BRAIN
ORIGINAL ARTICLE

Establishing gene regulatory networks from Parkinson's disease risk loci

Sophie L. Farrow,[1,2] William Schierding,[1,2] Sreemol Gokuladhas,[1] Evgeniia Golovina,[1] Tayaza Fadason,[1,2] Antony A. Cooper[3,4] and Justin M. O'Sullivan[1,2,3,5,6]

The latest meta-analysis of genome-wide association studies identified 90 independent variants across 78 genomic regions associated with Parkinson's disease, yet the mechanisms by which these variants influence the development of the disease remains largely elusive.

To establish the functional gene regulatory networks associated with Parkinson's disease risk variants, we utilized an approach combining spatial (chromosomal conformation capture) and functional (expression quantitative trait loci) data.

We identified 518 genes subject to regulation by 76 Parkinson's variants across 49 tissues, whicih encompass 36 peripheral and 13 CNS tissues. Notably, one-third of these genes were regulated via *trans*-acting mechanisms (distal; risk locus-gene separated by >1 Mb, or on different chromosomes). Of particular interest is the identification of a novel *trans*-expression quantitative trait loci–gene connection between rs10847864 and *SYNJ1* in the adult brain cortex, highlighting a convergence between familial studies and Parkinson's disease genome-wide association studies loci for *SYNJ1* (*PARK20*) for the first time. Furthermore, we identified 16 neurodevelopment-specific expression quantitative trait loci–gene regulatory connections within the foetal cortex, consistent with hypotheses suggesting a neurodevelopmental involvement in the pathogenesis of Parkinson's disease. Through utilizing Louvain clustering we extracted nine significant and highly intraconnected clusters within the entire gene regulatory network. The nine clusters are enriched for specific biological processes and pathways, some of which have not previously been associated with Parkinson's disease.

Together, our results not only contribute to an overall understanding of the mechanisms and impact of specific combinations of Parkinson's disease variants, but also highlight the potential impact gene regulatory networks may have when elucidating aetiological subtypes of Parkinson's disease.

1 Liggins Institute, The University of Auckland, Auckland, New Zealand
2 The Maurice Wilkins Centre, The University of Auckland, Auckland, New Zealand
3 Australian Parkinson's Mission, Garvan Institute of Medical Research, Sydney, New South Wales, Australia
4 St Vincent's Clinical School, UNSW Sydney, Sydney, New South Wales, Australia
5 Brain Research New Zealand, The University of Auckland, Auckland, New Zealand
6 MRC Lifecourse Epidemiology Unit, University of Southampton, UK

Correspondence to: Professor Justin O'Sullivan
Liggins Institute, University of Auckland
85 Park Road
Auckland 1024, New Zealand
E-mail: justin.osullivan@auckland.ac.nz

Received March 29, 2021. Revised December 02, 2021. Accepted December 20, 2021. Advance access publication January 30, 2022

Correspondence may also be addressed to: Antony A. Cooper
Garvan Institute of Medical Research
384 Victoria Street
Sydney NSW, Australia 2010
E-mail: a.cooper@garvan.org.au

Keywords: Parkinson's disease; genetics; gene regulation; spatial genomics; precision medicine

Abbreviations: eQTL = expression quantitative trait loci; GWAS = genome-wide association study; LOEUF = loss-of-function observed/expected upper bound fraction; PPI = protein–protein interaction; SNP = single nucleotide polymorphism

Introduction

Parkinson's disease is considered to be primarily an idiopathic neurodegenerative disorder, with monogenic forms contributing to just 5–10% of all cases.[1] However, the idiopathic nature of Parkinson's disease is being questioned, as evidence increasingly supports a complex involvement of genetics in the development of the majority of cases.[2,3] Genome-wide association studies (GWAS) have identified >200 Parkinson's disease risk loci,[4] but only 90 Parkinson's disease-associated single nucleotide polymorphisms (Parkinson-SNPs) across 78 risk loci were replicated in the largest meta-analysis to date.[5] As is typically observed with GWAS variants, the majority of the Parkinson-SNPs are located within non-coding regions of the genome, with no direct or obvious influence on protein structure or function.[6,7] Studies have shown that such non-coding disease-associated variants are more likely to be located within regulatory regions[8] and thus contribute to risk through influencing gene regulation and expression, either locally or distally. These regulatory interactions are likely to be tissue-specific, adding a further layer of complexity. Consequently, determining how these variants contribute to Parkinson's disease risk, both individually and in combination, poses a major scientific challenge.[9,10]

Although Parkinson's disease is defined as a neurodegenerative disease, mounting evidence demonstrates the role of non-CNS tissues in the development and presentation of such disorders (i.e. Huntington's disease[11] and Parkinson's disease[12,13]). Both alpha-synuclein protein pathology and modulation of Parkinson's disease-related genes have been identified in peripheral tissues (e.g. the gastrointestinal tract and heart) of patients with Parkinson's disease.[12,14–19] The contribution of peripheral tissue in the origins of Parkinson's disease warrants further research, and thus the consideration of how Parkinson-SNPs mediate risk should not be confined to tissues of the CNS.

Spatial gene regulatory interactions are hypothesized to be drivers of complex trait heritability,[20] acting through both *cis*- (nearby) and *trans*- (distal; locus-gene separated by >1 Mb, or on different chromosomes) mechanisms (Fig. 1).[19,21,22] These *cis*- and *trans*-acting elements can regulate the transcription of one or more genes, in a tissue-specific manner, and are commonly detected in the form of expression quantitative trait loci (eQTL).[23] Genetic variation within elements of gene regulatory networks likely confer risk at different developmental stages, including during foetal neurodevelopment—a critical stage that has a growing body of support in neurodegenerative diseases.[24]

Here we performed correlational analyses of experimentally derived data to identify eQTLs that physically connect Parkinson-SNPs to the genes that they control, in three dimensions, with the goal of understanding the putative functional impacts of known Parkinson-SNPs.[21] The integration of spatial and eQTL data allows for the identification of *trans*-eQTL–gene associations,[25] thereby nominating genes which have not previously been implicated in Parkinson's disease. Our analysis identified 518 genes subject to regulation by 76 Parkinson-SNPs across 49 tissues. Further, clustering analysis of the entire gene network revealed nine significant, intraconnected clusters, enriched for both novel and known Parkinson's disease biological pathways, highlighting putative disease-causative molecular mechanisms and areas for future research.

Materials and methods

Data and reference files

The 90 Parkinson-SNPs (across 78 genomic regions; Supplementary Table 1) investigated in this study were previously identified by a GWAS meta-analysis as being of genome-wide significance ($P < 5 \times 10^{-8}$).[5]

All coordinates presented within this manuscript are according to human reference genome GRCh38 (hg38). The coordinates for the 90 Parkinson-SNPs were converted from hg19 to hg38 using the UCSC LiftOver tool.

Identification of eQTL–gene pairs

The contextualize developmental SNPs using 3D information (CoDeS3D)[21] algorithm was used to identify genes whose transcript levels are putatively regulated by the 90 Parkinson-SNPs. CoDeS3D integrates data on spatial interactions between genomic loci (Hi-C data; Supplementary Table 2) with expression data (genotype–tissue expression database version 8; GTEx v8[26]) to identify genes whose transcript levels are associated with a physical connection to the SNP (i.e. spatial eQTL).

Hi-C captures regions of the genome that are physically interacting and can be covalently connected by a cross-linking agent.[27] The hg38 reference genome was digitally digested with MboI, DpnII and HindIII to obtain all possible Hi-C fragment locations for the 90 Parkinson-SNP loci. All identified SNP fragments (tagged by the Parkinson-SNPs) were then queried against the Hi-C databases (70 different cell lines from 12 studies; Supplementary Table 2) to identify distal fragments of DNA that spatially connect to the SNP loci. Spatial SNP–gene connections are established when the SNP-containing fragment spatially connects to a fragment that overlaps any region between the start and end of a gene as defined by GENCODE. There was no binning or padding around restriction fragments to obtain gene overlap. The resulting spatial SNP–gene pairs were subsequently used to query the GTEx v8 eQTL database[26] to identify spatial SNP–gene pairs with significant eQTLs [both *cis*- and *trans*-acting eQTL; false discovery

