the underlying genes and biological pathways that drive relationships between
359 neurodegenerative and neuropsychiatric diseases, experimental validation is required to establish
360 the extent to which these genes represent genuine intermediary phenotypes.

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

Materials and methods

Trait pre-processing

GWAS summary statistics from a total of 8 distinct traits were used, including 6 disease traits and 2 gene expression traits. Gene expression traits were used to facilitate functional interpretation of local genetic correlations (r_g) between disease 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).[3,19–23,53] Gene expression traits included expression quantitative trait loci (eQTLs) from eQTLGen and PsychENCODE, which represent large human blood and brain expression datasets, respectively.[32,33] 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.[54] 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) determine the global r_g and (ii) estimate sample overlap.[55,56] The latter was used as an input for LAVA, given that potential sample overlap between GWASs could impact estimated local r_gs.[18] 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.[57] For the LD reference panel, 1000 Genomes Project Phase 3 European population

SNPs were used.[58] Both HapMap Project Phase 3 SNPs and European LD Scores from the 1000 Genomes Project are made available by the developers of LDSC[55,56] from the following repository: https://alkesgroup.broadinstitute.org/LDSCORE/ (see **Key resources** for details). Any shared variance due to sample overlap was modelled as a residual genetic covariance. As performed by Werme *et al.*[18], 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 phenotype 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) blocks generated by Werme *et al.*[18] using a genome-wide partitioning algorithm. Briefly, each chromosome was recursively split into blocks using (i) a break point to minimise LD between the resulting blocks and (ii) a minimum size requirement. The resulting LD blocks represent approximately equal-sized, semi-independent blocks of SNPs, with a minimum size requirement of 2,500 SNPs (resulting in an average block size of around 1Mb). Only those LD blocks containing genome-wide significant GWAS loci from at least one trait were carried forward in downstream analyses, resulting in a total of 300 autosomal LD blocks. Of the 22 possible autosomes, 21 contained

LD blocks with overlapping loci, with the highest number of LD blocks located in chromosome 1 and 6 (**Supplementary Figure 1**).

Between disease and gene expression traits

A total of 5 LD blocks, as highlighted by bivariate local r_g analysis of disease traits, were used in this analysis (LD block 681, 1273, 1719, 2281 and 2351). From these LD blocks of interest, we defined genic regions for all protein-coding, antisense or lincRNA genes that overlapped an LD block of interest. Most lead *cis*-eQTL SNPs (i.e. the SNP with the most significant p-value in a SNP-gene association) lie outside the gene start and end coordinates and are located within 100 kb of the gene. Indeed, in eQTLGen, 55% of lead-eQTL SNPs were outside the gene body and 92% were within 100 kb from the gene.[32] Thus, we extended genic regions with a 100-kb window (100 kb upstream and 100 kb downstream of gene coordinates). These genic regions (n = 92) were carried forward in downstream analyses.

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 6, Supplementary Table 7**).

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 5**). 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 5**). 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.[32] In contrast, p-value estimates between two disease traits were comparable across the two window sizes.

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 and PD that excluded UKBB by-proxy cases.[20] 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).[59] 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[60]) or directly from GitHub. Figures were produced using *circlize*, *ggplot2* and *ggraph*.[61–63] 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 Ensembl v87	Ensembl genome browser[52]	http://ftp.ensembl.org/pub/grch37/current/gtf/homo_sapiens/
eQTLGen eQTLs	Vosã *et al.*, 2021[32]	https://www.eqtlgen.org/cis-eqtls.html
GWAS, Alzheimer's disease (clinically diagnosed + UK Biobank proxy cases and controls)	Jansen *et al.*, 2019[19]	https://ctg.cncr.nl/software/summary_statistics
GWAS, Alzheimer's disease (clinically diagnosed)	Kunkle *et al.*, 2019[20]	https://www.niagads.org/igap-rv-summary-stats-kunkle-p-value-data
GWAS, Bipolar disease	Mullins *et al.*, 2021[22]	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[21]	https://www.med.unc.edu/pgc/download-results/
GWAS, Schizophrenia	Pardiñas *et al.*, 2018[23]	https://www.med.unc.edu/pgc/download-results/
LAVA LD blocks	Werme *et al.*, 2021[18]	https://github.com/cadeleeuw/lava-partitioning

LDSC: HapMap Project Phase 3 SNPs	International HapMap 3 Consortium[57]	https://alkesgroup.broadinstitute.org/LDSCORE/; file name: `w_hm3_snplist`
LDSC: 1000 Genomes European LD Scores	1000 Genomes Project Consortium[58]	https://alkesgroup.broadinstitute.org/LDSCORE/; file name: `eur_w_ld_chr.tar.bz2`
PsychENCODE eQTLs	Wang *et al.*, 2018[33]	http://resource.psychencode.org/
Software		
Bioconductor		http://www.bioconductor.org; RRID:SCR_006442
BiocManager (v 1.30.16)	Morgan, 2021[60]	https://CRAN.R-project.org/package=BiocManager
CRAN		http://cran.r-project.org/; RRID:SCR_003005
circlize (v 0.4.13)	Gu *et al.*, 2014[61]	https://github.com/jokergoo/circlize; RRID:SCR_002141
cowplot (v 1.1.1)	Wilke, 2020[64]	https://CRAN.R-project.org/package=cowplot; RRID:SCR_018081
data.table (v 1.14.2)	Dowle and Srinivasan, 2021[65]	https://CRAN.R-project.org/package=data.table
doSNOW (v 1.0.19)	Microsoft and Weston, 2020[66]	https://CRAN.R-project.org/package=doSNOW
foreach (v 1.5.1)	Microsoft and Weston, 2020[67]	https://CRAN.R-project.org/package=foreach
GenomicRanges (v 1.42.0)	Lawrence *et al.*, 2013[68]	https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html; RRID:SCR_000025
ggbeeswarm (v 0.6.0)	Clarke and Sherrill-Mix, 2017[69]	https://CRAN.R-project.org/package=ggbeeswarm
ggplot2 (v 3.3.5)	Wickham, 2016[62]	https://ggplot2.tidyverse.org; RRID:SCR_014601
ggpubr (v 0.4.0)	Kassambara, 2020[70]	https://CRAN.R-project.org/package=ggpubr; RRID:SCR_021139
ggraph (v 2.0.5)	Pedersen, 2021[63]	https://CRAN.R-project.org/package=ggraph; RRID:SCR_021239
gtools (v 3.9.2)	Warnes *et al.*, 2021[71]	https://CRAN.R-project.org/package=gtools
here (v 1.0.1)	Müller, 2020[72]	https://CRAN.R-project.org/package=here
janitor (v 2.1.0)	Firke, 2021[73]	https://CRAN.R-project.org/package=janitor

LAVA (v 0.0.6; commit #7be342)	Werme *et al.*, 2021[18]	https://github.com/josefin-werme/LAVA
LDSC (v 1.0.1)	Bulik-Sullivan *et al.*, 2015[74]	https://github.com/bulik/ldsc
openxlsx (v 4.2.4)	Schauberger and Walker, 2021[75]	https://CRAN.R-project.org/package=openxlsx; RRID:SCR_019185
qdapTools (v 1.3.5)	Rinker, 2015[76]	http://github.com/trinker/qdapTools
readxl (v 1.3.1)	Wickham and Bryan, 2019[77]	https://CRAN.R-project.org/package=readxl; RRID:SCR_018083
R (v 4.0.5)	R Core Team[59]	http://www.r-project.org/; RRID:SCR_001905
rtracklayer (v 1.50.0)	Lawrence *et al.*, 2009[78]	https://bioconductor.org/packages/release/bioc/html/rtracklayer.html; RRID:SCR_021325
tidyverse (v 1.3.1)	Wickham et al., 2019[79]	https://www.tidyverse.org/; RRID:SCR_019186

505

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.

