ARTICLE   OPEN

Moving beyond neurons: the role of cell type-specific gene regulation in Parkinson's disease heritability

Regina H. Reynolds [1], Juan Botía[1,2], Mike A. Nalls[3,4], International Parkinson's Disease Genomics Consortium (IPDGC), System Genomics of Parkinson's Disease (SGPD), John Hardy[1,5], Sarah A. Gagliano Taliun[6] and Mina Ryten[1]

Parkinson's disease (PD), with its characteristic loss of nigrostriatal dopaminergic neurons and deposition of α-synuclein in neurons, is often considered a neuronal disorder. However, in recent years substantial evidence has emerged to implicate glial cell types, such as astrocytes and microglia. In this study, we used stratified LD score regression and expression-weighted cell-type enrichment together with several brain-related and cell-type-specific genomic annotations to connect human genomic PD findings to specific brain cell types. We found that PD heritability attributable to common variation does not enrich in global and regional brain annotations or brain-related cell-type-specific annotations. Likewise, we found no enrichment of PD susceptibility genes in brain-related cell types. In contrast, we demonstrated a significant enrichment of PD heritability in a curated lysosomal gene set highly expressed in astrocytic, microglial, and oligodendrocyte subtypes, and in LoF-intolerant genes, which were found highly expressed in almost all tested cellular subtypes. Our results suggest that PD risk loci do not lie in specific cell types or individual brain regions, but rather in global cellular processes detectable across several cell types.

INTRODUCTION

Late-onset sporadic forms of neurodegenerative diseases are devastating conditions imposing an increasing burden on health-care systems worldwide. Currently, 2–3% of the population over 65 years of age are living with Parkinson's disease (PD), making this disorder the most prevalent late-onset neurodegenerative disorder worldwide after Alzheimer's disease.[1] This progressive condition is characterised by the loss of dopaminergic neurons in the substantia nigra pars compacta manifesting clinically as a tremor at rest, muscle rigidity and bradykinesia.[1,2] Existing symptomatic treatments do not alter the course of the disease and their effectiveness declines with time, which makes the identification of potential therapeutic targets of key importance.

The primary focus of PD research to date has been on neurons and, more specifically, nigrostriatal dopaminergic neurons. This focus is driven in part because the death of dopaminergic neurons is primarily responsible for the motor features of PD, but also because the most prominent and distinctive neuropathological findings in PD are the presence of neuronal inclusions, termed Lewy bodies.[1,2] The findings that alpha-synuclein (encoded by the gene *SNCA*) is predominantly expressed in neurons,[2,3] is the major component of Lewy bodies,[3,4] and mutations in *SNCA* give rise to autosomal dominant PD,[5–8] provide a key link between *SNCA* function, neurons and disease pathogenesis. Furthermore, the identification of risk single nucleotide polymorphisms (SNPs) at the *SNCA* locus through genome-wide association studies (GWAS) of sporadic PD[9] provides support for the importance of *SNCA*-

related pathways and, by implication, neurons in both sporadic and Mendelian forms of PD. Despite this neuronal focus, there is also growing evidence to suggest the involvement of other cell types in PD pathogenesis. In particular, astrocytes and microglia have been highlighted;[10,11] for instance, with a recent study demonstrating that blocking the microglial-mediated conversion of astrocytes to an A1 neurotoxic phenotype was neuroprotective in mouse models of sporadic and familial α-synucleinopathy.[12]

In previous work, we applied stratified LD score regression and gene-set enrichment methods to determine if particular functional marks for regulatory activity and gene-set lists were enriched for sporadic PD genetic heritability.[13] We did not observe enrichment for the various brain annotations assessed (this did not include brain-relevant cell types) and in fact found further evidence for the importance of the adaptive and innate immune system.

The increasing power of GWASs (with the most recently published PD GWAS including 37.7K cases, 18.6K 'proxy-cases' and 1.4M controls, resulting in 90 associated loci[14]) coupled with the increased availability of cell-specific gene expression data provides a new opportunity to address the potential cellular specificity of disease heritability, as was elegantly demonstrated for schizophrenia in a study by Skene et al.[15] Brain regions contain a mixture of cell types, such as neurons, microglia and astrocytes, which may exhibit their own specific regulatory features that could be masked when averaging features across cell types. Resolving this question has become increasingly important; with the advent of induced pluripotent stem cell models of disease,

modelling PD in vitro is now possible, and this implies some decision about the cell type of interest. In this study, we addressed cellular heterogeneity through the analysis of genomic regions overlapping regulatory marks or gene expression from cell types within the brain, including neurons. We focus on PD GWAS datasets and use schizophrenia (SCZ) GWAS datasets for comparisons purposes.

RESULTS

Overview of methods

To study the cellular specificity of the heritability of sporadic PD, we compiled brain-related genomic annotations denoting tissue- and cell-type-specific markers of activity. We used several approaches to capture the expression profiles of human brain-related cell types (Fig. 1, see Methods). This was because no single data set had all the desirable properties; namely, data that was human in origin, covered multiple brain regions, had high cellular detail and was derived from large numbers of individuals. Using the largest publicly available GWASs of PD[16] and SCZ,[17] we applied stratified LD score regression (LDSC) to assess enrichment of the common-SNP heritability of PD and SCZ, respectively, for each annotation category. SCZ heritability has been previously shown to be enriched in genes expressed within the central nervous system (CNS) and, more specifically, neuronal cell types,[15,18] and was therefore included as a measure of robustness. For all stratified-LDSC analyses, we report coefficient p-values, which test whether the regression coefficient of an annotation category positively contributes to trait heritability, conditional upon the LDSC baseline model, which accounts for underlying genetic architecture.

PD heritability in human brain

It is well recognised that regional differences in gene expression within human brain and related co-expression modules are driven by differences in the type and density of specific cell types.[19] Therefore, we first used regional data as a means of capturing major cellular profiles. This information, comprising sets of human tissue-specific genes generated by Finucane et al. with GTEx gene expression,[18,20] had the advantage of being comprehensive in terms of sampling across the human CNS, and being robust in that greater than 63 independent samples contributed to the generation of each profile. We confirmed that SCZ heritability was significantly enriched in all 13 brain regions relative to all other tissues, as previously demonstrated by Finucane et al.[18] using the 2014 SCZ GWAS (Fig. 2a, Supplementary Table 1). In contrast, no tissues were enriched for PD heritability, although spinal cord and substantia nigra approached the Bonferroni significance threshold (threshold p-value $= 4.72 \times 10^{-4}$; spinal cord (cervical c-1), p-value $= 1.36 \times 10^{-3}$; substantia nigra, p-value $= 8.96 \times 10^{-4}$). Our comparison of PD and SCZ GWAS iterations across the years revealed the robust nature of the CNS enrichment in SCZ, which was apparent in the first and smallest SCZ GWAS (Supplementary Figure 1). Furthermore, increasing GWAS sample sizes were associated with coefficient p-values becoming more significant, particularly for CNS-related tissues. Interestingly, we also observed an ordering of tissues, with brain regions of greater relevance to disease pathology demonstrating the most significant coefficient p-values in the largest GWAS iterations (e.g. substantia nigra in PD and frontal cortex in SCZ).

However, due to the way these annotations were constructed, related tissues (e.g. brain regions) have overlapping gene sets and therefore may appear enriched as a group. To differentiate among brain regions, we used fine-scale brain expression data generated by Finucane et al. from a brain-only analysis of the 13 GTEx brain regions.[18] We confirmed significant enrichments in the cortex

relative to other brain regions for SCZ, but saw no enrichments for PD (Fig. 2b, Supplementary Table 1).

We also compared the PD and SCZ GWAS results to sets of blood- and brain-specific eQTLs derived from GTEx. We demonstrated an enrichment of SCZ heritability in brain-specific eQTLs and blood-specific eQTLs, but no enrichment of PD heritability in either eQTL annotation (Fig. 3a, Supplementary Table 2). A comparison of eQTLs specific to each brain region revealed no preferential enrichment of disease heritability in one region relative to the others (Fig. 3b, Supplementary Table 2). In summary, these analyses revealed no enrichment of PD heritability in brain annotations, while in contrast, SCZ heritability was highly enriched in both global and specific regional brain annotations.

PD heritability in brain-related cell-type annotations

Given the lack of enrichment of PD heritability in global and regional brain annotations, we wondered whether cellular heterogeneity may be masking signals, and provided more cell-type-specific information the enrichment would become more apparent. Thus, to address the relative importance of brain cell types in PD and SCZ, we generated cell-type-specific annotations from three types of brain-related cell-type-specific data: bulk RNA-sequencing from the Barres group of immunopanned cell types from human temporal lobe cortex;[21] single-cell RNA-sequencing from the Linnarsson group of the adolescent mouse nervous system;[22] and finally, cell-type modules inferred from human tissue-level co-expression networks.[23] Genes were assigned to cell types by fold enrichment (i.e. mean expression in one cell type divided by the mean expression in all other cell types) or module membership in the case of co-expression (module membership is a measure of how correlated a gene's expression is with respect to a module's eigengene).

Each of these datasets came with advantages and disadvantages, which motivated our decision to use all three. The Barres data was based on the analysis of human tissue; however, it covered only one brain region, was derived from a small number of individuals ($n = 14$) who all had an underlying neurological disorder (epilepsy, stroke and glioma), and lacked cellular detail. While the cell-type-specific data provided by the Linnarsson group covered both the central and peripheral nervous system, and contained remarkable cellular detail, it was mouse in origin. Cell-type modules also covered several brain regions, were based on large sample sizes, and importantly, were human in origin. Nevertheless, they were inferred cell types, the definition of which was strongly dependent on the quality of the cell-type markers used to identify them.