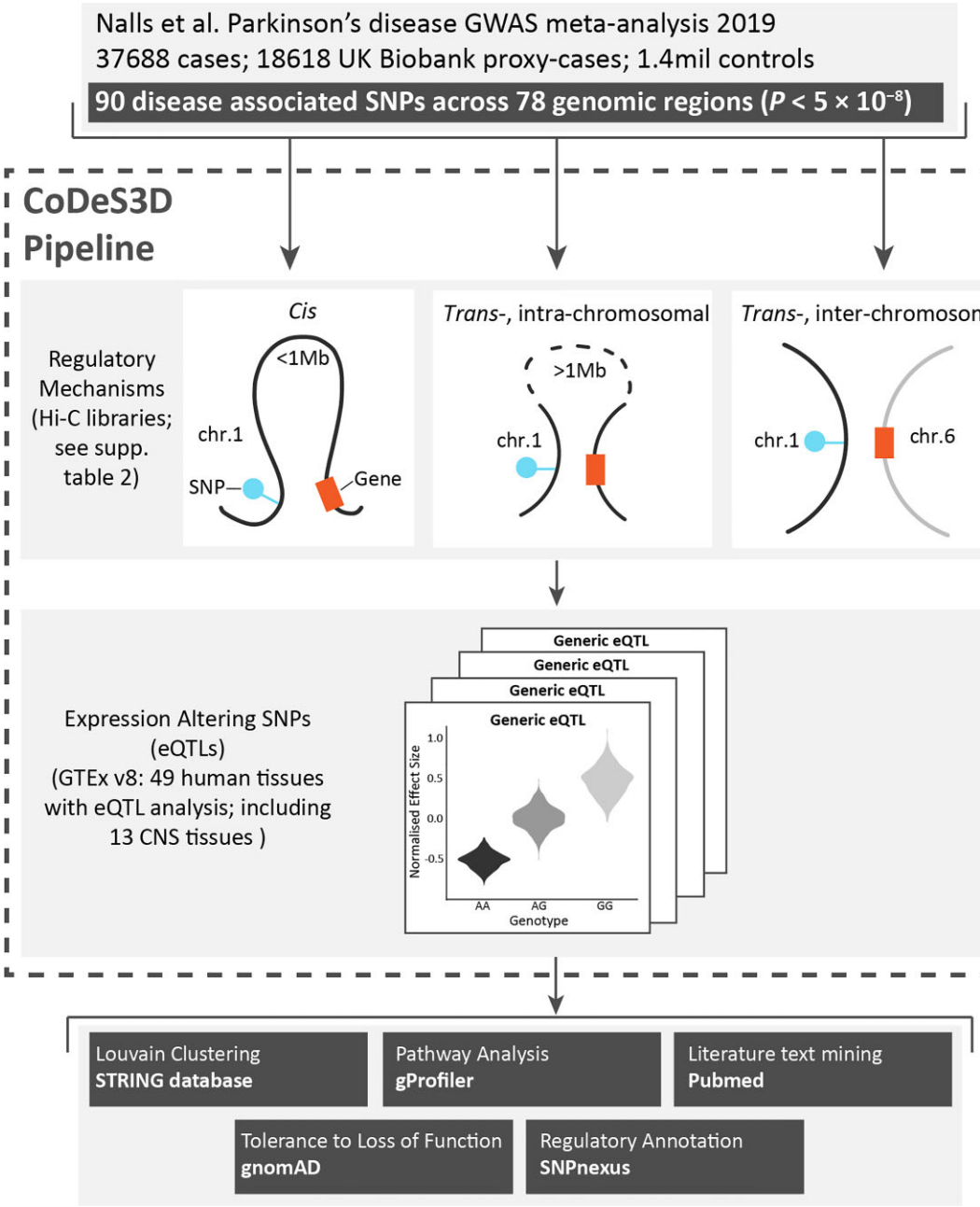

Figure 1 Methods workflow. Ninety Parkinson-SNPs were obtained from Nalls *et al.*[5] Spatial interactions between the 90 Parkinson-SNPs and genes were identified from Hi-C libraries (Supplementary Table 2). The resulting spatial SNP–gene pairs were then used to query GTEx v8 to identify significant eQTLs. The significant spatial SNP–gene pairs were then analysed for functional relevance using multiple tools and databases ('Materials and methods' section). Figure adapted from Schierding *et al.*[19]

rate (FDR) adjusted $P < 0.05$]. Using the same parameters, we also identified spatial eQTL interactions for two sets of null distributions generated from (i) all called variants in the GTEx v8; and (ii) trait-associated SNPs in the GWAS Catalogue (accessed 18 November 2021). Each set contained 1000 simulations of 90 randomly selected SNPs without replication. We then intersected the resulting gene sets with the original 518 Parkinson's disease genes. The proportion overlap for each simulation set was calculated.

We performed a 'brain-specific' analysis by interrogating only the subset of Hi-C libraries derived from brain-specific cell lines (11 cell lines from four studies, highlighted in red in Supplementary Table 2)[28–31] and only expression data from the 13 brain-specific tissues in GTEx v8.

Identification of neurodevelopmental-specific eQTL–gene pairs

We performed a neurodevelopmental stage-specific analysis by interrogating Hi-C libraries from foetal-specific brain cell lines (cortical plate neurons; germinal zone neurons; Supplementary Table 2; datasets 1 and 2) with expression data from a foetal cortex eQTL dataset.[32]

Functional analysis of eQTL SNPs

SNPnexus v4[33] (https://www.snp-nexus.org/v4/; accessed 22 June 2020) was used to obtain known epigenomic annotations for the eQTL SNPs.

Probability of gene loss of function intolerance

The loss-of-function observed/expected upper bound fraction (LOEUF)[34] score for the genes, within the significant SNP–gene pairs, was obtained from gnomAD v2.1.1 (https://gnomad.broadinstitute.org/; accessed 21 July 2020) to determine the level of constraint on the identified genes.

Protein–protein interaction and modularity clustering

STRING[35] (Search Tool for Retrieval of Interaction Genes/proteins; https://string-db.org, accessed 22 July 2020) was queried to identify published information on interactions between genes or their respective proteins. Only protein–protein interactions (PPIs) with a high confidence level (>0.700, as defined in STRING) were used for this analysis, and interactions identified only through text-mining were excluded. A Louvain method was used to determine the syntality of each node, following four different criteria: (i) immediate connection; (ii) shortest path (i.e. the minimum number of edges connecting any two nodes); (iii) node acting as a bridge; (iv) connections that nodes have in common. The proteins were then hierarchically clustered using the Louvain algorithm[36]; clusters were defined as significant if $P < 0.05$.

Pathway analyses

The g:Profiler[37] database was used to identify enriched pathways. Queries were run on: (i) all genes; (ii) the '*cis*'; and (iii) the '*trans*' subsets of genes (i.e. genes regulated only in *cis*, or only in *trans*). To test the specificity of the identified pathways to Parkinson's disease, we compared their occurrence in the previously described 1000 simulations of trait-associated SNPs in the GWAS Catalogue. We calculated a *P*-value for each pathway from the set of Parkinson's genes as the number of its occurrence in the simulations divided by 1000. We also performed a similar bootstrapping on 1000 simulations of 518 random genes from the GENCODE gene reference (v26).

Additional analyses of genes and variants identified by Makarious *et al.*

Makarious *et al.*[38] recently utilized a multimodality approach to identify genetic and transcriptomic features that contribute to risk predictions of Parkinson's disease. They highlighted two SNPs (rs10835060 and rs4238361) and 29 genes. We performed CoDeS3D analysis on the two SNPs, across all Hi-C cell lines and GTEx tissues (as previously described). The resulting eQTL–gene pairs, along with the 29 genes highlighted through transcriptomic analysis, were combined with our set of 523 genes. Louvain clustering and PPI analysis were re-run on this combined list of genes to see how or if the subset of genes co-locate within the networks.

URLs

CoDeS3D pipeline: https://github.com/Genome3d/codes3d-v2
gnomAD: https://gnomad.broadinstitute.org/
gProfiler: https://biit.cs.ut.ee/gprofiler/
UCSC: https://genome.ucsc.edu/index.html
STRING: https://string-db.org/
SNPNexus: https://www.snp-nexus.org/
Louvain clustering analysis: https://github.com/Genome3d/PPI-network-analysis

Data availability

All data generated during this study are included in the supplementary information. Datasets analysed and tools used in this study were all derived from publicly available resources (see 'URLs' section).

Results

Parkinson's disease GWAS SNPs regulate the expression of >500 genes

Nalls *et al.*[5] identified 90 SNPs that were associated with Parkinson's disease at the level of genome-wide significance (Supplementary Table 1), yet the mechanisms by which these variants influence the development of the disease remains largely elusive. We used the CoDeS3D algorithm[21] to identify SNPs that have evidence of physical interaction with the gene as captured by Hi-C (Supplementary Table 2) and also associate with changes in gene expression (hereafter eQTLs) and the genes whose transcript levels were affected (Fig. 1).

Seventy-six (84%) of the 90 Parkinson's disease-risk SNPs were identified as eQTLs associated with the regulation of 518 genes through 542 unique eQTL–gene pairs across the 49 tissues (Table 1, Supplementary Table 3 and Supplementary Fig. 1). The

Table 1 Summary statistics for the spatial eQT–gene regulatory network for the 90 Parkinson-SNPs

	Parkinson's disease SNPs		1000 simulation SNP sets [Mean (min–max)][a]	
	Brain-specific[b]	All tissues[b]	GWAS SNPs (all tissues)	All SNPs (all tissues)
No. SNPs	90	90	90	90
No. eQTL SNPs[c]	55	76	54 (33–74)	7 (0–17)
No. genes[d]	165	518	235 (93–478)	18 (0–135)
No. eQTL–gene pairs[e]	167	542	244 (93–518)	19 (0–136)
No. *trans* eQTL–gene pairs	30	178	37 (13–63)	2 (0–15)

SNPs were downloaded from the Nalls *et al.*[5] GWAS (download date: 18 June 2020). eQTLs (both *cis*- and *trans*-acting eQTL) were only called if: (i) a spatial interaction connecting the SNP and gene had been captured; and (ii) if the adjusted *P*-value for the eQTL association was $P < 0.05$ following a step-wise Bonferroni Hochberg correction for the number of tests that were performed during the eQTL calling. All adjusted *P*-values are presented for all eQTLs (*cis* and *trans*; Supplementary Tables 3 and 4).

[a]Spatial eQTL interactions identified for two sets of null distributions generated from (i) all called variants in the GTEx v8; and (ii) trait-associated SNPs in the GWAS Catalogue (accessed 18 November 2021). Each set contained 1000 simulations of 90 randomly selected SNPs without replication.

[b]For full list of tissue eQTL–gene interactions see Supplementary Tables 3 and 4.

[c]eQTL SNPs were defined as having significant spatial interactions (FDR ≤ 0.05) with at least one gene.

[d]Genes were those whose expression was shown to be affected by an eQTL SNP.