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.

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.

References

1. Robinson JL, Lee EB, Xie SX, et al. Neurodegenerative disease concomitant proteinopathies are prevalent, age-related and APOE4-associated. *Brain*. 2018;141(7):2181-2193. doi:10.1093/brain/awy146

2. De Jager PL, Yang HS, Bennett DA. Deconstructing and targeting the genomic architecture of human neurodegeneration. *Nat Neurosci*. 2018;21(10):1310-1317. doi:10.1038/s41593-018-0240-z

3. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18(12):1091-1102. doi:10.1016/S1474-4422(19)30320-5

4. Guerreiro R, Ross OA, Kun-Rodrigues C, et al. Investigating the genetic architecture of dementia with Lewy bodies: a two-stage genome-wide association study. *Lancet Neurol*. 2018;17(1):64-74. doi:10.1016/S1474-4422(17)30400-3

5. Chia R, Sabir MS, Bandres-Ciga S, et al. Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nat Genet*. 2021;53(3):294-303. doi:10.1038/s41588-021-00785-3

6. Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL. Alzheimer's disease. *Nat Rev Dis Prim*. 2015;1:15056. doi:10.1038/nrdp.2015.56

7. Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. *Nat Rev Dis Prim*. 2017;3:17013. doi:10.1038/nrdp.2017.13

8. Kuring JK, Mathias JL, Ward L. Prevalence of Depression, Anxiety and PTSD in People with Dementia: a Systematic Review and Meta-Analysis. *Neuropsychol Rev*. 2018;28(4):393-416. doi:10.1007/s11065-018-9396-2

9. Reijnders JSAM, Ehrt U, Weber WEJ, Aarsland D, Leentjens AFG. A systematic review of

562 prevalence studies of depression in Parkinson's disease. *Mov Disord*. 2008;23(2):183-189;
563 quiz 313. doi:10.1002/mds.21803

564 10. Weintraub D, Aarsland D, Chaudhuri KR, et al. The neuropsychiatry of Parkinson's disease:
565 advances and challenges. *Lancet Neurol*. 2022;21(1):89-102. doi:10.1016/S1474-
566 4422(21)00330-6

567 11. Ribe AR, Laursen TM, Charles M, et al. Long-term risk of dementia in persons with
568 schizophrenia: A danish population-based cohort study. *JAMA Psychiatry*. 2015;72(11):1095-
569 1101. doi:10.1001/jamapsychiatry.2015.1546

570 12. Stroup TS, Olfson M, Huang C, et al. Age-Specific Prevalence and Incidence of Dementia
571 Diagnoses among Older US Adults with Schizophrenia. *JAMA Psychiatry*. 2021;78(6):632-641.
572 doi:10.1001/jamapsychiatry.2021.0042

573 13. Gustafsson H, Nordström A, Nordström P. Depression and subsequent risk of Parkinson
574 disease A nationwide cohort study. *Neurology*. 2015;84(24):2422-2429.
575 doi:10.1212/WNL.0000000000001684

576 14. Kazmi H, Walker Z, Booij J, et al. Late onset depression: Dopaminergic deficit and clinical
577 features of prodromal Parkinson's disease: A cross-sectional study. *J Neurol Neurosurg*
578 *Psychiatry*. 2021;92(2):158-164. doi:10.1136/jnnp-2020-324266

579 15. Bellou E, Stevenson-Hoare J, Escott-Price V. Polygenic risk and pleiotropy in
580 neurodegenerative diseases. *Neurobiol Dis*. 2020;142(January):104953.
581 doi:10.1016/j.nbd.2020.104953

582 16. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, et al. Analysis of shared heritability in
583 common disorders of the brain. *Science*. 2018;360(6395):eaap8757.
584 doi:10.1126/science.aap8757

585 17. van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic

586 disease traits: from theory to practice. *Nat Rev Genet*. 2019;20(10):567-581.
587 doi:10.1038/s41576-019-0137-z

588 18. Werme J, van der Sluis S, Posthuma D, de Leeuw CA. An integrated framework for local
589 genetic correlation analysis. *Nat Genet*. 2022;54(3):274-282. doi:10.1038/s41588-022-01017-
590 y

591 19. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and
592 functional pathways influencing Alzheimer's disease risk. *Nat Genet*. 2019;51(3):404-413.
593 doi:10.1038/s41588-018-0311-9

594 20. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's
595 disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat*
596 *Genet*. 2019;51(3):414-430. doi:10.1038/s41588-019-0358-2

597 21. Howard DM, Adams MJ, Clarke TK, et al. Genome-wide meta-analysis of depression identifies
598 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat*
599 *Neurosci*. 2019;22(3):343-352. doi:10.1038/s41593-018-0326-7

600 22. Mullins N, Forstner AJ, O'Connell KS, et al. Genome-wide association study of more than
601 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet*.
602 2021;53(6):817-829. doi:10.1038/s41588-021-00857-4

603 23. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in
604 mutation-intolerant genes and in regions under strong background selection. *Nat Genet*.
605 2018;50(3):381-389. doi:10.1038/s41588-018-0059-2

606 24. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in
607 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of
608 Disease Study 2019. *Lancet (London, England)*. 2020;396(10258):1204-1222.
609 doi:10.1016/S0140-6736(20)30925-9

610 25. Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. *Nat Rev Dis Prim*. 2017;3:17013.
611 doi:10.1038/nrdp.2017.13

612 26. Erkkinen MG, Kim M ohk, Geschwind MD. Clinical Neurology and Epidemiology of the Major
613 Neurodegenerative Diseases. *Cold Spring Harb Perspect Biol*. 2018;10(4).
614 doi:10.1101/cshperspect.a033118

615 27. Urs NM, Peterson SM, Caron MG. New Concepts in Dopamine D2Receptor Biased Signaling
616 and Implications for Schizophrenia Therapy. *Biol Psychiatry*. 2017;81(1):78-85.
617 doi:10.1016/j.biopsych.2016.10.011

618 28. Underwood R, Wang B, Carico C, Whitaker RH, Placzek WJ, Yacoubian TA. The GTPase Rab27b
619 regulates the release, autophagic clearance, and toxicity of α-synuclein. *J Biol Chem*.
620 2020;295(23):8005-8016. doi:10.1074/jbc.RA120.013337

621 29. Foster EM, Dangla-Valls A, Lovestone S, Ribe EM, Buckley NJ. Clusterin in Alzheimer's Disease:
622 Mechanisms, Genetics, and Lessons From Other Pathologies. *Front Neurosci*.
623 2019;13(FEB):164. doi:10.3389/fnins.2019.00164

624 30. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of
625 Alzheimer disease genetic association studies: The AlzGene database. *Nat Genet*.
626 2007;39(1):17-23. doi:10.1038/ng1934