Using immunopanning data, we identified a neuronal enrichment for SCZ heritability, but no cell-type enrichment for PD (Fig. 4a, Supplementary Table 3). We questioned whether this lack of cell-type enrichment in PD may result from sampling a tissue which is typically affected only in the later stages of sporadic PD.[2] Thus, we analysed a subset of mouse single-cell data representing tissues affected in earlier stages of sporadic PD, including the enteric nervous system, the substantia nigra and the basal ganglia. Once again, we found no cell-type enrichment for PD heritability (Fig. 4b, Supplementary Table 3). Conversely, we demonstrated a significant enrichment of SCZ heritability in three types of GABAergic medium spiny neurons (MSNs): MSN2, MSN3 and MSN5. This is consistent with the findings reported by Skene et al.[15] Common to all three types of MSN is that they express the D_2 dopamine receptor, a common target of antipsychotic drugs used in SCZ therapy.[24]

To our knowledge, there is currently no single-cell RNA-sequencing data for human striatum or substantia nigra, so we sought to validate our findings using cell-type modules inferred from co-expression networks constructed from human tissue-level
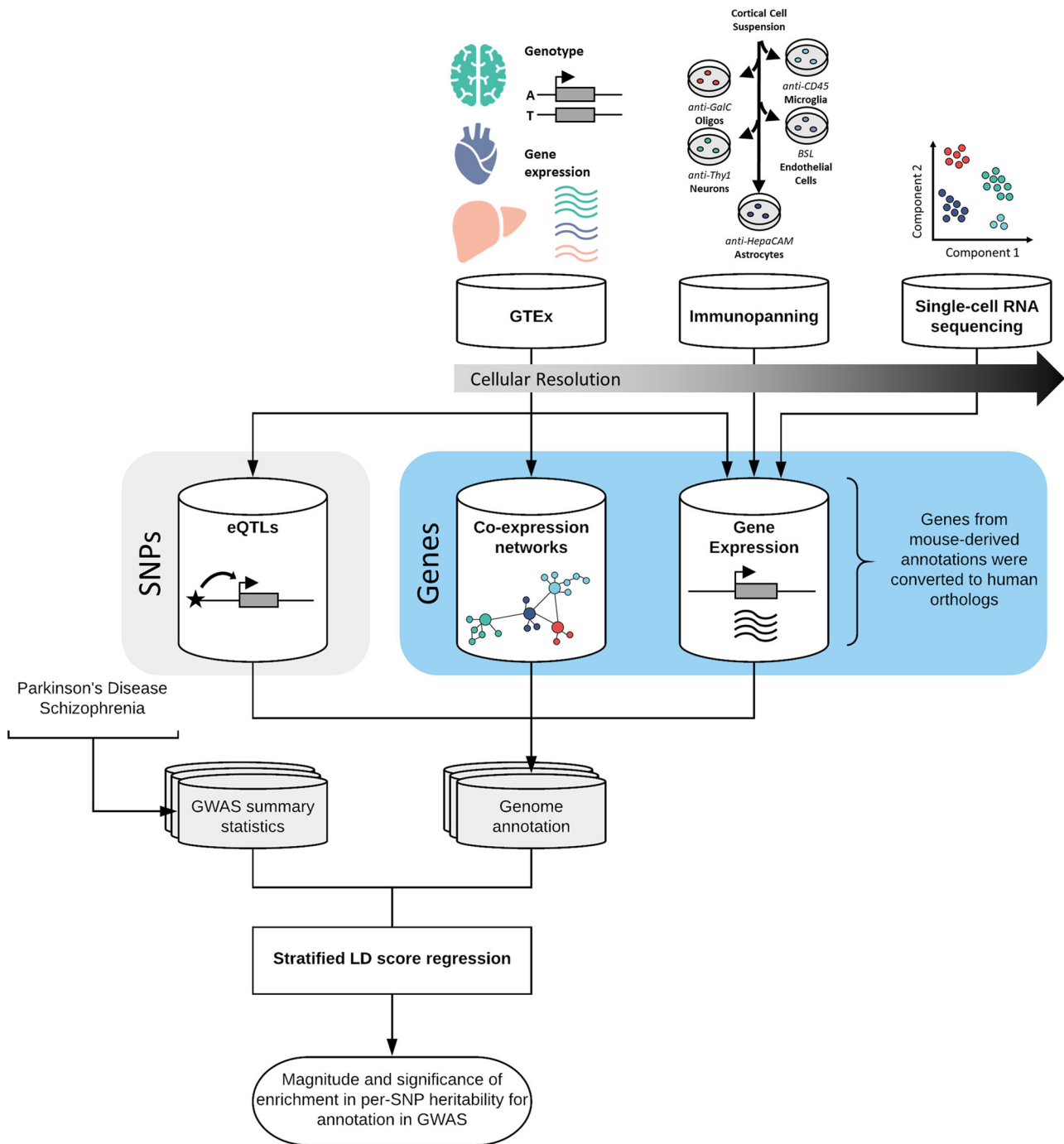
Fig. 1 Overview of approach and datasets used. This study compiled several brain-related genomic annotations reflecting tissue- and cell-type-specific activity, using data generated by the GTEx project,[20] the Barres group[21] and the Linnarsson group.[22] These annotations, each of which varied in their cellular resolution, included: tissue-specific eQTLs (reflecting the effect of genetic variation on gene expression); tissue-specific co-expression networks (reflecting the connectivity of a gene to all other expressed genes in the tissue), and tissue- and cell-type-specific gene expression. All annotations were constructed in a binary format (1 if the SNP is present within the annotation and 0 if not). For annotations where the primary input was a gene, all SNPs with a minor allele frequency > 5% within ± 100 kb of the transcription start and end site were assigned a value of 1. For more details of how each individual annotation was generated see Methods. Stratified LDSC was then used to test whether an annotation was significantly enriched for the common-SNP heritability of PD or SCZ

expression data of the frontal cortex, putamen and substantia nigra. We observed no significant enrichments for PD heritability in any modules, while SCZ heritability was enriched in several neuronal modules, including: brown and turquoise modules in the frontal cortex; blue and dark magenta modules in the putamen; and cyan and darkgrey modules in the substantia nigra (Fig. 5, Supplementary Table 4). All of these modules were enriched for markers of pyramidal S1 neurons, which have previously been associated with SCZ.[15] Furthermore, some modules (brown, turquoise, dark magenta and cyan) were enriched for markers of interneurons and dopaminergic neurons, both of which are implicated in SCZ.[15,24]

a
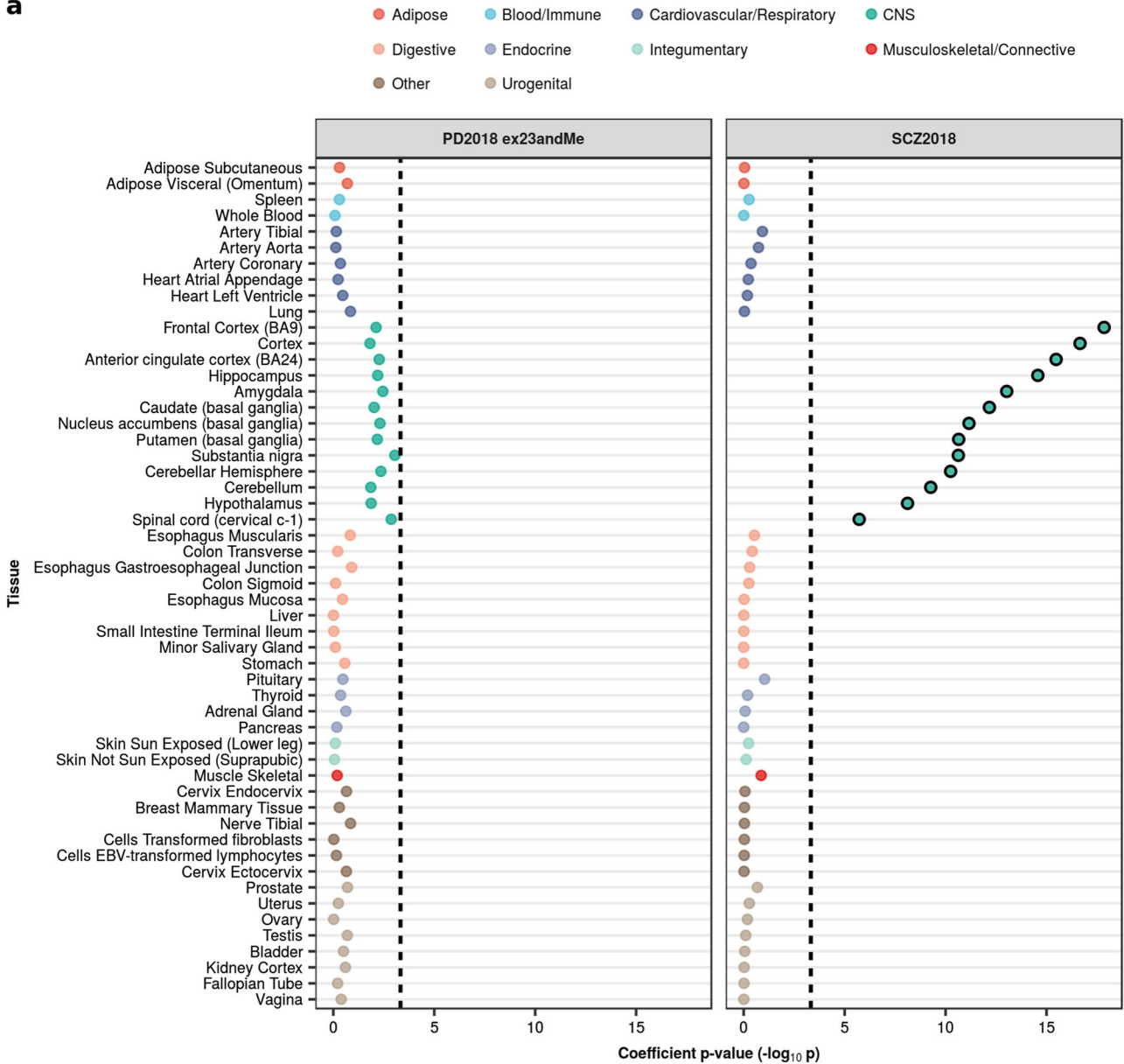
b
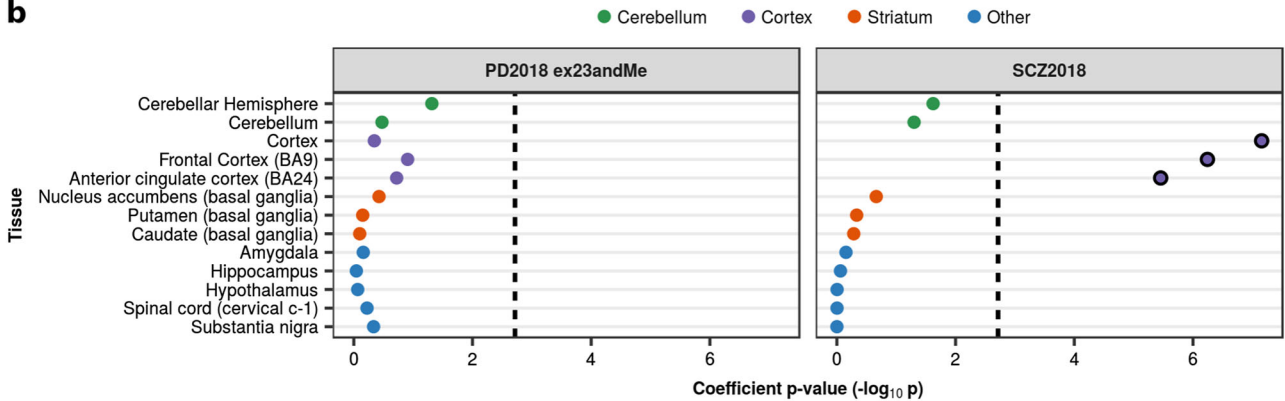
PD susceptibility genes in brain-related cell types

To ensure rigour, we attempted to identify cell types of importance in PD in a separate analysis using expression-

weighted cell-type enrichment (EWCE). This method statistically evaluates whether a set of genes has higher expression in one cell type than expected by chance. Using the same subset of clusters

from the Linnarsson single-cell RNA-sequencing, cell-type specificity values were computed for each gene (i.e. proportion of expression of a gene in a given cell type), and cell-type enrichments of PD susceptibility genes implicated by common-variant studies were estimated (Fig. 6a, see Methods). Susceptibility genes were derived using MAGMA,[25] which estimates the association between each gene and a phenotype while accounting for LD, and from a study by Li et al.[26] using TWAS[27] and coloc,[28] which evaluate the association between eQTLs and GWAS risk loci. We found no significant enrichment of PD susceptibility genes in any of the major cell-type classes (Fig. 6b, Supplementary Table 5) or their cell subtypes (Fig. 6c, Supplementary Table 5).