[e]The total number of SNP-gene pairs reflects interactions with FDR ≤ 0.05 in at least one GTEx tissue.

identified gene sets were specifically enriched for the 90 Parkinson's SNPs ($P < 0.001$ based on 1000 bootstrap permutations; Table 1 and Supplementary Fig. 2). The 76 eQTLs were individually associated with the regulatory impacts of as few as one, or as many as 39 genes in *cis* and *trans*. We identified 178 of the 542 genes as being associated with Parkinson's disease through *trans*-eQTL–gene connections. Consistent with previous studies,[39–41] the effect sizes associated with the *trans*-eQTLs are significantly smaller than those associated with *cis*-eQTLs [Supplementary Fig. 3; Kruskal–Wallis test, $H(2) = 82.628$, $P < 2.2 \times 10^{-16}$]. Bootstrapping confirmed that the *trans*-eQTLs at Parkinson's disease-associated SNPs occur more frequently than expected at random. The proportion of *trans*-eQTLs was similar to what was observed for an analysis of autoimmune conditions that was also performed on the latest GTEx release (v8).[42] These results are consistent with an increasing body of research that is noting the tissue-specific nature of *trans*-interactions and their importance for gene regulation.[41,43]

We did not identify eQTL interactions for 14 of the 90 SNPs. Conversely, eight of these 14 SNPs are annotated as being eQTLs in the IPDGC GWAS Locus Browser[44] and a further four are eQTLs in GTEx (Supplementary Table 1). However, in these 12 instances, the eQTLs occur in *cis*- (i.e. within 1 Mb) and are not supported by Hi-C data. The apparent contradictions between the iPDGC and GTEx datasets lie in the underlying eQTL sources. iPDGC harnesses multiple brain-specific and blood-only eQTL datasets, whereas GTEx includes data for 49 CNS and peripheral tissues, providing the most comprehensive eQTL dataset across all tissues. Furthermore, over half of the genes we identified, but which iPDGC has not, involved the inclusion and analysis of *trans*-eQTLs in GTEx. Such analysis was not conducted in the iPDGC browser and thus only *cis*-regulated genes could be identified.

To highlight eQTL associations that may be of particular interest for future studies, we looked to replicate our findings in a second eQTL dataset. We used the eQTLGen dataset[41] for replication. eQTLGen performed *cis*- and *trans*-eQTL analyses using blood-derived expression from 31 684 individuals, and thus we looked to replicate only the CoDeS3D eQTLs that we identified in the whole blood. We found that 71.3% (62 of 87) of the whole blood *cis*-eQTLs identified by CoDeS3D were replicated in the eQTLGen dataset (Supplementary Table 3B). We were unable to test for replication of our *trans*-eQTLs within the eQTLGen dataset for the following reason. Specifically, only 11 of the Nalls *et al.*[5] SNPs (rs10513789, rs10797576, rs11158026, rs117896735, rs12456492, rs147045, rs34311866, rs356182, rs35749011, rs76904798, rs823118) are present within the curated Parkinson's SNPs that were tested in eQTLGen (Supplementary Table 3C; note that rs34778348 was filtered out of the eQTLGen list prior to their analysis[41]). Of the 11 SNPs within the overlapping set, CoDeS3D only identifies rs1474055 as a whole blood *trans*-acting eQTL targeting *TLK1*, a distance of 2 736 939 bp. Unfortunately, this 2.7 Mb distance is below the minimum intra-chromosomal distance of >5 Mb set by eQTLGen.[41] Therefore, we were unable to test for any CoDeS3D *trans*-eQTL validations. We note that eQTLGen did identify rs35749011 as a whole blood *trans*-eQTL targeting *HIST1H3H*, *HIST1H2BH* and *HIST1H2BD*. However, in our analyses CoDeS3D did not identify any spatial interactions for this SNP and thus it was filtered out.

Consistent with observations for SNPs associated with other traits,[45] at least one *trans*-regulatory interaction was identified for 81.6% (62 of 76) of the eQTLs. Moreover, 92.7% (165 of 178) of these *trans*-eQTL–gene interactions were identified in only one tissue. By contrast, the *cis*-interactions were identified in eight tissues on average (range of 1 to 49 tissues). Of the eQTLs, 11.8% (9 of 76; Supplementary Table 1) were exclusively involved in *trans*-regulatory interactions. *Trans*-eQTL interactions regulated 18.1% of the genes identified in the brain (30 of 166; Supplementary Table 4) and 32.8% of the genes among all 49 tissues (178 of 518; Table 2 and Supplementary Table 3). Collectively, these results highlight the importance of looking beyond the nearest gene to identify the regulatory effects of disease-associated variants.

Table 2 Proportion of genes subject to *cis*- and *trans*-regulation

	Genes subject to *cis*- or *trans*-regulation	
	Brain-specific[a]	All tissues[b]
Cis eQTL–gene pairs	136 (82.0%)	364 (67.2%)
Trans-intrachromosomal eQTL–gene pairs	10 (6.0%)	56 (10.3%)
Trans-interchromosomal eQTL–gene pairs	20 (12.0%)	122 (22.5%)

The proportion of eQTL-gene pairs that are either *cis*-, *trans*-intrachromosomal or *trans*-interchromosomal in 13 GTEx brain-specific tissues[a] and all 49 GTEx tissues[b]. Brain-specific indicates the eQTL dataset obtained through analysing Hi-C cell lines only from the brain and eQTLs only from the brain tissues in GTEx. All-tissues indicates the eQTL dataset obtained through analysing all Hi-C cell lines and eQTLs from all tissues in GTEx. There is a significant difference (chi square test P-value < 0.01) between brain tissues and all tissues for the proportions of the *cis* versus *trans* eQTLs.
For detailed information on the specific eQTL–gene pairs see Supplementary Tables 3 and 4.

We reasoned that SNPs that are involved in eQTLs likely mark enhancer or promoter sites.[46] We queried SNPnexus[33] to identify those eQTLs that were marked by histone modifications or fell within open chromatin regions, as indicated by DNAse accessibility. Consistent with our hypothesis, 91% (69 of 76) of the SNPs were marked by histone modifications associated with either enhancers (58) and/or promoters (27). Of the SNPs, 27.6% were within accessible chromatin (Supplementary Table 5). Collectively, these results are consistent with the hypothesis that the loci marked by these eQTLs may be involved in the regulation of gene expression.

Pathway analysis was conducted on the complete set of 518 genes that were impacted by the eQTLs (Supplementary Table 6). g:Profiler[37] identified significant (adjusted $P < 0.05$) enrichment within 10 known biological pathways (g:GOSt), including organelle organization, synaptic vesicle recycling and endocytosis. Bootstrapping analyses (see 'Materials and methods' section) identified which of the pathways were specifically enriched for Parkinson's disease (Supplementary Table 6, yellow highlight). Of note, enriched pathways that are driven predominantly by genes within the human leucocyte antigen (HLA)-region, such as antigen processing and presentation, are not exclusive to Parkinson's disease and were identified in ~17% of the pathway bootstrapping simulations. This does not affect the relevance of these pathways, but rather indicates the importance of these pathways across multiple cellular functions and responses.

Brain-specific regulatory impact of Parkinson-SNPs

Fifty-five (61%) of the 90 Parkinson-SNPs were identified as eQTLs associated with brain-specific regulation of 165 genes through 169 unique eQTL–gene pairs across the 13 GTEx brain tissues (Table 1 and Supplementary Table 4; refer to the 'Materials and methods' section for details). Notably, g:Profiler[37] analysis only identified regulation of neuron death as the one significantly

enriched biological pathway within the GTEx brain tissues (Supplementary Table 6B). This enrichment is consistent with neuronal cell death being one of the primary pathological characteristics seen in Parkinson's disease.[47] We propose that the identified eQTLs associated with this pathway are likely contributing to the dysregulation of neuron death seen in Parkinson's disease.

The regulatory impact of Parkinson-SNPs extends beyond the CNS

Although Parkinson's disease is considered a degenerative disease of the brain, it has become apparent that dysfunction and or alpha-synuclein pathology is observed in non-CNS tissues of Parkinson's disease patients.[13–15] Our spatial eQTL analysis included an assessment of the tissue distribution of the effects of the identified eQTLs within 13 CNS and 36 peripheral tissues. We identified peripheral tissue-specific eQTLs for 28% of the Parkinson-SNPs (21 of 76). Only 2 of 76 Parkinson-SNPs (i.e. rs10756907–*SH3GL2*, brain cortex; rs873786–*SLC26A1*, brain cerebellum) had eQTLs that impacted gene expression levels exclusively in the brain. This supports a possible role for peripheral tissues in Parkinson's disease risk (Supplementary Fig. 5).

The ability to detect eQTLs in specific tissues is known to correlate with tissue sample size within GTEx.[26] Consistent with this, we identified highly significant correlations between tissue sample numbers and (i) all-eQTLs in the brain tissues (Fig. 2A; identified using brain-specific Hi-C and eQTL data; Supplementary Table 4); or (ii) all tissues (i.e. the 49 tissues included within GTEx; Fig. 2E). These highly significant correlations remained when analysing the *cis*-eQTL subsets in the brain ($R = 0.93$, $P = 3.6 \times 10^{-6}$; Fig. 2B) and all tissues ($R = 0.9$, $P < 2.2 \times 10^{-16}$; Fig. 2F). Similarly, the correlation was evident for *trans*-intrachromosomal eQTLs detected in all tissues ($R = 0.67$, $P < 1.8 \times 10^{-6}$; Fig. 2G). By contrast, there was no observable correlation between the number of *trans*-interchromosomal interactions and tissue sample number (Fig. 2D and H). The substantia nigra and brain cerebellar hemisphere exhibited more *trans*-interchromosomal-eQTLs (Fig. 2D), while the thyroid exhibited more eQTLs than expected across all three categories (Fig. 2E–H).