627 31. Desikan RS, Schork AJ, Wang Y, et al. Genetic overlap between Alzheimer's disease and
628 Parkinson's disease at the MAPT locus. *Mol Psychiatry*. 2015;20(12):1588-1595.
629 doi:10.1038/mp.2015.6

630 32. Võsa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify
631 thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat*
632 *Genet*. 2021;53(9):1300-1310. doi:10.1038/s41588-021-00913-z

633 33. Wang D, Liu S, Warrell J, et al. Comprehensive functional genomic resource and integrative

634 model for the human brain. *Science*. 2018;362(6420). doi:10.1126/science.aat8464

635 34. Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by
636 assayed gene expression levels. *Nat Genet*. 2020;52(6):626-633. doi:10.1038/s41588-020-
637 0625-2

638 35. Ponce G, Pérez-González R, Aragüés M, et al. The ANKK1 kinase gene and psychiatric
639 disorders. *Neurotox Res*. 2009;16(1):50-59. doi:10.1007/s12640-009-9046-9

640 36. Li JY, Paragas N, Ned RM, et al. Scara5 Is a Ferritin Receptor Mediating Non-Transferrin Iron
641 Delivery. *Dev Cell*. 2009;16(1):35-46. doi:10.1016/j.devcel.2008.12.002

642 37. Crielaard BJ, Lammers T, Rivella S. Targeting iron metabolism in drug discovery and delivery.
643 *Nat Rev Drug Discov*. 2017;16(6):400-423. doi:10.1038/nrd.2016.248

644 38. Nho K, Kim S, Horgusluoglu E, et al. Association analysis of rare variants near the APOE region
645 with CSF and neuroimaging biomarkers of Alzheimer's disease. *BMC Med Genomics*.
646 2017;10(Suppl 1). doi:10.1186/s12920-017-0267-0

647 39. Zhou X, Chen Y, Mok KY, et al. Non-coding variability at the APOE locus contributes to the
648 Alzheimer's risk. *Nat Commun*. 2019;10(1):1-16. doi:10.1038/s41467-019-10945-z

649 40. Locascio JJ, Eberly S, Liao Z, et al. Association between α-synuclein blood transcripts and
650 early, neuroimaging-supported Parkinson's disease. *Brain*. 2015;138(9):2659-2671.
651 doi:10.1093/brain/awv202

652 41. Jellinger KA. Dementia with Lewy bodies and Parkinson's disease-dementia: current concepts
653 and controversies. *J Neural Transm*. 2018;125(4):615-650. doi:10.1007/s00702-017-1821-9

654 42. Guerreiro R, Escott-Price V, Darwent L, et al. Genome-wide analysis of genetic correlation in
655 dementia with Lewy bodies, Parkinson's and Alzheimer's diseases. *Neurobiol Aging*.
656 2016;38:214.e7-214.e10. doi:10.1016/j.neurobiolaging.2015.10.028

657 43. Moskvina V, Harold D, Russo GC, et al. Analysis of genome-wide association studies of

658 alzheimer disease and of parkinson disease to determine if these 2 diseases share a common
659 genetic risk. *JAMA Neurol*. 2013;70(10):1268-1276. doi:10.1001/jamaneurol.2013.448

660 44. Drange OK, Bjerkehagen Smeland O, Shadrin AA, et al. Genetic overlap between alzheimer's
661 disease and bipolar disorder implicates the MARK2 and VAC14 genes. *Front Neurosci*.
662 2019;13(March):1-11. doi:10.3389/fnins.2019.00220

663 45. Lutz MW, Sprague D, Barrera J, Chiba-Falek O. Shared genetic etiology underlying Alzheimer's
664 disease and major depressive disorder. *Transl Psychiatry*. 2020;10(1). doi:10.1038/s41398-
665 020-0769-y

666 46. Gibson J, Russ TC, Adams MJ, et al. Assessing the presence of shared genetic architecture
667 between Alzheimer's disease and major depressive disorder using genome-wide association
668 data. *Transl Psychiatry*. 2017;7(4). doi:10.1038/tp.2017.49

669 47. McKeith IG, Boeve BF, Dickson DW, et al. Diagnosis and management of dementia with Lewy
670 bodies: Fourth consensus report of the DLB Consortium. *Neurology*. 2017;89(1):88-100.
671 doi:10.1212/WNL.0000000000004058

672 48. Ozaki Y, Yoshino Y, Yamazaki K, et al. DRD2 methylation to differentiate dementia with Lewy
673 bodies from Parkinson's disease. *Acta Neurol Scand*. 2020;141(2):177-182.
674 doi:10.1111/ane.13186

675 49. Piggott MA, Ballard CG, Rowan E, et al. Selective loss of dopamine D2 receptors in temporal
676 cortex in dementia with Lewy bodies, association with cognitive decline. *Synapse*.
677 2007;61(11):903-911. doi:10.1002/syn.20441

678 50. Foo JN, Chew EGY, Chung SJ, et al. Identification of Risk Loci for Parkinson Disease in Asians
679 and Comparison of Risk Between Asians and Europeans: A Genome-Wide Association Study.
680 *JAMA Neurol*. 2020;77(6):746-754. doi:10.1001/jamaneurol.2020.0428

681 51. Brown BC, Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye CJ, Price AL,

682 Zaitlen N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am J Hum*
683 *Genet*. 2016;99(1):76-88. doi:10.1016/j.ajhg.2016.05.001

684 52. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;46(D1):D754-
685 D761. doi:10.1093/nar/gkx1098

686 53. Chia R, Sabir MS, Bandres-Ciga S, et al. Genome sequencing analysis identifies new loci
687 associated with Lewy body dementia and provides insights into the complex genetic
688 architecture. *bioRxiv*. Published online January 1, 2020:2020.07.06.185066.
689 doi:10.1101/2020.07.06.185066

690 54. Pagès H. SNPlocs.Hsapiens.dbSNP144.GRCh37: SNP locations for Homo sapiens (dbSNP Build
691 144). *R Packag version 09920*. Published online 2017.

692 55. Bulik-Sullivan B, Loh PR, Finucane HK, et al. LD score regression distinguishes confounding
693 from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.
694 doi:10.1038/ng.3211

695 56. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human
696 diseases and traits. *Nat Genet*. 2015;47(11):1236-1241.