In summary, our EWCE and stratified LDSC analyses would suggest that PD heritability/susceptibility cannot be attributed to a specific cell type (amongst those tested), unlike what has been observed by us and others for SCZ,[15] wherein a limited set of neuronal cell types have been implicated.

PD heritability in PD-relevant genes sets

Risk loci can operate in several manners, including: a cell-type-/tissue-specific manner, which is only detectable if measured in the "correct" cell type/tissue, or in a pathway-specific manner, which one might expect to be detectable across more than one cell type/tissue. Given our inability to implicate a cell type in PD, we wondered whether the latter scenario of pathway-specific risk might be applicable in PD.

To address this question, we applied stratified LDSC to gene sets implicated in PD by Mendelian forms of PD, functional assays performed in the context of PD-associated mutations, such as the A53T missense mutation in *SNCA*, and rare-variant studies of sporadic PD.[29–34] In particular, we focused on gene sets associated with autophagy,[30,31] the lysosomal system[32] and mitochondrial function.[33,34] Our gene sets were derived either from Gene Ontology terms (autophagy) or curated gene databases (lysosomal, hLGDB; mitochondrial, MitoCarta 2.0; see Methods), developed using literature curation (with a focus on unbiased proteomic studies) and experimental approaches. As with previous stratified LDSC analyses, we included SCZ for comparison purposes. In addition, we used a gene set comprising loss-of-function (LoF)-intolerant genes, as defined by the Exome Aggregation Consortium (ExAC)[35] using their gene-level constraint metric (pLI ≥ 0.9), which has been previously shown enriched for SCZ heritability[17] and thus would serve as a positive control. The overlap between these gene sets was relatively low (Supplementary Figure 15). We identified a significant enrichment of PD heritability in the lysosomal and LoF-intolerant gene set, while SCZ heritability was only found enriched in LoF-intolerant genes, as expected (Fig. 7a, Supplementary Table 6). We ran these analyses with and without a category accounting for "all genes" and found little difference between the estimates provided by stratified LDSC (Supplementary Figure 16), thus we report those results that do not account for "all genes".

Using the same gene sets together with EWCE, we also evaluated whether these PD-implicated gene sets were highly expressed in any of the Linnarsson cell-type classes and their cell

subtypes. Autophagy and lysosomal gene sets were significantly enriched in a limited number of major cell-type classes, with autophagy enriched in oligodendrocytes and cholinergic/monoaminergic neurons, and lysosomal enriched in microglia (Fig. 7b, Supplementary Table 6). The mitochondrial gene set, on the other hand, was significantly enriched in almost all cell-type classes, including astrocytes, oligodendrocytes, oligodendrocyte precursor cells, cholinergic/monoaminergic neurons and telencephalon projecting inhibitory neurons. Likewise, the LoF-intolerant gene set was found significantly enriched in all cell-type classes. As expected, analyses performed on cell subtypes predominantly reflected those performed on the overarching cell-type classes, with significant pathway enrichments observed in cell subtypes associated with the pathway-enriched cell-type classes (Fig. 7c, Supplementary Table 6). For example, all three microglial subtypes (MGL1-3, representing one baseline and two activated microglial subtypes), were enriched for lysosomal genes. Interestingly, the subtype analyses also revealed a significant enrichment of the lysosomal gene set in two astrocytic subtypes, ACNT1 and ACNT2 (both non-telencephalon astrocytes, but protoplasmic and fibrous, respectively), and in two oligodendrocyte subtypes, MFOL1 and MOL1 (myelin forming oligodendrocytes and mature oligodendrocytes, respectively), which was not reflected when using the major cell-type classes, suggesting that analyses performed with cellular subtypes provide greater resolution. It is worth noting that despite higher than expected expression of the lysosomal gene set in astrocytic, microglial and oligodendrocyte subtypes, the lysosomal gene set is ubiquitously expressed across all cellular subtypes (Supplementary Figure 19). In other words, higher than expected expression is not necessarily equivalent to exclusive expression in a cell type.

Taken together these findings provide support for the view that in contrast to the genetic structure of SCZ, PD risk loci operate in a more global manner enriching within pathways and gene sets, which are highly expressed across several cell subtypes.

DISCUSSION

One of the most striking features of PD is the specificity of its neuropathology and clinical symptoms, which has implicated α-synuclein biology in dopaminergic neurons of the substantia nigra pars compacta as a key component of the disease.[1,2] This stands in stark contrast to SCZ, which has a very heterogeneous clinical phenotype and lacks a characteristic neuropathology,[36,37] with a notable absence of pathological lesions and no reported overall neuronal loss.[38] The apparent cellular specificity of PD has encouraged researchers to hypothesise that selective vulnerability is prompted by the action of risk loci in specific cell types; in other words, it is the nature of the cell type itself, which renders it vulnerable. However, given the interrelated nature of brain regions, apparently specific and reproducible patterns of abnormality could also be the result of a more global effect that exposes functional systems (e.g. neural networks) at different times along a disease's natural history, a view now put forward by several independent groups.[39–41] That is, risk loci may not necessarily lie in cellular subtypes or individual brain regions, but in global
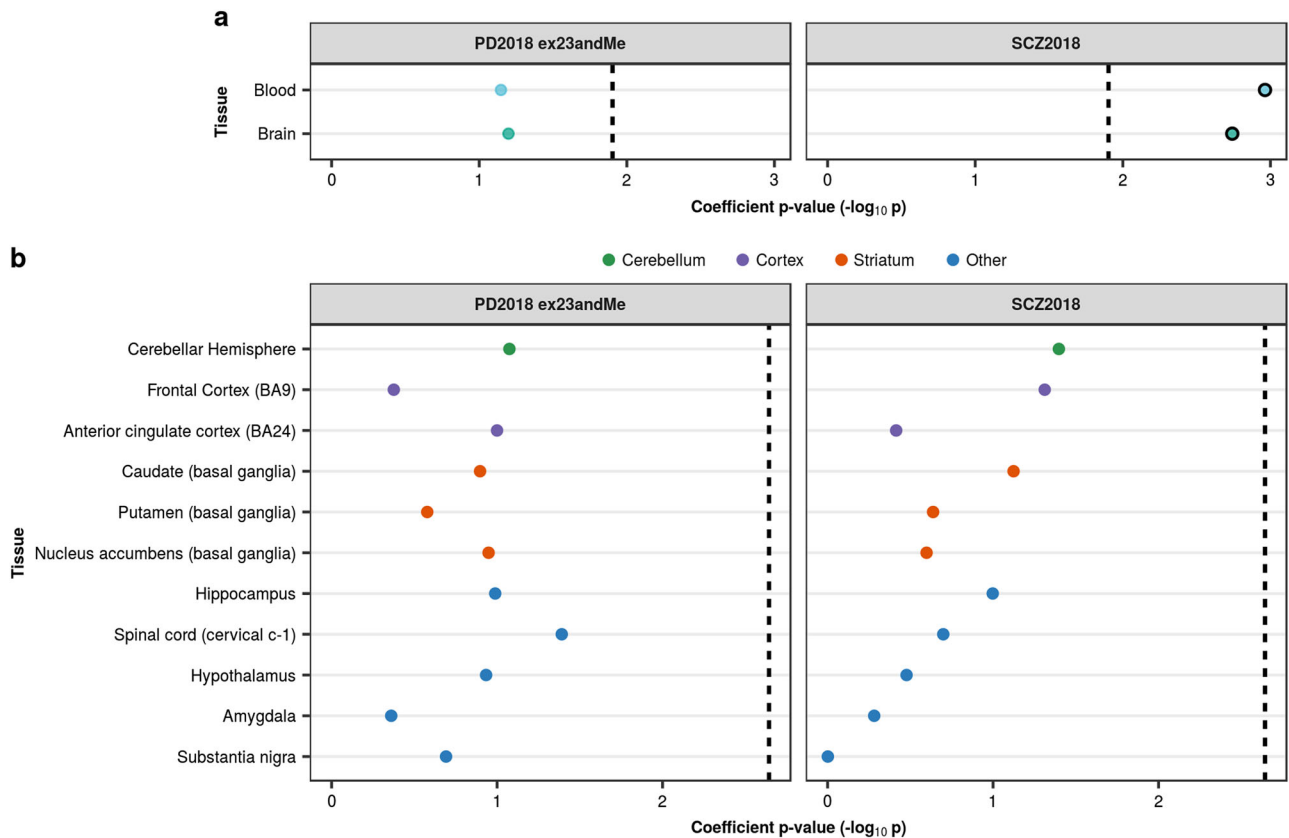
Fig. 3 Enrichment of PD and SCZ common-SNP heritability in tissue-specific eQTL annotations. **a** Stratified LDSC analyses showed significant enrichment of SCZ heritability in brain-specific and blood-specific GTEx eQTLs. **b** A within-brain analysis of GTEx eQTLs showed no significant enrichment of PD and SCZ heritability in one region relative to others. In both analyses, eQTLs were assigned to a tissue/brain region based on their effect size (i.e. the absolute value of the linear regression slope). Tissues were ordered within each tissue category by the coefficient *p*-value obtained for SCZ. The black dashed lines indicate the cut-off for Bonferroni significance (**a**, $p < 0.05/(2 \times 2)$; **b**, $p < 0.05/(2 \times 11)$). Bonferroni-significant results are marked with black borders. The proportion of SNPs accounted for by each annotation (compared to the baseline model), the regression coefficient calculated for the latest PD and SCZ GWASs, and the coefficient *p*-values for previous iterations of the PD and SCZ GWASs are displayed in Supplementary Figures 5–7. Numerical results are reported in Supplementary Table 2