Genes subject to *trans*-regulation by Parkinson-SNPs are more likely to be intolerant to loss-of-function mutations

Genes that are intolerant to inactivation by loss-of-function variants are deemed essential for healthy development.[48] Intolerance to loss of function variants leaves changes to regulation as one of the few mechanisms that can be modified to introduce variation at a population level. The 117 *trans*-interchromosomal-eQTL regulated genes were significantly ($P < 0.01$, Kruskal–Wallis test) more intolerant to loss-of-function mutations [LOEUF 0.42 (median); a low LOEUF score is indicative of evolutionary constraint] than those regulated by *cis*- or *trans*-intrachromosomal acting eQTLs [LOEUF 0.83 and 0.85, respectively (median); Fig. 3, Supplementary Fig. 4 and Supplementary Table 7]. This result is consistent with earlier observations that *trans*-eQTLs are enriched in regulating constrained genes with low LOEUF scores.[19]

Parkinson's disease GWAS SNPs regulate expression of a subset of genes within the foetal cortex

Emerging evidence suggests Parkinson's disease has a neurodevelopmental aspect,[49] similar to recent observations in Huntington's disease.[24] Therefore, we analysed the regulatory impacts of the Parkinson-SNPs using foetal cell line Hi-C (i.e. cortical plate neurons and germinal zone neurons)[29] and foetal cortex eQTL datasets[32] (Supplementary Table 8). Thirty-three genes were found to be regulated by 22 Parkinson-SNPs in the foetal cortex. Of these, 16 genes were regulated by eQTLs involving Parkinson-SNPs in the foetal cortex, without evidence of any eQTLs in adult brain tissues (Fig. 4 and Supplementary Table 4). Ten genes were affected by eQTLs involving Parkinson-SNPs in both the foetal and adult cortex, with effect sizes that were similar in both (Fig. 4). Finally, seven genes were regulated by *cis*-eQTLs in the foetal cortex and adult non-cortical brain tissues (Fig. 4). These findings are consistent with the hypothesis that development stage-specific eQTL patterns impact on disease-relevant mechanisms and thus may contribute to the proposed temporal phases of Parkinson's disease pathogenesis.[50]

Louvain clustering highlights nine intraconnected protein clusters, enriched for disease-relevant, biological pathways

Network representations of complex datasets can aid the identification of biological relationships that are often not identified by enrichment analyses.[51] We used a Louvain clustering algorithm to identify clusters of interacting genes and proteins from within a PPI network generated from the 523 eQTL regulated genes (518 adult tissue eQTLs and the five unique foetal cortex eQTLs; Supplementary Table 9). Nine significant ($P < 0.05$) clusters consisting of 122 genes were identified within the high-confidence PPI network (Fig. 5). The genes within each cluster were regulated by between five and 18 Parkinson-SNPs (Supplementary Table 9) and every cluster contained at least two genes that were co-regulated by a single SNP (Supplementary Fig. 6). Notably, genes that were subject to *trans*-acting eQTLs were central to the definition and identification of several clusters (Fig. 5). Pathway analysis (g:Profiler; $P < 0.05$) of the genes within the individual clusters revealed enrichment in categories that included immunological surveillance (cluster 7), synaptic vesicle recycling (cluster 5) and microtubule polymerization (cluster 3; Supplementary Table 10).

We analysed the nine clusters to identify clusters that were significantly associated with increased numbers of risk or protective SNPs (determined by GWAS-defined odds ratio). Analysis according to the proportions of risk:protective SNPs found that no clusters significantly differed from the expected proportion (Supplementary Table 9B). However, in many of the clusters, a single SNP may have multiple impacts, meaning a protective SNP can act as an eQTL for several genes. Therefore, we analysed the proportions of genes that were affected by risk or protective SNPs. This analysis highlighted a notable shift for cluster 7 (binomial test, Bonferroni adjusted $P < 0.01$; risk proportion = 0.105), consistent with the idea that the immune pathways associated with cluster 7 are protective against Parkinson's disease.

Makarious *et al.*[38] recently used a multimodal machine learning approach, incorporating multi-omics datasets, to inform and improve predictions of Parkinson's disease. Beyond the 90 GWAS SNP signals (which collectively were the top genetic feature), they also identified rs10835060 and rs4238361 as two SNPs that impact on Parkinson's disease biology. CoDeS3D analysis identified eQTLs for both rs10835060 (*KRTAP5-AS1*, *CCDC88A*, *KRTAP5-5*, *BRSK2*) and rs4238361 (*USP47*; and *RP11-507B12.2* and *RP11-259A24.1*; Supplementary Table 11). Notably, BRSK2 co-locates with cluster 2 through an established interaction with the Tau-encoding *MAPT* gene.[52] The model also highlighted 29 genes through transcriptomic

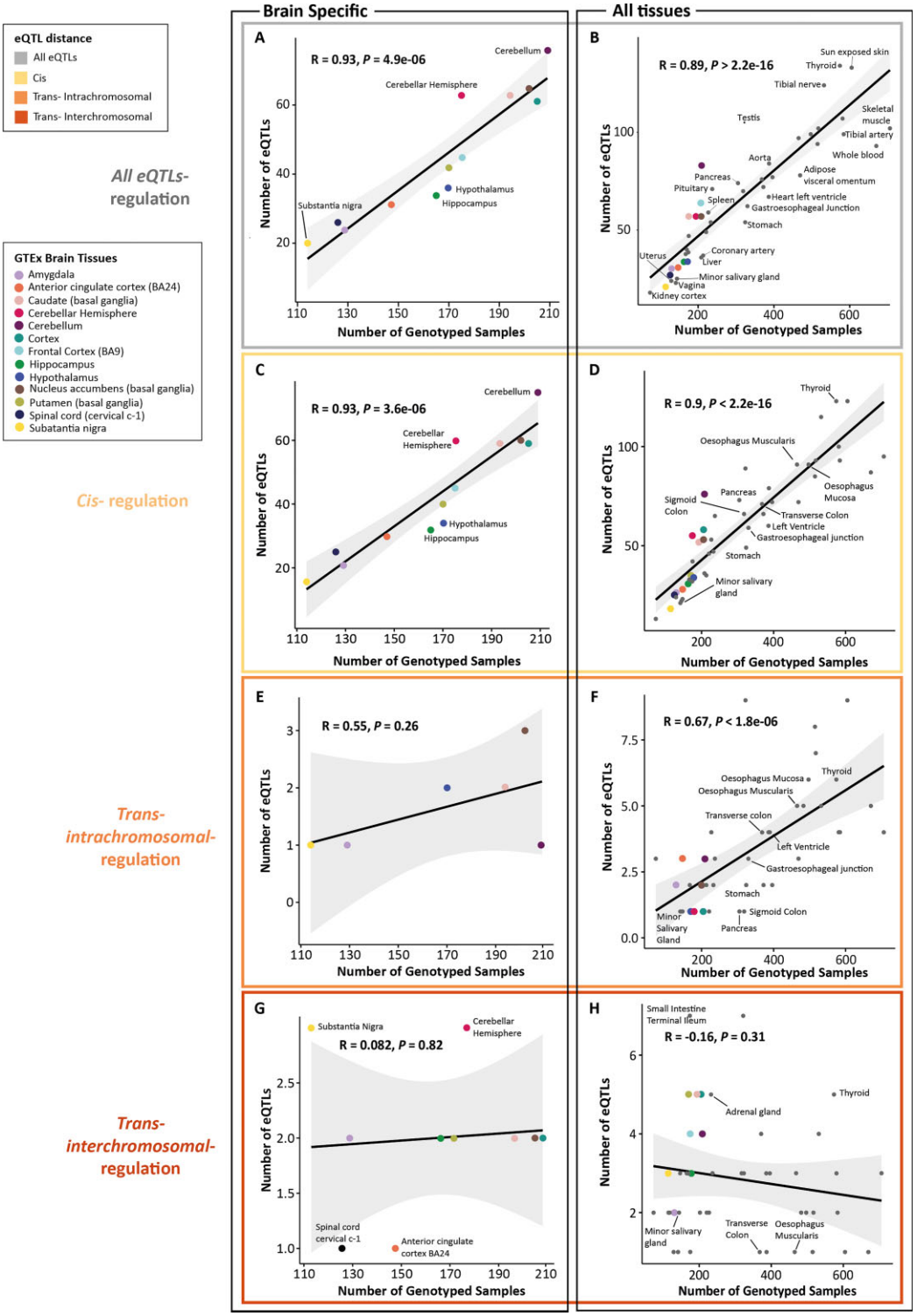

Figure 2 Correlation between genotype samples per tissue and number of eQTLs present in the tissue. (**A**) Correlation between the number of genotyped samples per tissue (in GTEx) and the number of eQTLs (including *cis*, *trans*-intrachromosomal and *trans*-interchromosomal) per tissue, in 13 brain-specific tissues. (**B**) Correlation between the number of genotyped samples per tissue (in GTEx) and the number of eQTLs (including *cis*, *trans*-intrachromosomal and *trans*-interchromosomal) per tissue, in all 49 tissues. (**C**) Correlation between the number of genotyped samples per tissue (in GTEx) and the number of *cis*-eQTLs per tissue, in 13 brain-specific tissues. (**D**) Correlation between the number of genotyped samples per tissue (in GTEx) and the number of *cis*-eQTLs per tissue, in all 49 tissues. (**E**) Correlation between the number of genotyped samples per tissue (in GTEx) and the number of *trans*-intrachromosomal-eQTLs per tissue, in 13 brain-specific tissues. (**F**) Correlation between the number of genotyped samples per tissue (in GTEx) and the number of *trans*-intrachromosomal-eQTLs per tissue, in all 49 tissues. (**G**) Correlation between the number of genotyped samples per tissue (in GTEx) and the number of *trans*-interchromosomal-eQTLs per tissue, in 13 brain-specific tissues. (**H**) Correlation between the number of genotyped samples per tissue (in GTEx) and the number of *trans*-interchromosomal-eQTLs per tissue, in all 49 tissues. The tissues that fall furthest from the confidence interval are annotated. The grey dots show the correlation for all GTEx tissues. The 13 brain tissues (from GTEx) are indicated by the coloured dots, as shown in the legend. For information on all tissues outside of the 95% confidence interval, see Supplementary Table 12.