697 57. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, et al. Integrating common and
698 rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52-58.
699 doi:10.1038/nature09298

700 58. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human
701 genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

702 59. R Core Team. R: A language and environment for statistical computing. Published online
703 2021. http://www.r-project.org/

704 60. Morgan M. BiocManager: Access the Bioconductor Project Package Repository. Published
705 online 2021. https://cran.r-project.org/package=BiocManager

706 61. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular
707 visualization in R. *Bioinformatics*. 2014;30(19):2811-2812. doi:10.1093/bioinformatics/btu393

708 62. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016.
709 https://ggplot2.tidyverse.org

710 63. Pedersen TL. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks.
711 Published online 2021. https://cran.r-project.org/package=ggraph

712 64. Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2." Published
713 online 2020. https://cran.r-project.org/package=cowplot

714 65. Dowle M, Srinivasan A. data.table: Extension of `data.frame`. Published online 2021.
715 https://cran.r-project.org/package=data.table

716 66. Corporation M, Weston S. doSNOW: Foreach Parallel Adaptor for the "snow" Package.
717 Published online 2020. https://cran.r-project.org/package=doSNOW

718 67. Microsoft, Weston S. foreach: Provides Foreach Looping Construct. Published online 2020.
719 https://cran.r-project.org/package=foreach

720 68. Lawrence M, Huber W, Pagès H, et al. Software for Computing and Annotating Genomic
721 Ranges. Prlic A, ed. *PLoS Comput Biol*. 2013;9(8):e1003118. doi:10.1371/journal.pcbi.1003118

722 69. Clarke E, Sherrill-Mix S. ggbeeswarm: Categorical Scatter (Violin Point) Plots. Published online
723 2017. https://cran.r-project.org/package=ggbeeswarm

724 70. Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots. Published online 2020.
725 https://cran.r-project.org/package=ggpubr

726 71. Warnes GR, Bolker B, Lumley T. gtools: Various R Programming Tools. Published online 2021.
727 https://cran.r-project.org/package=gtools

728 72. Müller K. here: A Simpler Way to Find Your Files. Published online 2020. https://cran.r-
729 project.org/package=here

730 73. Firke S. janitor: Simple Tools for Examining and Cleaning Dirty Data. Published online 2021.
731 https://cran.r-project.org/package=janitor

732 74. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human
733 diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406

734 75. Schauberger P, Walker A. openxlsx: Read, Write and Edit xlsx Files. Published online 2021.
735 https://cran.r-project.org/package=openxlsx

736 76. Rinker TW. {qdapTools}: Tools to Accompany the qdap Package. Published online 2015.
737 http://github.com/trinker/qdapTools

738 77. Wickham H, Bryan J. readxl: Read Excel Files. Published online 2019. https://cran.r-
739 project.org/package=readxl

740 78. Lawrence M, Gentleman R, Carey V. rtracklayer: An R package for interfacing with genome
741 browsers. *Bioinformatics*. 2009;25(14):1841-1842. doi:10.1093/bioinformatics/btp328

742 79. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *J Open Source Softw*.
743 2019;4(43):1686. doi:10.21105/joss.01686

744

Figures

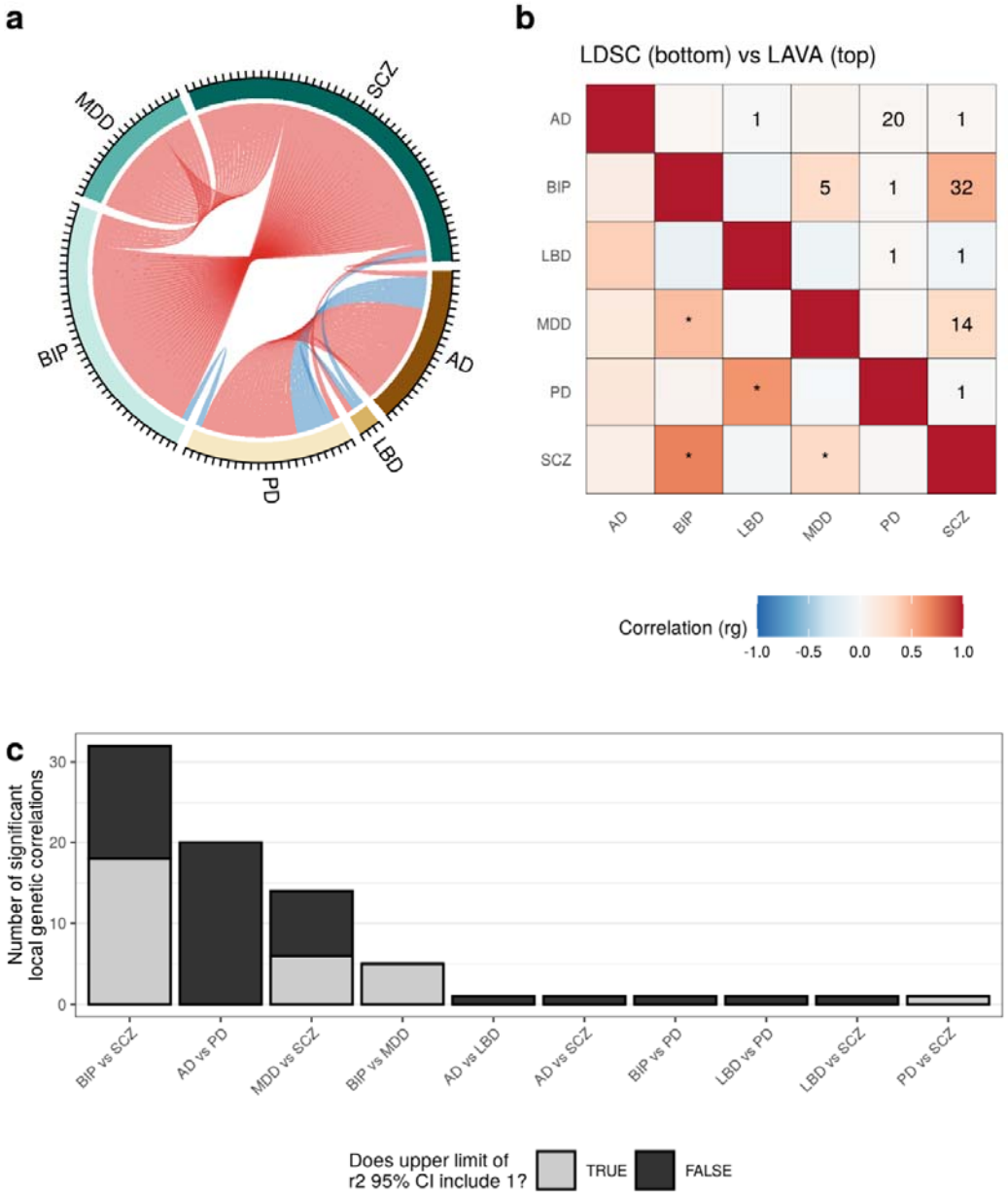


Figure 1 Overview of local and global genetic correlations between neurodegenerative diseases and neuropsychiatric disorders.
(a) Chord diagram showing the number of significant bivariate local s ($p < 0.05/1603$) between each of the disease traits across all LD blocks. Positive and negative correlations 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.