cellular processes to which cellular subtypes have varying vulnerability.
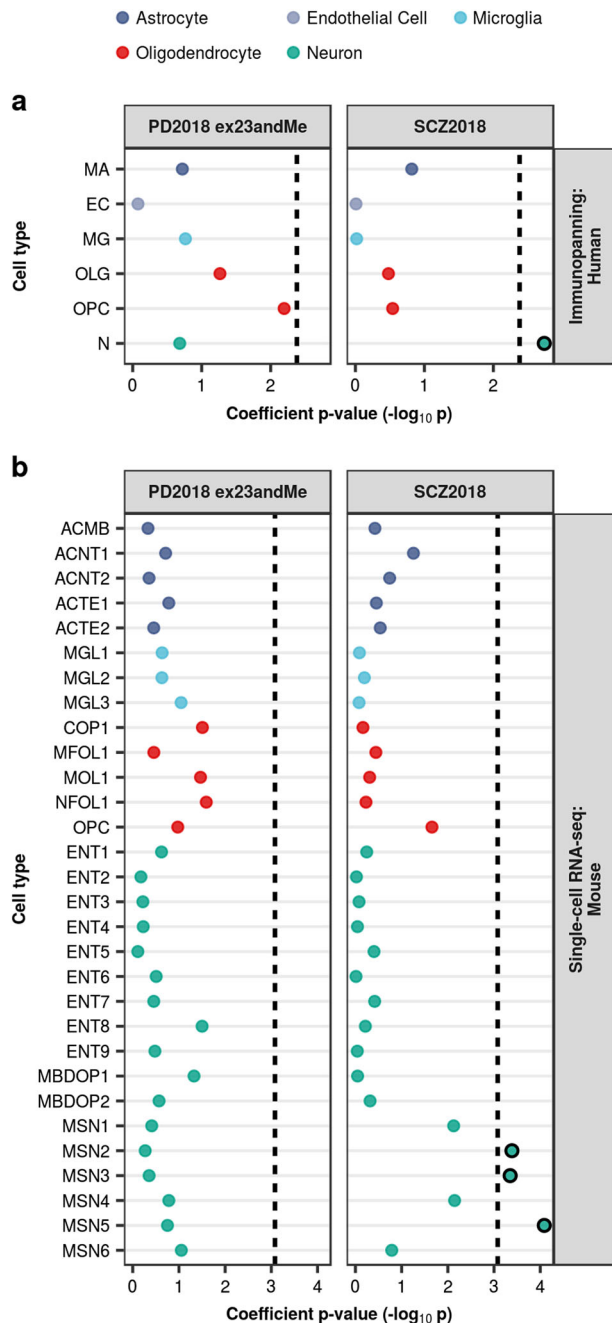
Addressing the question of cellular specificity in sporadic PD in a meaningful manner is now possible due to increasing GWAS sample sizes, increased availability of cell-type-specific gene expression data, and the recent development of robust methodologies. In this study, we used stratified LDSC and EWCE together with several brain-related genomic annotations to connect common-variant genetic findings for PD to specific brain cell types, with SCZ included for comparison purposes. We show that PD heritability does not enrich in global brain annotations or brain-related cell-type-specific annotations, as one might expect if cellular heterogeneity was masking the signal. In contrast, SCZ heritability significantly enriches in global and regional brain annotations and in select neuronal cell types, in line with previous results.[15,18]

One might argue that the lack of PD heritability enrichment in any cell-type-specific categories could be due to PD having a relatively low estimated total heritability; PD heritability estimates range between 20 and 27%[9,42,43] (by comparison, SCZ heritability estimates range between 25 and 80%, depending on whether it is estimated from common SNPs[44] or twin studies.[45,46]) However, we suggest that this is not a complete explanation as significant enrichments have been observed in other GWASs with relatively low overall heritability estimates. For example, in the original stratified LDSC paper they observe enrichment of genomic overlap of histone modifications for the CNS in the ever-smoked

GWAS, specifically in the inferior temporal lobe, and they observed enrichment of fetal brain regulatory features for age at menarche.[47]

Considering our inability to attribute PD heritability/susceptibility to a specific brain-related cell type, we also applied stratified LDSC and EWCE to gene sets implicated in PD (autophagy, lysosomal and mitochondrial gene sets) and SCZ (LoF-intolerant genes), all of which can be considered global pathways/gene sets. Here we show a significant enrichment of PD heritability in the lysosomal and LoF-intolerant gene set, with the former highly expressed in astrocytic, microglial, and oligodendrocyte subtypes and the latter highly expressed in almost all tested cellular subtypes, providing support for the view that PD is a disorder of global pathways working across various cell types, as opposed to specific cell types themselves driving disease risk.

With these results in mind, it is tempting to speculate that PD presents genetically as more of a systemic disorder, with a bias to brain pathology, as opposed to a primary brain disorder. In support of this view, PD-associated risk variants have been found associated with monocytes and the innate immune system,[13,26,48] in addition to lymphocytes, mesendoderm, liver- and fat-cells.[49] Recent work has also demonstrated a causal relationship between BMI and PD,[50] which together with the re-purposing of exenatide (a glucagon-like peptide-1 receptor agonist currently licensed for the treatment of type 2 diabetes) for the potential treatment of PD,[51] highlights the need to look beyond the brain and selective neuronal vulnerability.

Fig. 4 Enrichment of PD and SCZ common-SNP heritability in brain-related cell-type-specific gene expression annotations. Stratified LDSC analyses using cell-type-specific annotations derived from bulk RNA-sequencing of immunopanned cell types from human temporal lobe cortex (**a**) and single-cell RNA-sequencing of the adolescent mouse nervous system (**b**) demonstrated an enrichment of SCZ heritability in neuronal cell types (in particular, medium spiny neurons), but no cell-type enrichment for PD. All cell-type annotations were generated using the top 10% of enriched genes within a cell type compared to all others. Cell types were ordered alphabetically within each overarching cell type category. The black dashed lines indicate the cut-off for Bonferroni significance (**a**, $p < 0.05/(2 \times 6)$; **b**, $p < 0.05/(2 \times 30)$). Bonferroni-significant results are marked with black borders. The proportion of SNPs accounted for by each annotation (compared to the baseline model), the regression coefficient calculated for the latest PD and SCZ GWASs, and the coefficient p-values for previous iterations of the PD and SCZ GWASs are displayed in Supplementary Figs 8–10. Numerical results and cell-type abbreviations are reported in Supplementary Table 3

It is entirely possible that with growing sample sizes, our understanding of the contribution of tissue-specific and brain-region-specific eQTL annotations to PD and SCZ heritability may change. Likewise, given that both TWAS[27] and coloc[28] are dependent upon the accuracy and integrity of eQTL datasets, it is likely that growing samples sizes will alter the list of PD susceptibility genes, potentially resulting in a different cellular expression profile. It is worth noting that there are several tools which evaluate associations between eQTLs and GWAS risk loci to identify susceptibility genes (including coloc, TWAS and others like eCAVIAR[52]). These tools differ in their underlying algorithms and assumptions (e.g. one versus multiple causal variants at a locus), thus motivating the use of multiple methods followed by integration across the results, as performed in the study by Li et al.[26] used here.

Second, our strategy for creating cell-type-specific profiles primarily involved gene expression data and assumed disease relevance only if disease heritability enriched for SNPs within genes with high specific expression. This approach together with the use of the GTEx eQTL dataset (which is likely to be enriched only for eQTLs with larger effect sizes due to power limitations) should capture regulatory SNPs in close proximity to genes of interest. However, as demonstrated in a recent study from Hormozdiari et al. how one chooses to construct an eQTL annotation is fraught with challenges[53] and we recognise that our approach may have produced conservative enrichment estimates. Perhaps more importantly though, our strategy for creating cell-type-specific profiles does not account for the effect of regulatory SNPs that function at longer distances to impact upon gene expression. At present, our ability to address this issue is limited since detecting trans-acting eQTLs has proven to be challenging,[54] especially in human brain.

Third, our approach accounts only for cell type and pathway and, moreover, builds on the assumption that cellular diversity can be sufficiently described by discrete cells classes, which a recent single-cell RNA-sequencing study of the hippocampal CA1 area has called into question.[55] In this study, it was suggested that characterisation of cells requires continuous modes of variation in addition to discrete cell classes; that is, some cell classes exist on a common genetic continuum. Inherent within this spectrum is cellular state, which reflects the physiological condition of a given cell, whether it be the degree of differentiation or activation in response to a stimulus. There may be cellular states that we have not assayed or captured which harbour PD heritability enrichments. Furthermore, one would expect preferential enrichment of pathways in specific cell types/subtypes to vary dependent on their physiological profile. In view of increasing evidence for the

There are several caveats to this study aside from the limitations posed by the current PD GWAS, which does not capture the impact of rare variation or some forms of structural variation. These caveats include the quality of our annotations, the strategies employed to generate them and, perhaps most critically, the annotations we cannot account for.

First, the quality of our annotations is especially pertinent in the case of the gene sets used to reflect various PD-implicated pathways. While lysosomal and mitochondrial gene sets were derived from rigorously curated gene databases, with a focus on unbiased proteomic and localisation studies, the autophagy list stemmed from Gene Ontology, which has not undergone the same meticulous curation. The noise introduced by potentially inaccurate annotation could affect our ability to detect heritability enrichments. Furthermore, the quality of eQTL datasets, reflected in their power to detect an eQTL, is dependent upon sample size.
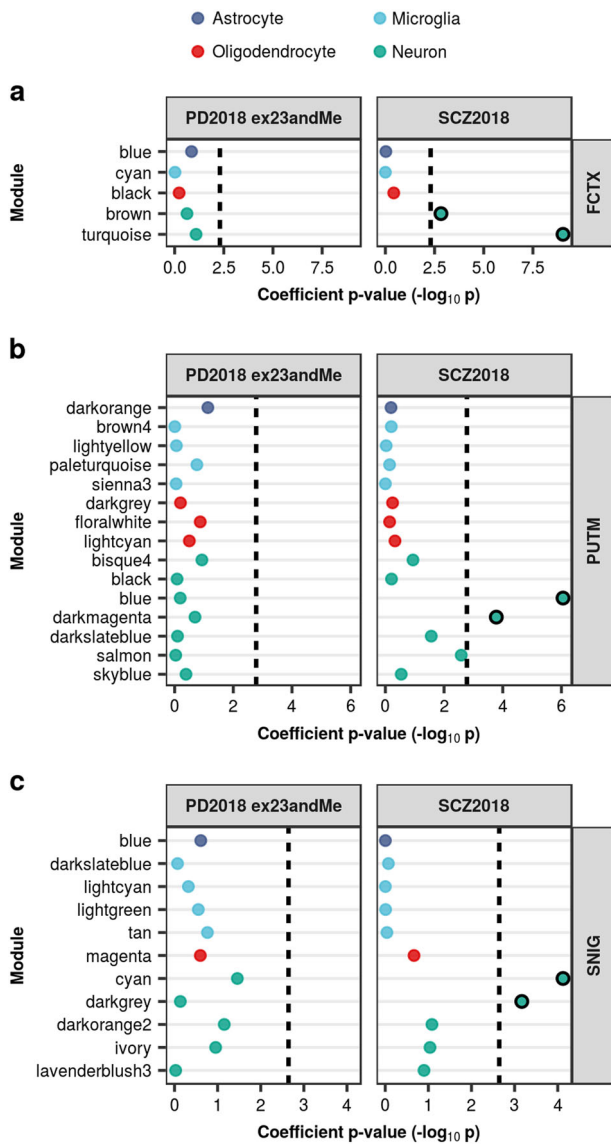
Fig. 5 Enrichment of PD and SCZ common-SNP heritability in cell-type modules inferred from human tissue-level co-expression networks. Stratified LDSC analyses using cell-type-specific co-expression modules from frontal cortex (**a**), putamen (**b**), and substantia nigra (**c**) demonstrated significant enrichment of SCZ heritability in certain neuronal modules across all three tissues, but no enrichment for PD heritability. Genes were assigned to cell-type modules by module membership. Cell-type-specific modules were ordered alphabetically within each overarching cell type category. The black dashed lines indicate the cut-off for Bonferroni significance (**a**, $p < 0.05/(2 \times 5)$; **b**, $p < 0.05/(2 \times 15)$; **c**, $p < 0.05/(2 \times 11)$). Bonferroni-significant results are marked with black borders. The proportion of SNPs accounted for by each annotation (compared to the baseline model), the regression coefficient calculated for the latest PD and SCZ GWASs, and the coefficient p-values for previous iterations of the PD and SCZ GWASs are displayed in Supplementary Figures 11–14. Numerical results and module descriptions are reported in Supplementary Table 4. FCTX, frontal cortex; PUTM, putamen; SNIG, substantia nigra