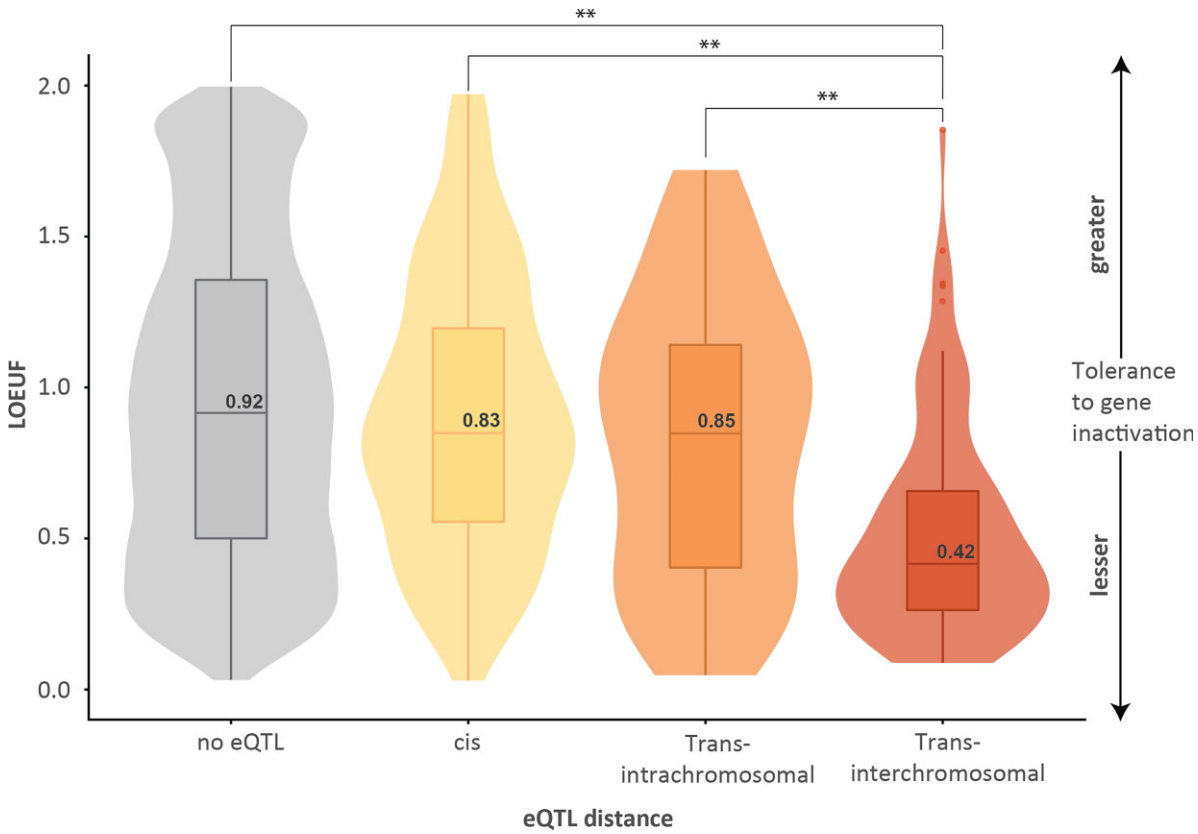

Figure 3 Genes subjected to *trans*-regulation by Parkinson-SNPs are enriched for loss-of-function intolerance. Genes that are loss-of-function intolerant, as measured by a continuous LOEUF score, are enriched in *trans*-regulatory interactions involving Parkinson-SNPs. The LOEUF score is a continuous value that indicates the tolerance of a given gene to inactivation. Low LOEUF scores indicate stronger selection against loss-of-function variation. The distribution is shown as a violin plot with the median (LOEUF) values for each eQTL group (black text). The groups were compared using a Kruskal–Wallis test (**P-value < 0.01); the absence of a significance value indicates the LOEUF values of the two groups were not significantly different. No eQTL = all genes in gnomAD with an assigned probability of being loss-of-function intolerant (pLI) or LOEUF for which an eQTL was not identified in this study (~18 500 genes). Not all genes had LOEUF scores (Supplementary Fig. 2 and Supplementary Table 7).

analysis. Three of these genes (*MMP9*, *TRIM4* and *SYS1*) integrate into clusters 1, 4 and 5, respectively (Supplementary Fig. 7). The co-location of genes and eQTLs, identified as being important for Parkinson's disease diagnosis,[38] within the nine clusters supports the potential importance of the gene–gene interactions and enriched pathways in Parkinson's disease.

Four hundred and two genes did not segregate in the nine clusters. Of note, 211 of the 402 genes (52.5%) had not previously been associated with Parkinson's disease GWAS loci (iPDGC Parkinson's disease browser).[44] Of the 211 genes, 123 were regulated by *trans*-eQTL–gene connections. Notably, iPDGC identified five genes (*DNAI1*, *EYA4*, *LYVE1*, *MYO5B*, *PDZRN4*) as being regulated by *cis*-eQTLs, yet these five genes were exclusively identified through *trans*-eQTL regulatory connections in our analysis.

Discussion

Assigning functionality to Parkinson-SNPs is a critical step towards determining how they contribute to the risk of Parkinson's disease development. In this study, we identified 518 genes whose expression was regulated in *cis* or *trans* by Parkinson-SNPs and the tissues where this regulation occurs. We also demonstrated that 22 Parkinson-SNPs impact the regulation of a subset of 16 genes solely in the foetal cortex and a further 10 genes in both the foetal and adult cortex. Of all 523 identified genes, a subset of 122 *cis*- and *trans*-regulated genes formed nine clusters within a PPIN that were enriched for specific biological pathways, some of which have not been previously associated with Parkinson's disease. Our findings support the hypothesis that both *cis*- and *trans*-dysregulation of gene expression contributes to the risk of Parkinson's disease and provide insight into possible disease-causing mechanisms.

Of note, the effect sizes associated with *trans*-eQTLs are generally smaller than those associated with *cis*-eQTLs, consistent with previous observations.[39–41] Our findings are consistent with predictions from the omnigenic model,[20] which suggests that the weak effects of *trans*-eQTLs significantly contribute to overall heritability of a complex disease. Further, the effects of the *trans*-eQTLs identified in our data converge on a smaller set of trait-relevant pathways, which are central to the regulatory networks underlying these pathways. Despite the relatively weak effects of the *trans*-eQTLs we identified, there are a couple of key examples that deserve further discussion.

SYNJ1 (PARK20) encodes synaptojanin-1, a presynaptic phosphoinositide phosphatase that dephosphorylates PI(4,5)P_2 to trigger the removal of the clathrin coat during synaptic vesicle recycling.[53] *SYNJ1* is a highly constrained gene (LOEUF score = 0.33) and rare missense mutations in *SYNJ1* have been linked to early-onset parkinsonism.[54] Despite *SYNJ1* being acknowledged as a high-

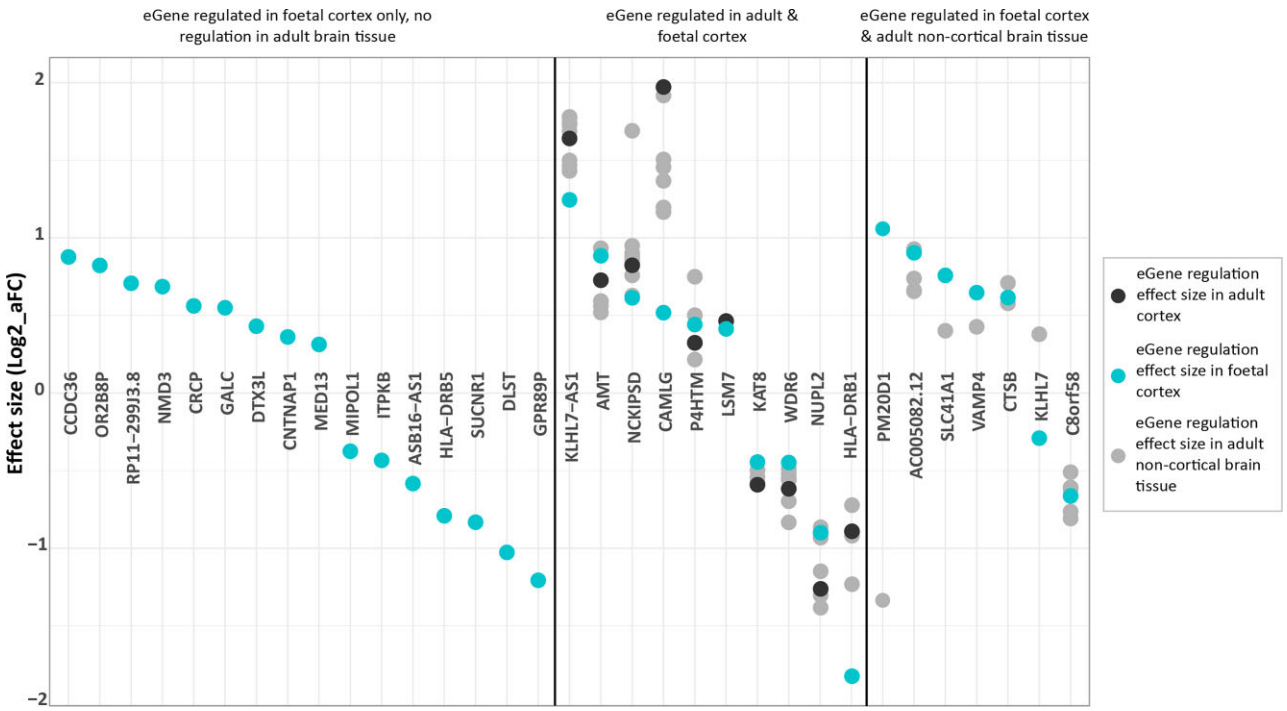

Figure 4 Gene regulation in the foetal cortex compared to the adult cortex. The *leftmost* section shows genes that are regulated only in the foetal cortex, with no eQTLs seen in any of the 13 adult brain tissues. The *middle* section shows genes that are regulated in both the foetal and adult cortex. The black dots show the regulation effect size of the gene in the adult cortex, and the grey dots show the regulation effect size of the gene across the different brain tissues (where an eQTL is seen). The *rightmost* section shows genes that are regulated in both the foetal cortex and adult non-cortical brain tissue.

confidence Parkinson's disease gene,[3,55] GWASs have not identified any SNPs proximal to *SYNJ1* as being significantly associated with Parkinson's disease, nor have they attributed any significant Parkinson-SNP (near or far) to *SYNJ1*. Critically, we identified that the Parkinson's disease-associated SNP rs10847864 acts as a *trans*-acting eQTL for *SYNJ1* expression. rs10847864 is intronic to, and also acts as a *cis*-eQTL with, *HIP1R*, another gene that is involved in clathrin-mediated endocytosis.[56–58] Our discovery of the *trans*-eQTL–*SYNJ1* connection merges observations from population level (i.e. GWAS) and familial studies, reinforcing the potential importance of *SYNJ1* in Parkinson's disease. Such convergence and pleomorphism has previously been reported for other Parkinson's disease genes such as *SNCA* (*PARK1*) and *LRRK2* (*PARK8*), with both genes being identified as risk genes through population-level and familial studies.[59]