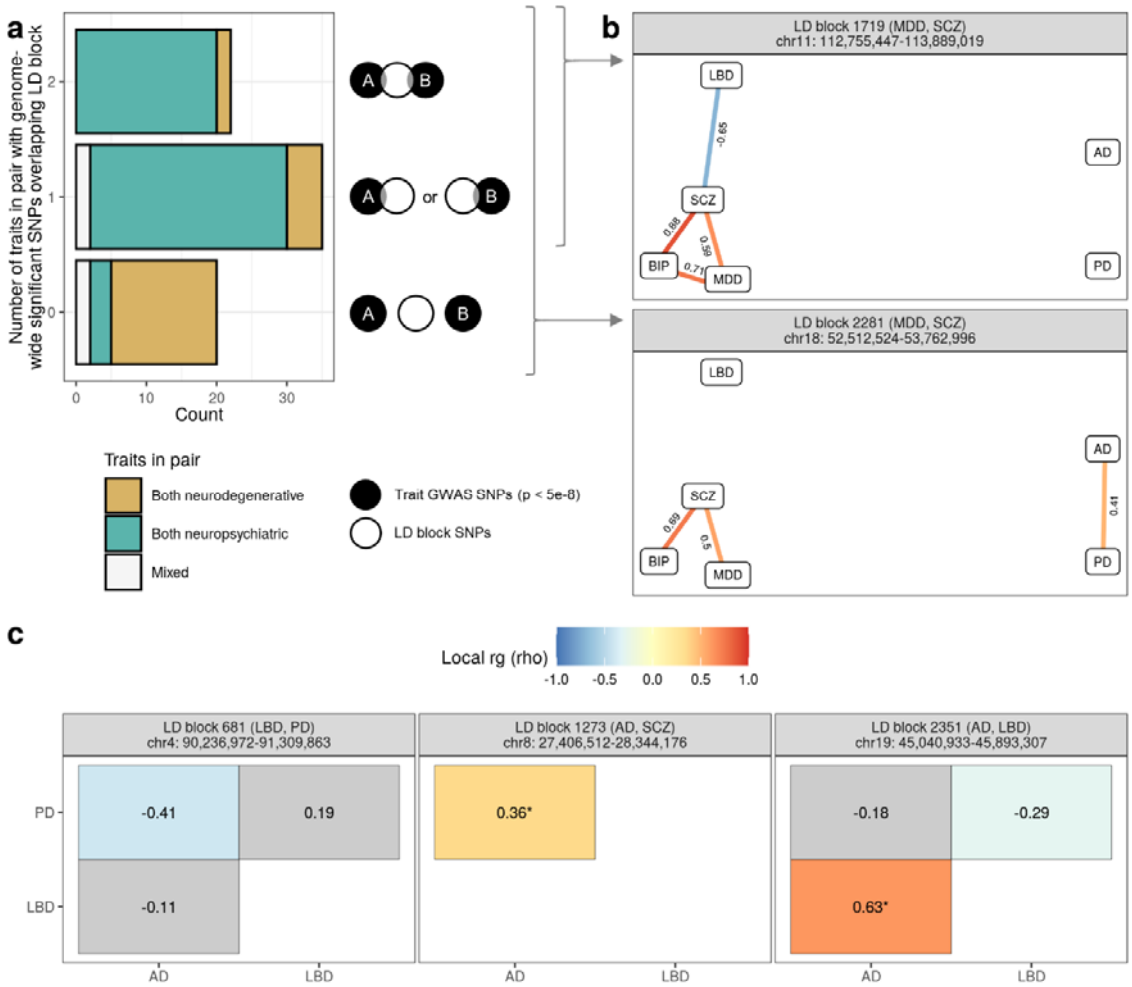


Figure 2 Local analyses associate disease-implicated genomic regions with previously unrelated traits.

(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).

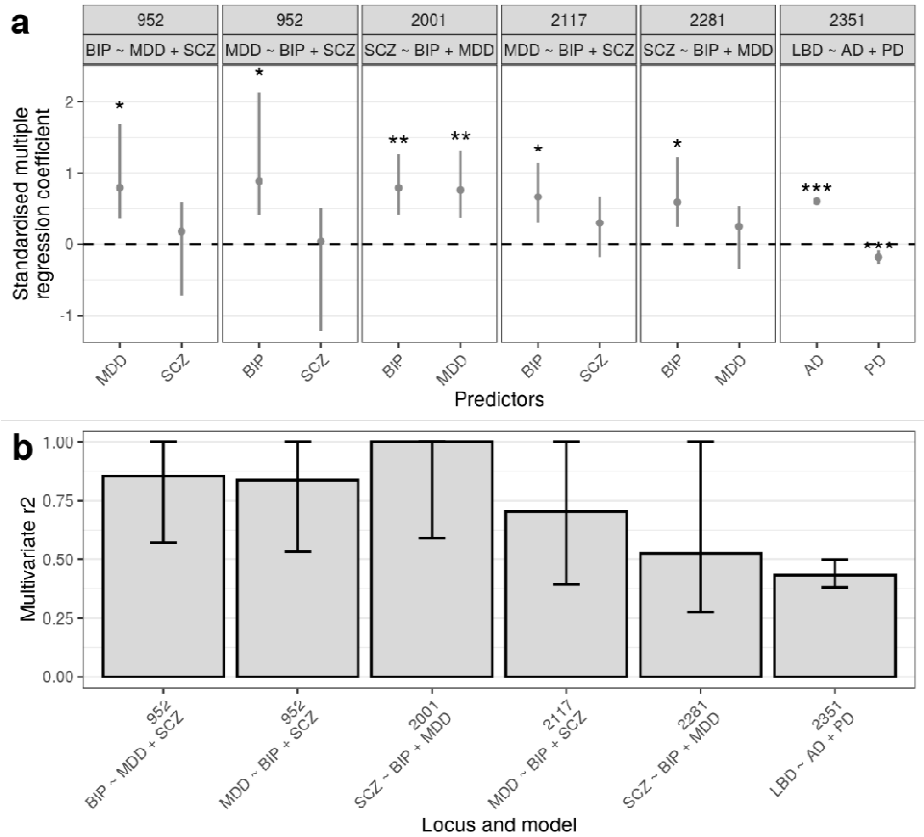
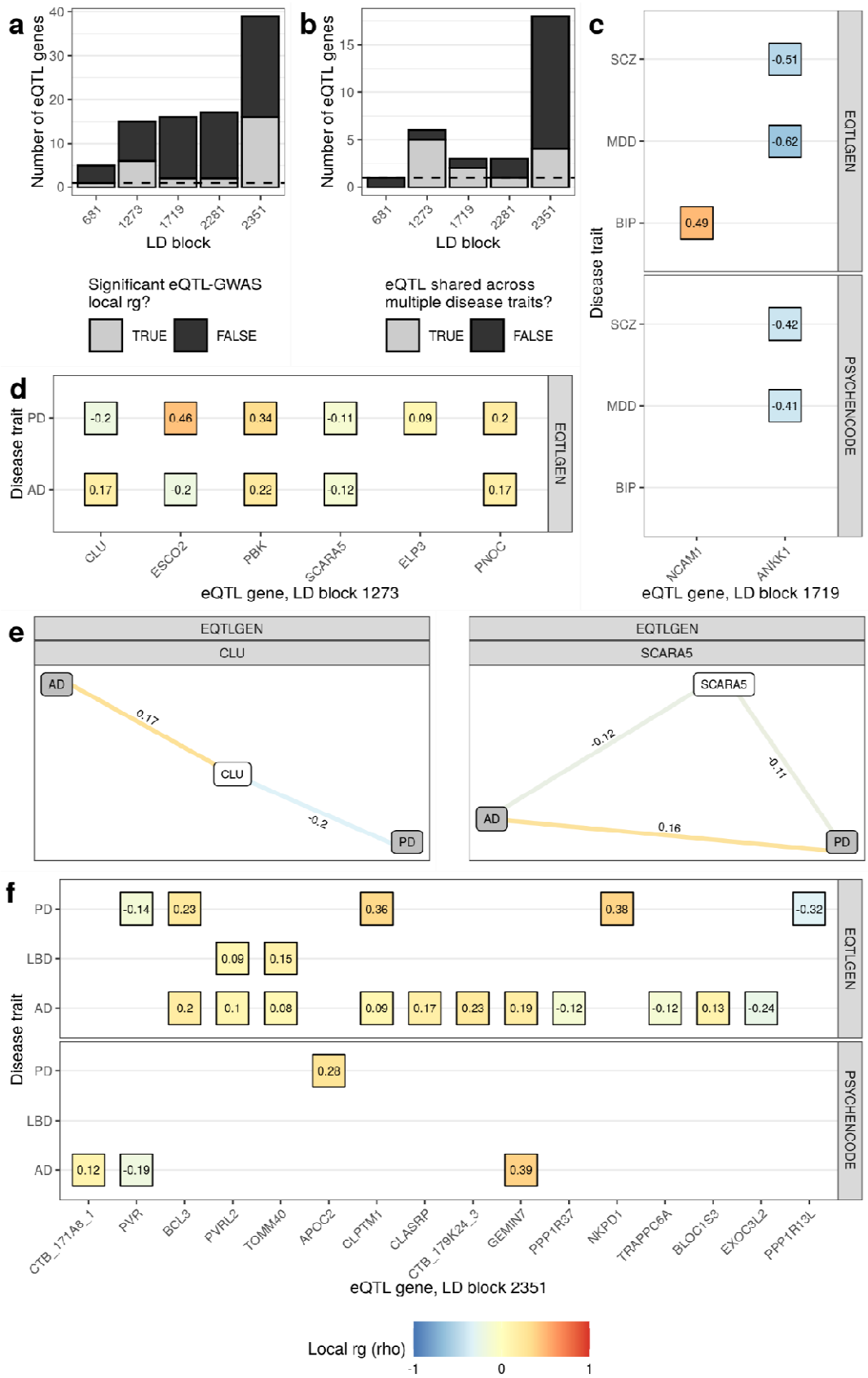