association between PD and the innate immune system,[13,26,48] we think that cellular state is likely to be an important factor, which we cannot fully assess at this stage.

In conclusion, our results add to a growing body of evidence in support of the view that PD risk loci may not lie entirely in those cell types that display the disease's characteristic neuropathology,

but instead in global cellular processes, with effects in a range of cellular subtypes. This view has significant implications for disease modelling, with a choice of model perhaps based upon the cell type, which best reflects the process of interest, as opposed to the cell type which demonstrates the highest burden of α-synuclein aggregates. Likewise, viewing PD as a systemic disorder may have implications for potential drug re-purposing, as in the case of exenatide. Thus, our work here may have wider implications in terms of understanding neurodegenerative disorders more generally as disorders of key cellular processes rather than disorders driven solely by specific cell types.

METHODS

Stratified LD score regression (LDSC)
We applied stratified LDSC[47] (see URLs, Supplementary information) to determine if various categories of genomic annotations (marking tissue- or cell-type-specific activity, as summarised in Annotation datasets) were enriched for heritability of various GWASs (see GWAS datasets below). LDSC exploits the expected relationships between true association signals and surrounding local linkage disequilibrium (LD) to correct out confounding biases, such as cryptic relatedness and population stratification, and arrive at unbiased estimates of genetic heritability within a given set of SNPs (here stratified according to whether they were located within genomic annotation regions). Following the procedure employed by Finucane et al.,[47] we added annotation categories individually to the baseline model (version 1.1, see URLs, Supplementary information). We used HapMap Project Phase 3 (HapMap3)[56] SNPs for the regression, and 1000 Genomes Project[57] Phase 3 European population SNPs for the LD reference panel. We only partitioned the heritability of SNPs with minor allele frequency >5%, and we excluded the MHC region from analysis due to the complex and long-range LD patterns in this region. To map SNPs to genes, we used the SNPlocs.Hsapiens.dbSNP144.GRCh37 R package (dbSNP build 144 and GRCh37 coordinates).[58]

For all stratified LDSC analyses, we report a one-tailed p-value (coefficient p-value) based on the coefficient z-score outputted by stratified LDSC. A one-tailed test was used as we were only interested in annotation categories with a significantly positive regression coefficient (i.e. the annotation positively contributed to trait heritability, conditional upon the baseline model, which accounts for the underlying genetic architecture). We looked at three versions of PD GWAS summary statistics and four versions of SCZ, and for each set of analyses we corrected for multiple testing of the GWASs across the number of annotation categories, resulting in Bonferroni significance thresholds for each set of analyses.

Annotation datasets
Tissue-specific gene expression. Annotation files were generated by Finucane et al.,[18] using GTEx V6P gene expression,[20] and obtained from Alkes Price's group data repository (see URLs, Supplementary information). Briefly, for each GTEx tissue, genes were ranked by a computed t-statistic reflecting their specific expression within that tissue versus all other tissues, excluding those that were from a similar tissue category (e.g. expression in cortex samples was compared to expression in all other tissues except other brain regions; see Supplementary Table 2 from Finucane et al.[18] for t-statistic tissue categories). The top 10% of expressed genes from each tissue was selected and a 100-kb window was added around their transcribed regions to obtain a tissue-specific gene expression annotation. As described by Finucane et al.,[18] the two parameters (proportion of genes selected and window size around each gene) were selected following testing of six different parameter settings, which identified 10% and 100 kb as the settings producing the most significant P-values for identifying enrichment in disease-relevant tissues. For the within-brain analysis, tissues were restricted to the 13 brain regions found in GTEx, including: amygdala, anterior cingulate cortex (BA24), caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens, putamen, spinal cord (cervical c-1), substantia nigra.

Tissue-specific eQTLs. From the GTEx Portal (V7, accessed 04/16/18, see URLs, Supplementary information), we downloaded all SNP-gene (expression quantitative trait loci, eQTL) association tests (including non-significant tests) for blood (to allow for a blood-brain comparison) and 11 of the 13 available brain regions.[20] To reduce redundancy across the
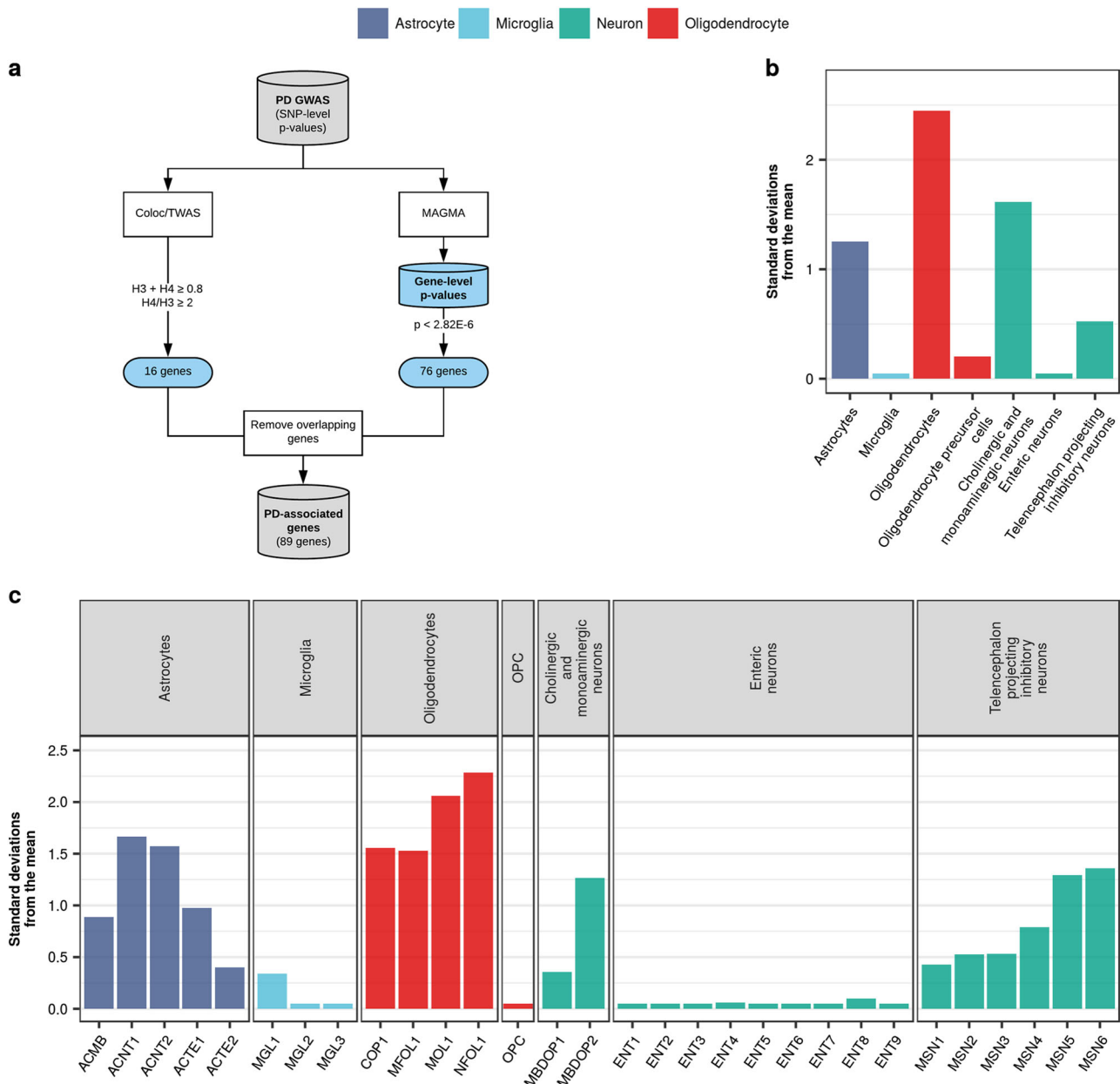
Fig. 6 PD susceptibility genes do not enrich in brain-related cell types. **a** PD susceptibility genes were derived from MAGMA analyses and a study attempting to prioritise genes in PD using TWAS and colocalisation analyses.[26] Genes overlapping between the two sets were removed, resulting in a list of 89 genes. Bootstrapping tests performed using the EWCE method revealed no enrichment of PD susceptibility genes in the major cell-type classes (**b**) or their cell subtypes (**c**) from the Linnarsson single-cell RNA-sequencing dataset. Gene lists and numerical results are available in Supplementary Table 5

brain regions, we excluded cortex and cerebellum, and instead included frontal cortex, anterior cingulate cortex and cerebellar hemisphere. We performed an FDR correction for each tissue and included all SNP-gene associations that passed FDR <5% in our downstream analyses. For the blood-brain comparison, eQTLs from all 11 brain regions were combined to form one brain category. eQTLs that replicated across brain regions were collapsed into one entry and allocated an effect size (i.e. the absolute value of the linear regression slope) equal to that of the maximum effect size observed across the brain regions. Finally, eQTLs were assigned to either blood or brain by their effect size. A similar approach was used for the within-brain analysis, where eQTLs were assigned to one of the 11 brain regions based on effect size.