Our analysis identified *trans*-eQTLs for approximately two-thirds of the known Parkinson-SNPs. Of note, *RAI14* (retinoic acid induced 14) is regulated by two *trans*-eQTLs, involving two independent Parkinson-SNPs (rs2251086, chr15 and rs55818311, chr19). Although not yet directly linked to Parkinson's disease, *RAI14* (and its encoded protein ankycorbin) has been shown to play a role in the inflammatory response in glial cells[60] and in the establishment of neuronal morphology,[61] both of which are pathways of known importance in Parkinson's disease pathogenesis.[62] Retinoic acid, a regulator of *RAI14* (one of multiple roles of retinoic acid), is being explored as a potential therapeutic target for Parkinson's disease.[63]

Our results provide support for the role of peripheral tissues in Parkinson's disease, notably the oesophagus and thyroid. First, the oesophagus is enriched for *cis* and *trans* regulatory eQTLs. rs76904798 [Parkinson's disease odds ratio (OR) = 1.155] is an eQTL that upregulates *LRRK2* expression in 19 peripheral tissues, including in the oesophagus. Notably, this *cis*-eQTL with *LRRK2* is not identified in any CNS tissues. Second, we identified the thyroid tissue as being enriched for eQTLs, many of which were not represented in CNS tissues. The thyroid is a component of the dopaminergic system and hypothalamic–pituitary–thyroid axis network.[64] A potential link between thyroid hormone disorders, Parkinson's disease risk and symptom severity has been suggested.[64] Specifically, one study identified patients with hypothyroidism to have a twofold elevated risk of developing Parkinson's disease.[65] Collectively, these findings support the growing body of evidence for the importance of the oesophagus[66,67] and thyroid[64] in Parkinson's disease.

We hypothesized that genes regulated by Parkinson-SNPs in foetal cortical tissue may contribute to potential neurodevelopmental aspects of Parkinson's disease.[24,68,69] Sixteen genes were regulated by Parkinson-SNP eQTLs within the foetal cortex. Two of these genes, *CNTNAP1* and *GALC*, are particularly notable. *CNTNAP1* encodes Caspr1, a Neurexin family membrane protein. Reductions in Caspr1 concentrations delay cortical neuron and astrocyte formation in the mouse developing cerebral cortex.[70] *GALC* encodes a lysosomal galactosylceramidase that ensures normal turnover of myelin[71] and has been linked to neuronal vulnerability.[72] While connections between the remaining 14 genes and Parkinson's disease development are less clear, we speculate that SNP-mediated regulation of these genes specific to the foetal cortex may contribute to early neurodevelopmental disturbances that render an individual more susceptible to Parkinson's disease.

Our analyses identified nine clusters that are enriched for specific biological processes and pathways, most of which have been previously associated with Parkinson's disease to some degree (e.g. synaptic vesicle recycling,[73] microtubule polymerization[74]

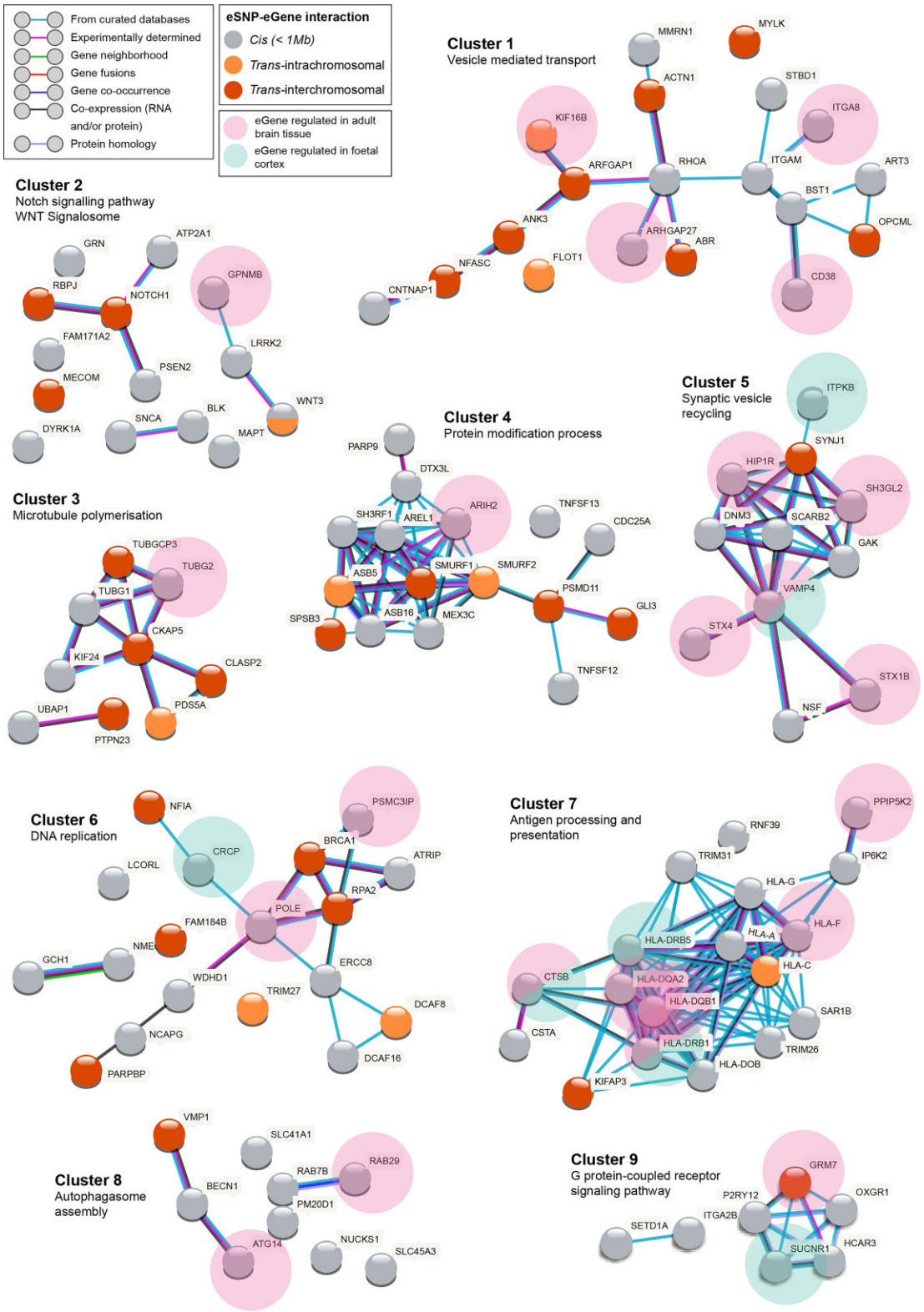

Figure 5 Louvain clustering analysis highlights nine significant clusters, indicative of biological connectivity. The grey and orange shading of the nodes is indicative of whether the gene is subject to regulation via *cis*- or *trans*-mechanisms. The pink- and turquoise-shaded circles indicate genes that are regulated in adult brain tissue and foetal cortex, respectively. The clusters were also analysed in STRING with an increased stringency [only PPIs with a high confidence level (>0.700, as defined in STRING) were used for this analysis, and interactions identified only through text-mining were excluded]. This exclusion led to very few changes, with cluster 6 the only cluster to lose any connectivity within the cluster (WDHD1, NCAPG and PARPBP no longer connect). Experimentally determined = imported from experimental repositories; gene neighbourhood = similar genomic context in different species suggest a similar function of the proteins; gene fusions = fused proteins are recognized by orthology of the fused parts to other, non-fused proteins; gene co-occurrence = indicates the presence of a specific gene pair is in agreement in all species—must be expressed together; co-expression = predicted association between genes/proteins based on RNA and/or protein expression.

and autophagosome assembly[75]; Supplementary Table 10B). Dysregulated expression of the components within these pathways is potentially the basis of the risk conferred by the Parkinson-SNPs. The clusters aid in understanding how Parkinson's disease SNPs mechanistically contribute to disease risk, and some interesting points can be drawn from these. For example, genes within cluster 6 are enriched for functions in DNA replication and repair, a pathway previously associated with the development of other neurodegenerative diseases.[76] Notably, *BRCA1* and *RPA2* (both previously linked to DNA damage response and repair)[77,78] are regulated in *trans* by Parkinson-SNPs (rs11950533; rs9568188; rs62053943) and are central to cluster 6. It is notable that cluster 6 contains several factors associated with PARP1 activity [e.g. the PARP1 binding protein (PARPBP) and BRCA1] that link this cluster to the repair of single stranded breaks, which are enriched at neuronal enhancers in post-mitotic neurons.[79] A further example is the regulation of autophagy initiation by Parkinson-SNPs, highlighted in cluster 8. Three interacting proteins within cluster 8, encoded by *VMP1*, *BECN1* and *ATG14*, are each regulated by a different Parkinson-SNP (rs12951632 chr17; rs11158026 chr14; rs10748818 chr10; Supplementary Table 3), including a *trans*-eQTL connection regulating *VMP1*. rs12951632 (OR = 0.93) and rs11158026 (OR = 0.919) are protective Parkinson-SNPs that, respectively, increase *BECN1* and *ATG14* expression, two interacting core components of the PI3–kinase complex, required for autophagosome formation.[80] Individuals with both of these variants would potentially have increased autophagic capacity relative to individuals with one variant.

The genes in cluster 7 are strongly enriched for antigen processing and presentation, which is increasingly being implicated in the progression of Parkinson's disease.[81] Both rs504594 and rs9261484 are associated with a reduced risk of developing Parkinson's disease (OR = 0.8457 and OR = 0.9385, respectively). We identified a spatial eQTL between rs504594 and *HLA-DRB1* in both the foetal cortex and adult brain (including cortex and substantia nigra), implicating this regulatory eQTL–gene connection in both the neurodevelopment and neurodegenerative stages of Parkinson's disease. Interestingly, rs504594 (previous ID: rs112485776) was recently validated in a SNP-level meta-analysis (OR = 0.87), with results displaying no residual HLA effect in adults after adjusting for the SNP.[82] Instead, three amino acid polymorphisms within the *HLA-DRB1* gene were identified as drivers of the association between the HLA region and Parkinson's disease risk.[82] We agree that the impacts of rs504594 are contingent upon the *HLA-DRB1* allele—both in terms of regulation and protein sequence. We contend that the effects of the rs504594 eQTL are developmental. Future studies must untangle these developmental effects and identify the neurodevelopmental stages that may prime certain individuals to be more vulnerable to later triggering mechanisms. We note that such interpretation should be taken with caution given the highly polymorphic nature of the HLA-region.