Figure 3 Multiple regression across LD blocks with multiple trait pair correlations.

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 . ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$.


a
Number of eQTL genes
LD block
681
1273
1719
2281
2351
Significant eQTL-GWAS local rg?
TRUE
FALSE
b
eQTL shared across multiple disease traits?
c
Disease trait
SCZ
MDD
BIP
EQTLGEN
PSYCHENCODE
NCAM1
ANKK1
eQTL gene, LD block 1719
d
PD
AD
CLU
ESCO2
PBK
SCARA5
ELP3
PNOC
eQTL gene, LD block 1273
e
EQTLGEN
CLU
SCARA5
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)

778

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

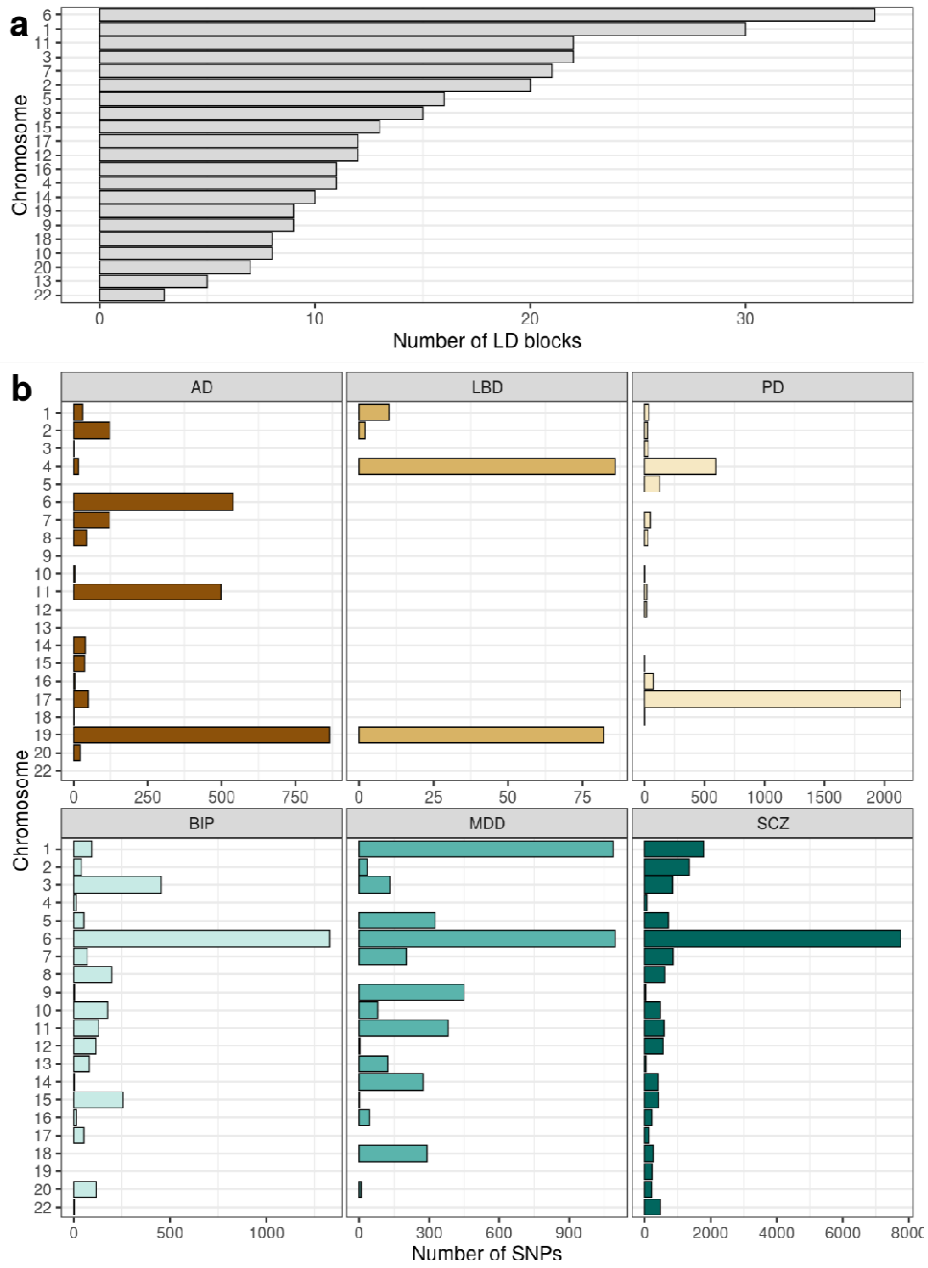
(a) Bar plot of the number of eQTL genes (as defined by their genic regions) tested in each LD block. The fill of the bars indicates whether eQTL genes were significantly correlated with at least one disease trait. **(b)** Bar plot of the number of eQTL genes that were significantly correlated with at 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.