Cell-type-specific gene expression. Cell-type-specific annotations were constructed using gene expression data from the Barres group[21] and the Linnarsson group[22] (see URLs, Supplementary information), which was generated using bulk RNA-sequencing and single-cell RNA-sequencing, respectively. Due to the disparate nature of the RNA-sequencing methods, each dataset was analysed separately. Common to both analyses was the calculation of an enrichment value for each gene in each cell type. Enrichment was calculated as: gene expression in one cell type divided by the average gene expression across all other cell types. We thereafter selected the top 10% of genes enriched within each cell type and added a 100 kb window to reflect the approach used by Finucane et al.[18] When using the Barres data, we averaged gene expression across samples of the same cell type, filtered genes on the basis of an FPKM ≥1 in at least one cell type (this equates to ~66% of all genes with FPKM >0.1, which was set by Zhang et al.[21] as the threshold for minimum gene expression), and then calculated gene enrichment. Our detection threshold of FPKM ≥1 was employed on the basis that smaller thresholds tend to produce large and misleading enrichments.[59] The Linnarsson data was available with gene expression aggregated by sub-cell type/cluster. Genes were filtered on the
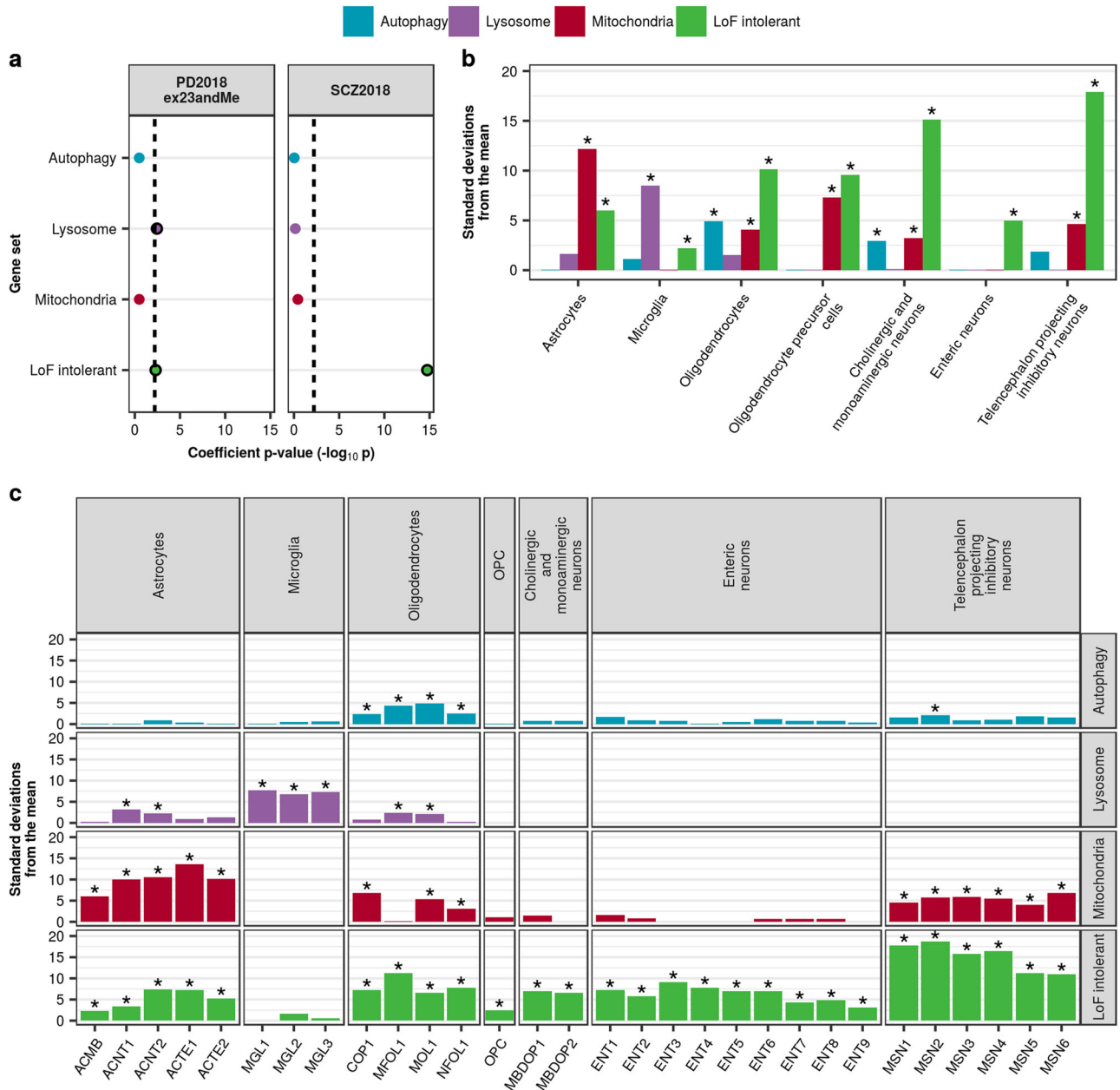
Fig. 7 PD heritability enriches in lysosomal and LoF-intolerant gene sets which are ubiquitously expressed. **a** Stratified LDSC analyses using gene sets implicated in PD demonstrated a significant enrichment of PD heritability in the lysosomal and LoF-intolerant gene sets. The black dashed lines indicate the cut-off for Bonferroni significance ($p < 0.05/(2 \times 4)$). Bonferroni-significant results are marked with black borders. The proportion of SNPs accounted for by each annotation (compared to the baseline model), the regression coefficient calculated for the latest PD and SCZ GWASs, and the coefficient p-values for previous iterations of the PD and SCZ GWASs are displayed in Supplementary Figures 17 and 18. Bootstrapping tests performed using the EWCE method demonstrated enrichment of autophagy, lysosomal and mitochondrial gene sets in specific cell-type classes (**b**) and their cell subtypes (**c**) from the Linnarsson single-cell RNA-sequencing dataset. Asterisks denote significance at $p < 0.05$ after correcting for multiple testing with the Benjamini-Hochberg method over all gene sets and cell types tested. Gene lists and numerical results are reported in Supplementary Table 6

basis of expression >0, enrichment was calculated and a subset of the 265 identified clusters were used as annotations. Mouse genes were converted to human orthologs using Biomart (of the 23,368 genes with expression >0 in at least one cluster, 16,025 genes were converted to human orthologs. A list of those genes that did not convert can be found in Supplementary Table 3).

Cell-type-specific co-expression modules. Co-expression networks for frontal cortex, putamen and substantia nigra were constructed using GTEx V6 gene expression,[20] the WGCNA R package[60] and post-processing with k-means[61] (see URLs, Supplementary information), as described by Botia

et al.[23] Modules were assigned to cell types using the userListEnrichment R function implemented in the WGCNA R package, which measures enrichment between module-assigned genes and defined brain-related lists[19,62–65] using a hypergeometric test. Genes assigned to modules significantly enriched for brain-related cell-type markers of predominantly one cell type, with a module membership of ≥0.5, were allocated a cell-type "label" of neuron, microglia, astrocyte or oligodendrocyte and considered cell-type specific. Module membership values range between 0 and 1, with 1 indicating that a gene's expression is highly correlated with the module eigengene. An eigengene is defined as the first principal component of a given module and can be considered representative of the

Table 1. Summary of GWAS datasets

Disease	First author, Year	N cases	N controls	PMID	Reference
PD	IPDGC Consortium, 2011	5,333	12,019	21738488	[71]
PD	Nalls, 2014	13,708	95,282	25064009	[14]
PD	Nalls, 2018 (excluding 23andMe contributions)[a]	33,674 (18,618 proxy cases from UK Biobank)	449,037		[16]
SCZ	SCZ Consortium, 2011	9,394	12,462	21926974	[72]
SCZ	Ripke, 2013	14,395	18,705	23974872	[73]
SCZ	SCZ Consortium, 2014 (EUR subset)	33,640	43,456	25056061	[74]
SCZ	Pardiñas, 2018	40,675	64,643	29483656	[17]

[a]Access to PD 2018 summary statistics (excluding 23andMe contributions) was provided by Mike A Nalls, with permissions from IPDGC and SGPD

gene expression profiles within the module, as it summarises the largest amount of variance in expression. As with tissue-specific and cell-type-specific gene expression annotations, we added a 100 kb window around each gene. To ensure we were using consistent and reproducible modules, we evaluated preservation of GTEx modules with modules derived from co-expression networks constructed for frontal cortex, putamen and substantia nigra using gene expression data from the United Kingdom Brain Expression Consortium (UKBEC).[61,66] Preservation values were calculated using the modulePreservation function from WGCNA. Preservation is a measure of how well the connectivity and correlation structures within the genes of a network´s module are preserved in another tissue expression dataset. The final preservation score we report is Z-summary-press, which is an aggregation of a number of measures (see Langfelder et al.[67] for a detailed explanation). Values under 2 denote no preservation whereas values greater than 2 suggest preservation, and values over 10 suggest strong preservation of the module in the tissue tested. We included only GTEx modules with a preservation value >2. Preservation values and module descriptions are reported in Supplementary Table 4; the latter can also be viewed at https://snca.atica.um.es/coexp/Run/Catalog/.

Gene sets. We investigated three gene sets with previous biological support for involvement in PD: autophagy, lysosomal and mitochondrial.[29–34] The autophagy gene set included all genes associated with the Gene Ontology terms: GO:0006914 ("autophagy") and GO:0005776 ("autophagosome"), as derived from the GO C5 collection of the MSigDB database (v5.2). Lysosomal genes were downloaded from the Human Lysosome Gene Database (hLGDB, see URLs, Supplementary information).[68] All genes reported lysosomal by any of the listed sources (9 of the 16 were unbiased proteomic studies) were used. Mitochondrial genes were obtained from Human MitoCarta 2.0, an inventory of human genes with the strong support of mitochondrial localisation based on literature curation, proteomic analyses and epitope tagging/microscopy (see URLs, Supplementary information).[69] In addition, we used a gene set comprising loss-of-function (LoF)-intolerant genes, as defined by the Exome Aggregation Consortium (ExAC)[35] using their gene-level constraint metric (pLI ≥0.9), and constructed an "all genes" gene set. The "all genes" gene set was constructed by extracting all genes from BioMart (using the homo sapiens GRCh37 library, as with all other annotations). For stratified LDSC analyses, an additional window of 100 kb was added around genes. The genes comprising these lists are available in Supplementary Table 6. Overlap between gene sets was determined using Intervene, a command line tool and web application that computes and visualises intersections of gene sets (see URLs, Supplementary information).[70]

GWAS datasets
Table 1[14,16,17,71–74]

MAGMA: assessing gene-level enrichment
Gene-level *p*-values were calculated with the Nalls et al.[16] PD GWAS (excluding 23andMe contributions) using MAGMA v1.06 (see URLs, Supplementary information),[25] which tests the joint association of all SNPs in a gene with the phenotype while accounting for LD between SNPs. SNPs were mapped to genes using NCBI definitions (GRCh37, annotation release 105); only genes in which at least one SNP mapped were included in downstream analyses. Gene boundaries were defined as the region from transcription start site to transcription stop site. In addition, we added a window of 35 kb upstream and 10 kb downstream of each gene. This choice was based on (1) most transcriptional regulatory elements fall within this interval,[75] (2) a fraction of GWAS risk loci lie outside gene boundaries and may regulate gene expression[76] and (3) previous work in the field of pathway analysis.[17,77,78] Furthermore, the MHC region on chromosome 6 (chr6: 25500000–33500000, human genome assembly GRCh37) was excluded. The gene *p*-value was computed based on the mean association statistic of SNPs within a gene, with genome-wide significance set to $p < 2.82 \times 10^{-6}$, and LD was estimated from the European subset of 1000 Genomes Phase 3[57].