The 90 SNPs used in our analyses explain 16–36% of the heritable risk of Parkinson's disease and contribute to a predictive polygenic risk score for Parkinson's disease [area under curve AUC) = 0.651; 95% confidence interval (CI) 0.617–0.684].[5] Nalls *et al.*[5] also conducted polygenic risk score analyses using a set of 1805 SNPs, identified from the NeuroX-dbGAP and Harvard Biomarker Study cohorts, using a *P*-value threshold for inclusion of 1.35×10^{-3}. Notably, the polygenic risk score calculation using 1805 versus 90 SNPs only improved the AUC to 0.692 (95% CI 0.660–0.725), and thus we did not include these additional SNPs in our analysis. However, statistical significance is simply a measure of confidence in the signal:noise ratio and does not prove function or causality. Therefore, because GWAS 'tag' SNPs are typically part of a larger linkage disequilibrium blocks, it remains possible that some of the 90 tag-SNPs are not the causal variant, but rather they are in strong linkage disequilibrium (i.e. co-inherited) with the causal variant. This means, that as noted by Nalls *et al.*,[5] there are thousands of SNPs that correlate with an increased risk of Parkinson's disease. Despite considerable progress in the development of methods to prioritize functional SNPs (e.g. using chromatin accessibility[83] or epigenetic marks[84,85]) there is still no definitive predictive method for distinguishing the causal variant. Thus, we used the 90 independent SNPs that reached genome-wide significance level as the proxy for determining larger associated risk gene networks.

We identified spatial eQTLs, under the premise that regulation occurs due to the physical connection that was captured by Hi-C. Additional *trans*-eQTLs may occur through indirect mediation as a result of *cis*-effects on a target gene (e.g. transcription factor).[39,86] While we did not observe this directly, we did not perform an in-depth exploration of eQTLs that might occur through this indirect mediation. Future research that incorporates the potential of indirect mediation may identify additional important Parkinson's disease-associated eQTL relationships.

We acknowledge several limitations to our analysis. First, the Hi-C cohort and GTEx libraries were generated from unrelated samples that were not age- or gender-matched. Second, the sampled donors in GTEx are predominantly of European descent, limiting the significance of our findings to this ethnicity. However, the GWAS cohort also used individuals of European ancestry, meaning for this analysis the datasets were congruent. Third, our eQTL analysis assumes that mRNA concentration correlates directly with protein levels. While it is true that protein levels are to some extent determined by their mRNA concentration, there are many post-transcriptional processes that can lead to a deviation from the expected correlation.[87] The fourth limitation is that eQTL data represent composite datasets across developmental periods (e.g. foetal samples were aged from 14 to 21 weeks post-conception and the adult samples were from individuals aged 21–70 years). Finally, the approach that we have taken throughout this study is correlational, and thus causality cannot be inferred. Future work may look at the utility of additional methods, such as Mendelian randomization, to affirm causative associations. Despite such limitations, the identification of *trans*-eQTLs is a particular strength of our methodology, relying on captured contacts within the genome organization to reduce the impact of multiple testing correction. As such, we contend that these limitations do not invalidate the significance of our findings of *trans*-acting eQTLs and the genes that they impact.

Conclusions

Understanding the functional impact of Parkinson-SNPs is critical to our understanding of how these variants contribute to the development and clinical presentation of Parkinson's disease. Our functional interpretation of Parkinson-SNPs integrates individual loci into a gene regulatory network, which includes genes with and without prior Parkinson's disease associations. The regulatory network includes clusters, and within them genes, that are enriched for biological functions that have known, putative or previously unknown roles in Parkinson's disease. Development-specific changes to this network (within the foetal cortex) are suggestive of roles for neurodevelopmental changes being early contributors to Parkinson's disease risk. Similarly, enrichments for regulatory

changes within peripheral tissues may indicate a greater role for these tissues in Parkinson's disease than is currently appreciated. Collectively, our findings not only contribute to an overall understanding of the multiple biological pathways associated with Parkinson's disease risk loci, but also highlight the potential utility of gene regulatory networks when considering aetiological subtypes of Parkinson's disease.

Funding

S.L.F., W.S., T.F., A.A.C. and J.M.O. were funded by the Michael J Fox Foundation for Parkinson's research and the Silverstein Foundation for Parkinson's with GBA—grant ID 16229 to J.M.O., S.L.F. and T.F. were funded by the Dines Family Charitable Trust. A.A.C. received grant funding from the Australian government. S.G. was funded by an MBIE Catalyst Grant (The New Zealand–Australia LifeCourse Collaboration on Genes, Environment, Nutrition and Obesity (GENO); UOAX1611). E.G. was funded by University of Auckland Doctoral Scholarship.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Bandres-Ciga S, Diez-Fairen M, Kim JJ, Singleton AB. Genetics of Parkinson's disease: An introspection of its journey towards precision medicine. *Neurobiol Dis*. 2020;137:104782.
2. Billingsley KJ, Barbosa IA, Bandrés-Ciga S, *et al*. Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset. *NPJ Parkinsons Dis*. 2019;5(1):8.
3. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol*. 2020;19:170–178.
4. Buniello A, Macarthur JAL, Cerezo M, *et al*. The NHGRI-EBIGWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005–D1012.
5. 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.
6. Porcu E, Rüeger S, Lepik K, *et al*. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat Commun*. 2019;10(1):3300.
7. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: Illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93(5):779–797.
8. Maurano MT, Humbert R, Rynes E, *et al*. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190–1195.
9. Pierce SE, Booms A, Prahl J, van der Schans EJC, Tyson T, Coetzee GA. Post-GWAS knowledge gap: The how, where, and when. *NPJ Parkinsons Dis*. 2020;6(1):23.
10. Sassone J, Colciago C, Cislaghi G, Silani V, Ciammola A. Huntington's disease: The current state of research with peripheral tissues. *Exp Neurol*. 2009;219(2):385–397.
11. French JD, Edwards SL. The role of noncoding variants in heritable disease. *Trends Genet*. 2020;36:880–891.
12. Horsager J, Andersen KB, Knudsen K, *et al*. Brain-first versus body-first Parkinson's disease: A multimodal imaging case-control study. *Brain*. 2020;143:3077–3088.
13. Borghammer P, Knudsen K, Fedorova TD, Brooks DJ. Imaging Parkinson's disease below the neck. *NPJ Parkinsons Dis*. 2017;3(1):15.
14. Wakabayashi K. Where and how alpha-synuclein pathology spreads in Parkinson's disease. *Neuropathology*. 2020;40(5):415–425.
15. Gelpi E, Navarro-Otano J, Tolosa E, *et al*. Multiple organ involvement by alpha-synuclein pathology in Lewy body disorders. *Mov Disord*. 2014;29(8):1010–1018.
16. Martinez-Valbuena I, Amat-Villegas I, Valenti-Azcarate R, *et al*. Interaction of amyloidogenic proteins in pancreatic β cells from subjects with synucleinopathies. *Acta Neuropathol*. 2018;135(6):877–886.
17. Homma T, Mochizuki Y, Mizutani T. Phosphorylated α-synuclein immunoreactivity in the posterior pituitary lobe. *Neuropathology*. 2012;32(4):385–389.
18. Kim S, Kwon SH, Kam TI, *et al*. Transneuronal propagation of pathologic α-synuclein from the gut to the brain models Parkinson's disease. *Neuron*. 2019;103(4):627–641.e7.
19. Schierding W, Farrow S, Fadason T, *et al*. Common variants coregulate expression of GBA and modifier genes to delay Parkinson's disease onset. *Mov Disord*. 2020;35:1346–1356.
20. Liu X, Li YI, Pritchard JK. *Trans* effects on gene expression can drive omnigenic inheritance. *Cell*. 2019;177(4):1022–1034.e6.
21. Fadason T, Schierding W, Lumley T, O'Sullivan JM. Chromatin interactions and expression quantitative trait loci reveal genetic drivers of multimorbidities. *Nat Commun*. 2018;9(1):5198.
22. Smemo S, Tena JJ, Kim K-H, *et al*. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014;507(7492):371–375.
23. Nica AC, Dermitzakis ET. Expression quantitative trait loci: Present and future. *Philos Trans R Soc B Biol Sci*. 2013;368(1620):20120362.
24. Barnat M, Capizzi M, Aparicio E, *et al*. Huntington's disease alters human neurodevelopment. *Science*. 2020;369(6505):787–793.
25. Fadason T, Schierding W, Kolbenev N, Liu J, Ingram J, O'Sullivan JM. Reconstructing the blood metabolome and genotype using long-range chromatin interactions. *Metab Open*. 2020;6:100035.
26. Aguet F, Barbeira AN, Bonazzola R, *et al*. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318–1330.
27. Kong S, Zhang Y. Deciphering Hi-C: From 3D genome to function. *Cell Biol Toxicol*. 2019;35(1):15–32.
28. Schmitt AD, Hu M, Jung I, *et al*. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep*. 2016;17(8):2042–2059.
29. Won H, de la torre-Ubieta L, Stein JL, *et al*. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. 2016;538:523–527.
30. Dixon JR, Jung I, Selvaraj S, *et al*. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518-(7539):331–336.
31. Dunham I, Kundaje A, Aldred SF, *et al*. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489-(7414):57–74.
32. Walker RL, Ramaswami G, Hartl C, *et al*. Genetic control of expression and splicing in developing human brain informs disease mechanisms. *Cell*. 2019;179(3):750–771.e22.
33. Ullah AZD, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C. SNPnexus: Assessing the functional relevance of genetic