Tables

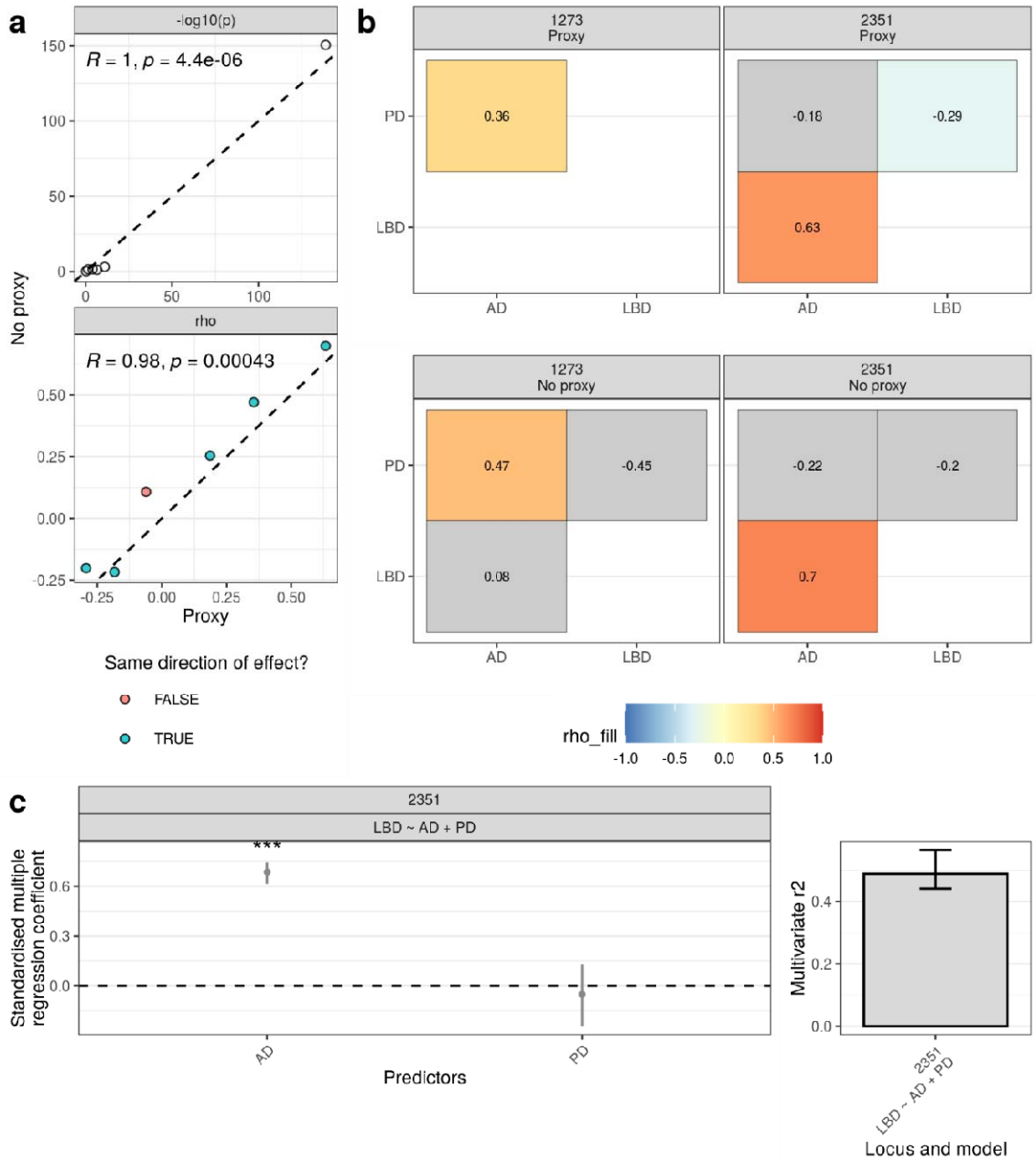
Table 1 Overview of traits included in this study. Global SNP heritability (h^2) for each trait was obtained using LDSC [74]. SE, standard error.

Trait type	Trait	Abbreviation	N	N cases	N controls	Global h^2 (SE)	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)	Jansen *et al.*, 2019[19]
Disease	Alzheimer's disease Clinically diagnosed	AD (no proxy)	63,926	21,982	41,944	7.1% (1.1)	Kunkle *et al.*, 2019[20]
Disease	Bipolar disorder	BIP	413,466	41,917	371,549	7.1% (0.3)	Mullins *et al.*, 2021[22]
Disease	Lewy body dementia Autopsy-confirmed + clinically diagnosed	LBD	6,618	2,591 (1,789 autopsy-confirmed)	4,027	17.1% (7.6)	Chia *et al.*, 2021[5]
Disease	Major depressive disorder	MDD	500,199	170,756	329,443	6% (0.2)	Howard *et al.*, 2019[21]
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)	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)	Nalls *et al.*, 2019[3]
Disease	Schizophrenia	SCZ	105,318	40,675	64,643	41% (1.4)	Pardiñas *et al.*, 2018[23]
Gene expression	eQTLGen Blood-derived eQTLs	eQTLGEN	31,684	-	-	-	Vosã *et al.*, 2021[32]
Gene expression	PsychENCODE Brain-derived eQTLs	PSYCHENCODE	1,387	-	-	-	Wang *et al.*, 2018[33]