Evaluating enrichment of PD-associated genes and gene sets
Expression-weighted cell-type enrichment (EWCE, see URLs, Supplementary information)[79] was used to determine whether PD-associated genes or gene sets have higher expression within a particular cell type than expected by chance. As our input, we used the same subset of clusters from the Linnarsson single-cell RNA-sequencing dataset used in stratified LDSC, in addition to a target gene list. For each gene in the Linnarsson dataset, we estimated its cell-type specificity i.e. the proportion of total expression of a gene found in one cell type compared to all cell types, using the 'generate.celltype.data' function of the EWCE package. EWCE with the target list was run with 100,000 bootstrap lists. We controlled for transcript length and GC-content biases by selecting bootstrap lists with comparable properties to the target list. P-values were corrected for multiple testing using the Benjamini-Hochberg method overall cell types and gene lists tested. We performed the analysis with major cell-type classes (e.g. "astrocyte", "microglia", "enteric neurons", etc.) and subtypes of these classes (e.g. ACNT1 ["Non-telencephalon astrocytes, protoplasmic"], ACNT2 ["Non-telencephalon astrocytes, fibrous"], etc.). Data are displayed as standard deviations from the mean, and any values < 0, which reflect a depletion of expression, are displayed as 0.

PD susceptibility genes. PD susceptibility genes were derived from our own MAGMA analyses and a study attempting to prioritise genes in PD using TWAS and colocalisation analyses (Supplementary Table 1 in ref.[26]). The genes comprising these lists are available in Supplementary Table 5. In the case of MAGMA, only those genes passing genome-wide significance ($p < 2.82 \times 10^{-6}$) were used. In the case of TWAS/coloc, only those eQTL-gene associations found within dorsolateral prefrontal cortex tissue, which were both TWAS and coloc hits (as defined in ref.[26]) were used.

Reporting Summary
Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY
Datasets analysed in this study are all derived from publicly available resources (see URLs, Supplementary information). All data generated during this study are included in this published article's supplementary information.

CODE AVAILABILITY

Open source software is available on GitHub for all tools used (see URLs, Supplementary information).

INTERNATIONAL PARKINSON'S DISEASE GENOMICS CONSORTIUM (IPDGC)

Alastair J. Noyce[1,7], Aude Nicolas[3], Mark R. Cookson[3], Sara Bandres-Ciga[3], J. Raphael Gibbs[3], Dena G. Hernandez[3], Andrew B. Singleton[3], Xylena Reed[3], Hampton Leonard[3], Cornelis Blauwendraat[3,8], Faraz Faghri[3,9], Jose Bras[5,10], Rita Guerreiro[5,10], Arianna Tucci[10], Demis A. Kia[10], Henry Houlden[10], Helene Plun-Favreau[10], Kin Y. Mok[10], Nicholas W. Wood[10], Ruth Lovering[10], Lea R'Bibo[10], Mie Rizig[10], Viorica Chelban[10], Daniah Trabzuni[10,11], Manuela Tan[12], Huw R. Morris[12], Ben Middlehurst[13], John Quinn[13], Kimberley Billingsley[13], Peter Holmans[14], Kerri J. Kinghorn[15], Patrick Lewis[16], Valentina Escott-Price[17], Nigel Williams[17], Thomas Foltynie[18], Alexis Brice[19], Fabrice Danjou[19], Suzanne Lesage[19], Jean-Christophe Corvol[19], Maria Martinez[19,20], Anamika Giri[21,22], Claudia Schulte[21,22], Kathrin Brockmann[21,22], Javier Simón-Sánchez[21,22], Peter Heutink[21,22], Thomas Gasser[21,22], Patrizia Rizzu[22], Manu Sharma[23], Joshua M. Shulman[24,25], Laurie Robak[24], Steven Lubbe[26], Niccolo E. Mencacci[27], Steven Finkbeiner[28,29], Codrin Lungu[30], Sonja W. Scholz[31], Ziv Gan-Or[32], Guy A. Rouleau[32], Lynne Krohan[33], Jacobus J. van Hilten[34], Johan Marinus[34], Astrid D. Adarmes-Gómez[35], Inmaculada Bernal-Bernal[35], Marta Bonilla-Toribio[35], Dolores Buiza-Rueda[35], Fátima Carrillo[35], Mario Carrión-Claro[35], Pablo Mir[35], Pilar Gómez-Garre[35], Silvia Jesús[35], Miguel A. Labrador-Espinosa[35], Daniel Macias[35], Laura Vargas-González[35], Carlota Méndez-del-Barrio[35], Teresa Periñán-Tocino[35], Cristina Tejera-Parrado[35], Monica Diez-Fairen[36], Miquel Aguilar[36], Ignacio Alvarez[36], María Teresa Boungiorno[36], Maria Carcel[36], Pau Pastor[36], Juan Pablo Tartari[36], Victoria Alvarez[37], Manuel Menéndez González[37], Marta Blazquez[37], Ciara Garcia[37], Esther Suarez-Sanmartin[37], Francisco Javier Barrero[38], Elisabet Mondragon Rezola[39], Jesús Alberto Bergareche Yarza[39], Ana Gorostidi Pagola[39], Adolfo López de Munain Arregui[39], Javier Ruiz-Martínez[39], Debora Cerdan[40], Jacinto Duarte[40], Jordi Clarimón[41,42], Oriol Dols-Icardo[41,42], Jon Infante[42,43,44], Juan Marín[42,45], Jaime Kulisevsky[42,45], Javier Pagonabarraga[42,45], Isabel Gonzalez-Aramburu[43], Antonio Sanchez Rodriguez[43], María Sierra[43], Raquel Duran[46], Clara Ruz[46], Francisco Vives[46], Francisco Escamilla-Sevilla[47], Adolfo Mínguez[47], Ana Cámara[48], Yaroslau Compta[48], Mario Ezquerra[48], Maria Jose Marti[48], Manel Fernández[48], Esteban Muñoz[48], Rubén Fernández-Santiago[48], Eduard Tolosa[48], Francesc Valldeoriola[48], Pedro García-Ruiz[49], Maria Jose Gomez Heredia[50], Francisco Perez Errazquin[50], Janet Hoenicka[51], Adriano Jimenez-Escrig[52], Juan Carlos Martínez-Castrillo[52], Jose Luis Lopez-Sendon[52], Irene Martínez Torres[53], Cesar Tabernero[54], Lydia Vela[55], Alexander Zimprich[56], Lasse Pihlstrom[57], Sulev Koks[58,59,60], Pille Taba[61], Kari Majamaa[62,63], Ari Siitonen[62,63], Njideka U. Okubadejo[64], Oluwadamilola O. Ojo[64]

[7]Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, QMUL, London, UK; [8]National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA; [9]Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA; [10]Department of Molecular Neuroscience, UCL, London, UK; [11]Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh 11211, Saudi Arabia; [12]Department of Clinical Neuroscience, University College London, London, UK; [13]Institute of Translational Medicine, University of Liverpool, Liverpool, UK; [14]Biostatistics & Bioinformatics Unit, Institute of Psychological Medicine and Clinical Neuroscience, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff, UK; [15]Institute of Healthy Ageing, University College London, London, UK; [16]University of Reading, Reading, UK; [17]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK; [18]UCL Institute of Neurology, London, UK; [19]Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, AP-HP, Pitié-Salpêtrière Hospital, Paris, France; [20]Paul Sabatier University, Toulouse, France; [21]Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany; [22]DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany; [23]Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tubingen, Tubingen, Germany; [24]Departments of Neurology, Neuroscience, and Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas, USA; [25]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas, USA; [26]Ken and Ruth Davee Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; [27]Northwestern University Feinberg School of Medicine, Chicago, IL, USA; [28]Departments of Neurology and Physiology, University of California, San Francisco, CA, USA; [29]Gladstone Institute of Neurological Disease, Taube/Koret Center for Neurodegenerative Disease Research, San Francisco, CA, USA; [30]National Institutes of Health Division of Clinical Research, NINDS, National Institutes of Health, Bethesda, MD, USA; [31]Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA; [32]Montreal Neurological Institute and Hospital, Departments of Neurology and Neurosurgery, Department of Human Genetics, McGill University, Montréal, QC H3A 0G4, Canada; [33]Department of Human Genetics, McGill University, Montréal, QC H3A 0G4, Canada; [34]Department of Neurology, Leiden University Medical Center, Leiden, Netherlands; [35]Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain; [36]Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona, Spain; [37]Hospital Universitario Central de Asturias, Oviedo, Spain; [38]Hospital Universitario Parque Tecnologico de la Salud, Granada, Spain; [39]Instituto de Investigación Sanitaria Biodonostia, San Sebastián, Spain; [40]Hospital General de Segovia, Segovia, Spain; [41]Memory Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain; [42]Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain; [43]Hospital Universitario Marqués de Valdecilla-IDIVAL, Santander, Spain; [44]University of Cantabria, Santander, Spain; [45]Movement Disorders Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain; [46]Centro de Investigacion Biomedica, Universidad de Granada, Granada, Spain; [47]Hospital Universitario Virgen de las Nieves, Instituto de Investigación Biosanitaria de Granada, Granada, Spain; [48]Hospital Clinic de Barcelona, Barcelona, Spain; [49]Instituto de Investigación Sanitaria Fundación Jiménez Díaz, Madrid, Spain; [50]Hospital Universitario Virgen de la Victoria, Malaga, Spain; [51]Institut de Recerca Sant Joan de Déu, Barcelona, Spain; [52]Hospital Universitario Ramón y Cajal, Madrid, Spain; [53]Department of Neurology, Instituto de Investigación Sanitaria La Fe, Hospital Universitario y Politécnico La Fe, Valencia, Spain; [54]Hospital General de Segovia, Segovia, Spain; [55]Department of Neurology, Hospital Universitario Fundación Alcorcón, Madrid, Spain; [56]Department of Neurology, Medical University of Vienna, Vienna, Austria; [57]Department of Neurology, Oslo University Hospital, Oslo, Norway; [58]Department of Pathophysiology, University of Tartu, Tartu, Estonia; [59]Department of Reproductive Biology, Estonian University of Life Sciences, Tartu, Estonia; [60]Perron Institute for Neurological and Translational Science, Perth, WA, Australia; [61]Departments of Neurology and Neurosurgery, University of Tartu, Tartu, Estonia; [62]Institute of Clinical Medicine, Department of Neurology, University of Oulu, Oulu, Finland; [63]Department of Neurology and Medical Research Center, Oulu University Hospital, Oulu, Finland and [64]University of Lagos, Lagos, Lagos State, Nigeria