variation to facilitate the promise of precision medicine. *Nucleic Acids Res*. 2018;46:109–113.
34. Karczewski KJ, Francioli LC, Tiao G, *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.
35. Szklarczyk D, Gable AL, Lyon D, *et al*. STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–D613.
36. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp*. 2008;2008(10):P10008.
37. Raudvere U, Kolberg L, Kuzmin I, *et al*. g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47(W1): W191–W198.
38. Makarious MB, Leonard HL, Vitale D, *et al*. Multi-modality machine learning predicting Parkinson's disease. *NPJ Parkinsons Dis*. 2022;8(1):35.
39. Yao C, Joehanes R, Johnson AD, *et al*. Dynamic role of *trans* regulation of gene expression in relation to complex traits. *Am J Hum Genet*. 2017;100(4):571–580.
40. McKenzie M, Henders AK, Caracella A, Wray NR, Powell JE. Overlap of expression Quantitative Trait Loci (eQTL) in human brain and blood. *BMC Med Genomics*. 2014;7(1):31.
41. Võsa U, Claringbould A, Westra H-J, *et al*. Large-scale *cis*- and *trans*-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300–1310.
42. Gokuladhas S, Schierding W, Golovina E, Fadason T, O'Sullivan J. Unravelling the shared genetic mechanisms underlying 18 autoimmune diseases using a systems approach. *Front Immunol*. 2021;12:693142.
43. de Klein N, Tsai EA, Vochteloo M, *et al*. Brain expression quantitative trait locus and network analysis reveals downstream effects and putative drivers for brain-related diseases. *bioRxiv*. [Preprint] https://doi.org/10.1101/2021.03.01.433439.
44. Grenn FP, Kim JJ, Makarious MB, *et al*. The Parkinson's disease genome-wide association study locus browser. *Mov Disord*. 2020;35(11):2056–2067.
45. Westra H-J, Peters MJ, Esko T, *et al*. Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45(10):1238–1243.
46. Spicuglia S, Vanhille L. Chromatin signatures of active enhancers. *Nucleus*. 2012;3(2):126–131.
47. Michel PP, Hirsch EC, Hunot S. Understanding dopaminergic cell death pathways in Parkinson disease. *Neuron*. 2016;90(4):675–691.
48. Karczewski KJ, Francioli LC, Tiao G, *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.
49. Schwamborn JC. Is Parkinson's disease a neurodevelopmental disorder and will brain organoids help us to understand it? *Stem Cells Dev*. 2018;27:968–975.
50. Johnson ME, Stecher B, Labrie V, Brundin L, Brundin P. Triggers, facilitators, and aggravators: Redefining Parkinson's disease pathogenesis. *Trends Neurosci*. 2019;42(1):4–13.
51. Lage K. Protein–protein interactions and genetic diseases: The interactome. *Biochim Biophys Acta*. 2014;1842(10):1971–1980.
52. Yoshida H, Goedert M. Phosphorylation of microtubule-associated protein tau by AMPK-related kinases. *J Neurochem*. 2012; 120(1):165–176.
53. Fasano D, Parisi S, Pierantoni GM, *et al*. Alteration of endosomal trafficking is associated with early-onset parkinsonism caused by SYNJ1 mutations. *Cell Death Dis*. 2018;9(3):385.
54. Quadri M, Fang M, Picillo M, *et al*. Mutation in the *SYNJ1* gene associated with autosomal recessive, early-onset parkinsonism. *Hum Mutat*. 2013;34(9):1208–1215.
55. Skrahina V, Gaber H, Vollstedt E, *et al*. The Rostock International Parkinson's Disease (ROPAD) Study: Protocol and initial findings. *Mov Disord*. 2021;36(4):1005–1010.
56. Le Clainche C, Pauly BS, Zhang CX, Engqvist-Goldstein AEY, Cunningham K, Drubin DG. A Hip1R–cortactin complex negatively regulates actin assembly associated with endocytosis. *EMBO J*. 2007;26(5):1199–1210.
57. Wilbur JD, Chen C-Y, Manalo V, *et al*. Actin binding by Hip1 (Huntingtin-interacting Protein 1) and Hip1R (Hip1-related Protein) is regulated by clathrin light chain. *J Biol Chem*. 2008;283:32870–32879.
58. Yang Q, Peng L, Wu Y, *et al*. Endocytic adaptor protein HIP1R controls intracellular trafficking of epidermal growth factor receptor in neuronal dendritic development. *Front Mol Neurosci*. 2018;11:447.
59. Reed X, Bandrés-Ciga S, Blauwendraat C, Cookson MR. The role of monogenic genes in idiopathic Parkinson's disease. *Neurobiol Dis*. 2019;124:230–239.
60. Shen XG, Zhang JR, Zhang XL, Wang YF, Hu YF, Guo J. Retinoic Acid-Induced Protein 14 (RAI14) promotes mTOR-mediated inflammation under inflammatory stress and chemical hypoxia in a U87 glioblastoma cell line. *Cell Mol Neurobiol*. 2019;39(2):241–254.
61. Wolf D, Hofbrucker-MacKenzie SA, Izadi M, *et al*. Ankyrin repeat-containing N-Ank proteins shape cellular membranes. *Nat Cell Biol*. 2019;21(10):1191–1205.
62. Orr CF, Rowe DB, Halliday GM. An inflammatory review of Parkinson's disease. *Prog Neurobiol*. 2002;68(5):325–340.
63. Clark JN, Whiting A, McCaffery P. Retinoic acid receptor-targeted drugs in neurodegenerative disease. *Expert Opin Drug Metab Toxicol*. 2020;16(11):1097–1108.
64. Mohammadi S, Dolatshahi M, Rahmani F. Shedding light on thyroid hormone disorders and Parkinson disease pathology: Mechanisms and risk factors. *J Endocrinol Invest*. 2021;44:1–13.
65. Chen SF, Yang YC, Hsu CY, Shen YC. Risk of Parkinson's disease in patients with hypothyroidism: A nationwide population-based cohort study. *Park Relat Disord*. 2020;74:28–32.
66. Houser MC, Tansey MG. The gut–brain axis: Is intestinal inflammation a silent driver of Parkinson's disease pathogenesis? *NPJ Parkinsons Dis*. 2017;3(1):3.
67. Suttrup I, Suttrup J, Suntrup-Krueger S, *et al*. Esophageal dysfunction in different stages of Parkinson's disease. *Neurogastroenterol Motil*. 2017;29(1):e12915.
68. Schwamborn JC. Is Parkinson's disease a neurodevelopmental disorder and will brain organoids help us to understand it? *Stem Cells Dev*. 2018;27(14):968–975.
69. Sulzer D. Multiple hit hypotheses for dopamine neuron loss in Parkinson's disease. *Trends Neurosci*. 2007;30(5):244–250.
70. Wu ZQ, Li D, Huang Y, *et al*. Caspr controls the temporal specification of neural progenitor cells through notch signaling in the developing mouse cerebral cortex. *Cereb Cortex*. 2017;27(2):1369–1385.
71. Spratley SJ, Hill CH, Viuff AH, Edgar JR, Skjødt K, Deane JE. Molecular mechanisms of disease pathogenesis differ in Krabbe disease variants. *Traffic*. 2016;17(8):908–922.
72. Marshall MS, Bongarzone ER. Beyond Krabbe's disease: The potential contribution of galactosylceramidase deficiency to neuronal vulnerability in late-onset synucleinopathies. *J Neurosci Res*. 2016;94(11):1328–1332.
73. Esposito G, Ana Clara F, Verstreken P. Synaptic vesicle trafficking and Parkinson's disease. *Dev Neurobiol*. 2012;72(1):134–144.
74. Pellegrini L, Wetzel A, Grannó S, Heaton G, Harvey K. Back to the tubule: Microtubule dynamics in Parkinson's disease. *Cell Mol Life Sci*. 2017;74(3):409–434.

75. Hou X, Watzlawik JO, Fiesel FC, Springer W. Autophagy in Parkinson's disease. *J Mol Biol*. 2020;432(8):2651–2672.
76. Sun Y, Curle AJ, Haider AM, Balmus G. The role of DNA damage response in amyotrophic lateral sclerosis. *Essays Biochem*. 2020; 64(5):847–861.
77. Wu J, Lu LY, Yu X. The role of BRCA1 in DNA damage response. *Protein Cell*. 2010;1(2):117–123.
78. Liaw H, Lee D, Myung K. DNA-PK-dependent RPA2 hyperphosphorylation facilitates DNA repair and suppresses sister chromatid exchange. *PLoS One*. 2011;6(6):e21424.
79. Wu W, Hill SE, Nathan WJ, *et al*. Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature*. 2021;593:440–444.
80. Mei Y, Su M, Sanishvili R, Chakravarthy S, Colbert CL, Sinha SC. Identification of BECN1 and ATG14 coiled-coil interface residues that are important for starvation-induced autophagy. *Biochemistry*. 2016;55(30):4239–4253.
81. Tan EK, Chao YX, West A, Chan LL, Poewe W, Jankovic J. Parkinson disease and the immune system—Associations, mechanisms and therapeutics. *Nat Rev Neurol*. 2020;16(6):303–318.
82. Yu E, Ambati A, Andersen MS, *et al*. Fine mapping of the HLA locus in Parkinson's disease in Europeans. *NPJ Parkinsons Dis*. 2021;7:84.
83. Booms A, Pierce SE, Coetzee GA. Parkinsons disease genetic risk evaluation in microglia highlights autophagy and lysosomal genes. *bioRxiv*. [Preprint] https://doi.org/10.1101/2020.08.17.254276
84. Farh KK-H, Marson A, Zhu J, *et al*. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015; 518(7539):337–343.
85. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res*. 2012;22(9):1748.
86. Yang F, Wang J, Consortium T, *et al*. Identifying *cis*-mediators for *trans*-eQTLs across many human tissues using genomic mediation analysis. *Genome Res*. 2017;27(11):1859–1871.
87. Koussounadis A, Langdon SP, Um IH, Harrison DJ, Smith VA. Relationship between differentially expressed mRNA and mRNA–protein correlations in a xenograft model system. *Sci Rep*. 2015;5(1):1–9.