Supplementary Figures

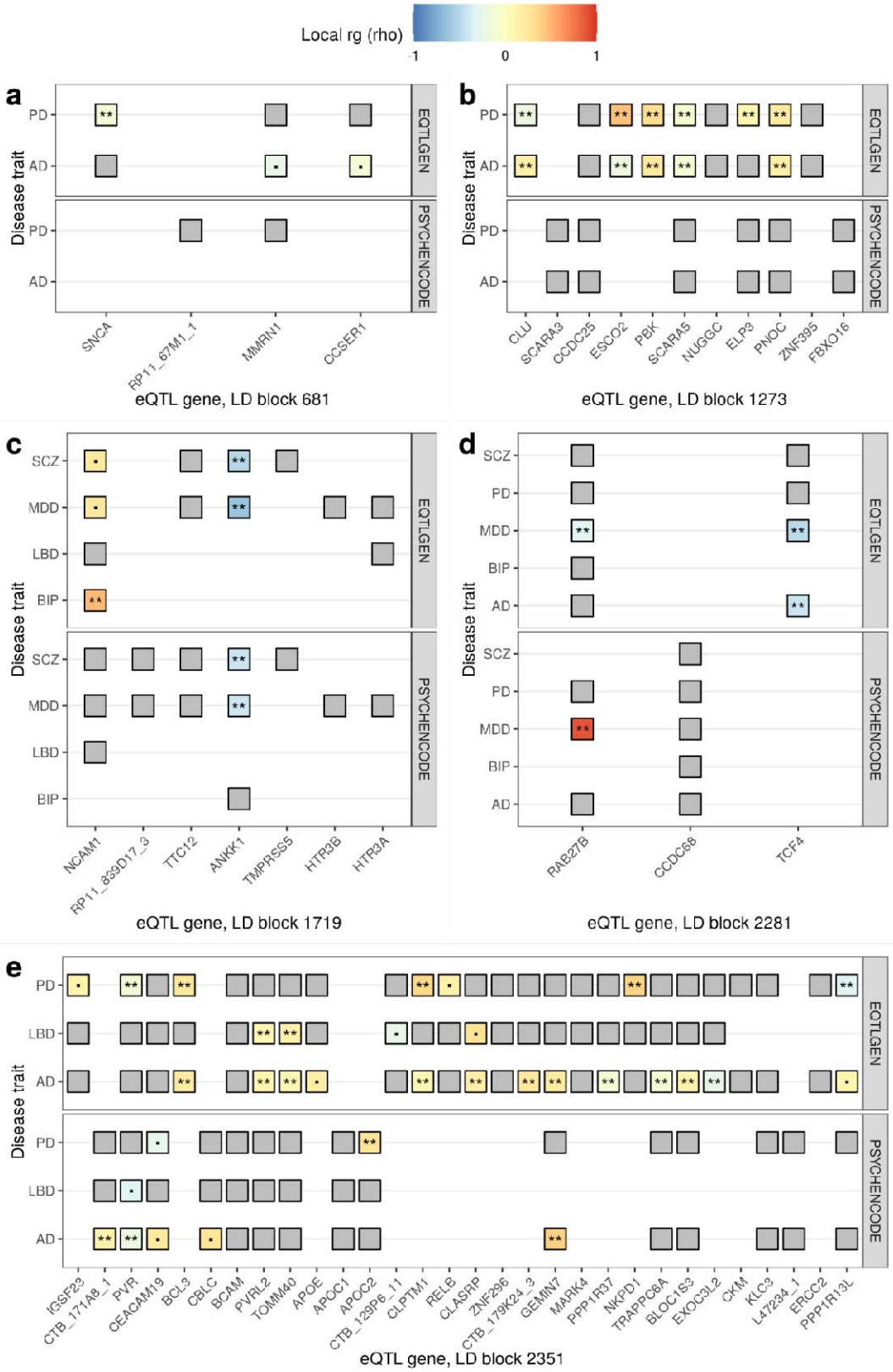


Supplementary Figure 1 (a) Number of LD blocks containing genome-wide significant loci per chromosome. Chromosomes have been ordered by the total number of LD blocks in each chromosome. **(b)** Number of genome-wide significant AD, BIP, LBD, MDD, PD and SCZ SNPs per autosome.



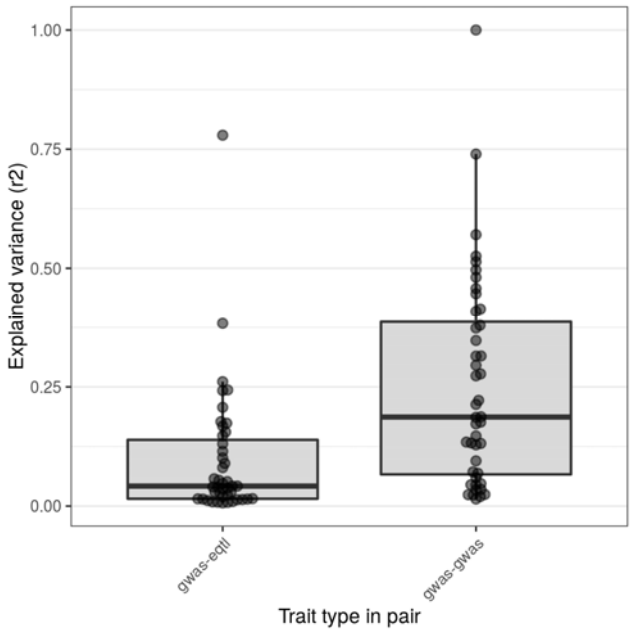
Supplementary Figure 2 Impact of excluding UK Biobank by-proxy cases on local genetic correlations and multiple regression.

(a) Scatter plot of -log10(p-value) and the standardised coefficient for (rho, ρ) for each pair of phenotypes 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.


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

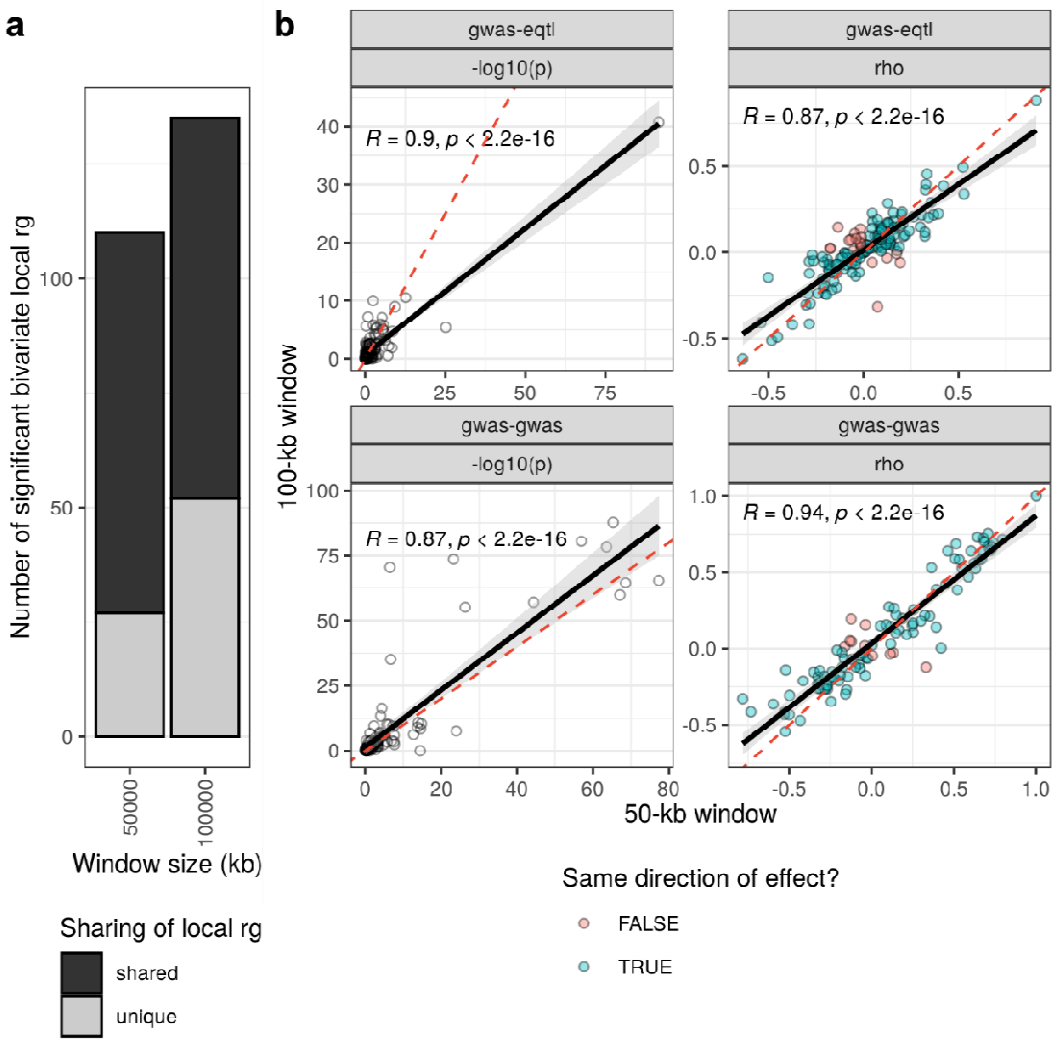
Supplementary Figure 3 Local gene expression and disease trait correlations across 5 LD blocks of interest.

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.



Supplementary Figure 4 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 5 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 phenotypes that could be tested across genic regions with a 50-kb or 100-kb window. Panels indicate whether the pair of phenotypes 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 Results of LAVA using GWASs for AD and PD that exclude UK Biobank by-proxy cases.

Supplementary Table 5 Results of multiple regression analyses.

Supplementary Table 6 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 7 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.