SYSTEM GENOMICS OF PARKINSON'S DISEASE (SGPD)

Toni Pitcher[65,66], Tim Anderson[65,66], Steven Bentley[67], Javed Fowdar[67], George Mellick[67], John Dalrymple-Alford[68], Anjali K. Henders[69], Irfahan Kassam[69], Grant Montgomery[69], Julia Sidorenko[69], Futao Zhang[69], Angli Xue[69], Costanza L. Vallerga[69], Leanne Wallace[69], Naomi R. Wray[69,70], Jian Yang[69,70], Peter M. Visscher[69,70], Jacob Gratten[69,70], Peter A. Silburn[70], Glenda Halliday[71], Ian Hickie[71], John Kwok[71], Simon Lewis[71], Martin Kennedy[72], John Pearson[72]

[65]New Zealand Brain Research Institute, Christchurch, New Zealand; [66]Department of Medicine, University of Otago, Christchurch, New Zealand; [67]Griffith Institute for Drug Discovery, Griffith University, Brisbane, Australia; [68]Dept. Psychology, University of Canterbury, New Zealand Brain Research Institute, Christchurch, New Zealand; [69]Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; [70]Queensland Brain Institute, University of Queensland, Brisbane, Australia; [71]Brain and Mind Centre, Sydney Medical School, University of Sydney, Sydney, Australia and [72]Department of Pathology, University of Otago, Christchurch, New Zealand

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

R.H.R., J.H., S.A.G.T. and M.R. conceived and designed the study. R.H.R. analysed data, drafted the figures, and together with S.A.G.T. and M.R. wrote the first draft of the manuscript. J.B. constructed the co-expression networks. M.A.N., IPDGC and SGPD helped with the use of Parkinson's disease GWAS data. All authors contributed to the critical analysis of the manuscript.

ADDITIONAL INFORMATION

REFERENCES

1. Poewe, W. et al. Parkinson disease. *Nat. Rev. Dis. Prim.* **3**, 17013 (2017).
2. Del Tredici, K. & Braak, H. Review: Sporadic Parkinson's disease: development and distribution of α-synuclein pathology. *Neuropathol. Appl. Neurobiol.* **42**, 33–50 (2016).
3. Bendor, J. T., Logan, T. P. & Edwards, R. H. The function of α-synuclein. *Neuron* **79**, 1044–1066 (2013).
4. Spillantini, M. G. et al. Alpha-synuclein in Lewy bodies. *Nature* **388**, 839–840 (1997).
5. Polymeropoulos, M. H. et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045–2047 (1997).
6. Krüger, R. et al. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat. Genet.* **18**, 106–108 (1998).
7. Singleton, A. B. et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841 (2003).
8. Zarranz, J. J. et al. The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann. Neurol.* **55**, 164–173 (2004).
9. Chang, D. et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* (2017). https://doi.org/10.1038/ng.3955
10. Brück, D., Wenning, G. K., Stefanova, N. & Fellner, L. Glia and alpha-synuclein in neurodegeneration: A complex interaction. *Neurobiol. Dis.* **85**, 262–274 (2016).
11. Booth, H. D. E., Hirst, W. D. & Wade-Martins, R. The Role of Astrocyte Dysfunction in Parkinson's Disease Pathogenesis. *Trends Neurosci.* **40**, 358–370 (2017).
12. Yun, S. P. et al. Block of A1 astrocyte conversion by microglia is neuroprotective in models of Parkinson's disease. *Nat. Med.* https://doi.org/10.1038/s41591-018-0051-5 (2018).
13. Gagliano, S. A. et al. Genomics implicates adaptive and innate immunity in Alzheimer's and Parkinson's diseases. *Ann. Clin. Transl. Neurol.* **3**, 924–933 (2016).
14. Nalls, M. A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
15. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
16. Nalls, M. A. et al. Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk. *bioRxiv* (2019). https://doi.org/10.1101/388165
17. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
18. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
19. Oldham, M. C. et al. Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271–1282 (2008).
20. GTEx Consortium et al. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
21. Zhang, Y. et al. Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37–53 (2016).
22. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).
23. Botia, J. A. et al. G2P: Using machine learning to understand and predict genes causing rare neurological disorders. *bioRxiv* (2018). https://doi.org/10.1101/288845
24. Urs, N. M., Peterson, S. M. & Caron, M. G. New Concepts in Dopamine D2Receptor Biased Signaling and Implications for Schizophrenia Therapy. *Biol. Psychiatry* **81**, 78–85 (2017).
25. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, 1–19 (2015).
26. Li, Y. I., Wong, G., Humphrey, J. & Raj, T. Prioritizing Parkinson's Disease genes using population-scale transcriptomic data. *Nat. Commun.* **10**, 994 (2019).
27. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
28. Giambartolomei, C. et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, e1004383 (2014).
29. Hernandez, D. G., Reed, X. & Singleton, A. B. Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. *J. Neurochem.* **139**, 59–74 (2016).
30. Manzoni, C. The LRRK2-macroautophagy axis and its relevance to Parkinson's disease. *Biochem. Soc. Trans.* **45**, 155–162 (2017).
31. Denny, P. et al. Exploring autophagy with Gene Ontology. *Autophagy* **14**, 419–436 (2018).
32. Robak, L. A. et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain* **140**, 3191–3203 (2017).
33. Haelterman, N. A. et al. A mitocentric view of Parkinson's disease. *Annu. Rev. Neurosci.* **37**, 137–159 (2014).
34. Ryan, B. J., Hoek, S., Fon, E. A. & Wade-Martins, R. Mitochondrial dysfunction and mitophagy in Parkinson's: From familial to sporadic disease. *Trends Biochem. Sci.* **40**, 200–210 (2015).
35. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
36. Kahn, R. S. et al. *Schizophr. Nat. Rev. Dis. Prim.* **1**, 15067 (2015).
37. Birnbaum, R. & Weinberger, D. R. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat. Rev. Neurosci.* **18**, 727–740 (2017).
38. Harrison, P. J. Postmortem studies in schizophrenia. *Dialog-. Clin. Neurosci.* **2**, 349–357 (2000).
39. Ilieva, H., Polymenidou, M. & Cleveland, D. W. Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. *J. Cell Biol.* **187**, 761–772 (2009).
40. Eisen, A. & Turner, M. R.Does variation in neurodegenerative disease susceptibility and phenotype reflect cerebral differences at the network level. *Amyotroph. Lateral Scler. Front. Degener.* **14**, 487–493 (2013).
41. Warren, J. D. et al. Molecular nexopathies: A new paradigm of neurodegenerative disease. *Trends Neurosci.* **36**, 561–569 (2013).
42. Do, C. B. et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for parkinson's disease. *PLoS Genet.* **7**, e1002141 (2011).
43. Keller, M. F. et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum. Mol. Genet.* **21**, 4996–5009 (2012).
44. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. *Sci. (80-.).* **360**, eaap8757 (2018).
45. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).
46. Hilker, R. et al. Heritability of schizophrenia and schizophrenia spectrum based on the nationwide Danish twin register. *Biol. Psychiatry* **83**, 492–498 (2018).
47. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
48. Raj, T. et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014).
49. Coetzee, S. G. et al. Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease etiology. *Sci. Rep.* **6**, 1–11 (2016).
50. Noyce, A. J. et al. Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study. *PLoS Med.* **14**, e1002314 (2017).
51. Athauda, D. et al. Exenatide once weekly versus placebo in Parkinson's disease: a randomised, double-blind, placebo-controlled trial. *Lancet (Lond., Engl.)* **390**, 1664–1675 (2017).
52. Hormozdiari, F. et al. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
53. Hormozdiari, F. et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat. Genet.* (2018). https://doi.org/10.1038/s41588-018-0148-2
54. Brynedal, B. et al. Large-scale trans-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation. *Am. J. Hum. Genet.* **100**, 581–591 (2017).
55. Harris, K. D. et al. Classes and continua of hippocampal CA1 inhibitory neurons revealed by single-cell transcriptomics. *PLoS Biol.* **16**, e2006387 (2018).
56. International HapMap 3 Consortium et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
57. 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
58. Pagès, H. SNPlocs.Hsapiens.dbSNP144.GRCh37: SNP locations for Homo sapiens (dbSNP Build 144). R Packag. version 0.99.20 (2017).
59. Sheng, Q. et al. Multi-perspective quality control of Illumina RNA sequencing data analysis. *Brief. Funct. Genom.* **16**, 194–204 (2017).
60. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
61. Botía, J. A. et al. An additional k-means clustering step improves the biological features of {WGCNA} gene co-expression networks. *BMC Syst. Biol.* **11**, 47 (2017).

62. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

63. Cahoy, J. D. et al. A Transcriptome database for astrocytes, neurons, and oligo-dendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).

64. Winden, K. D. et al. The organization of the transcriptional network in specific neuronal classes. *Mol. Syst. Biol.* **5**, 1–18 (2009).

65. Miller, J. A., Horvath, S. & Geschwind, D. H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc. Natl. Acad. Sci.* **107**, 12698–12703 (2010).

66. Ramasamy, A. et al. Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. *Nucleic Acids Res.* **41**, e88–e88 (2013).

67. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module pre-served and reproducible? *PLoS Comput. Biol*. **7**, e1001057 (2011).

68. Brozzi, A., Urbanell, L., Germain, P. L., Magini, A. & Emiliani, C. hLGDB: A database of human lysosomal genes and their regulation. *Database* **2013**, bat024 (2013).

69. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2016).

70. Khan, A. & Mathelier, A. Intervene: A tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinforma.* **18**, 1–8 (2017).

71. International Parkinson Disease Genomics Consortium et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet (Lond., Engl.)* **377**, 641–649 (2011).

72. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* **43**, 969–976 (2011).

73. Ripke, S. et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).

74. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

75. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional Regulatory Elements in the Human Genome. *Annu. Rev. Genom. Hum. Genet.* **7**, 29–59 (2006).

76. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genet*. **6**, e1000888 (2010).

77. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015).

78. Moss, D. J. H. et al. Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurol.* **16**, 701–711 (2017).

79. Skene, N. G. & Grant, S. G. N. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front. Neurosci.* **10**, 1–11 (2016).