



# Machine Learning Identifies Six Genetic Variants and Alterations in the Heart Atrial Appendage as Key Contributors to PD Risk Predictivity

Daniel Ho<sup>1</sup>, William Schierding<sup>1,2</sup>, Sophie L. Farrow<sup>1,2</sup>, Antony A. Cooper<sup>3,4</sup>,  
Andreas W. Kempa-Liehr<sup>5\*</sup> and Justin M. O'Sullivan<sup>1,2,6,7\*</sup>

<sup>1</sup>Liggins Institute, The University of Auckland, Auckland, New Zealand, <sup>2</sup>MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom, <sup>3</sup>Australian Parkinsons Mission, Garvan Institute of Medical Research, Sydney, NSW, Australia, <sup>4</sup>St Vincent's Clinical School, UNSW Sydney, Sydney, NSW, Australia, <sup>5</sup>Department of Engineering Science, The University of Auckland, Auckland, New Zealand, <sup>6</sup>Brain Research New Zealand, The University of Auckland, Auckland, New Zealand, <sup>7</sup>The Maurice Wilkins Centre, The University of Auckland, Auckland, New Zealand

## OPEN ACCESS

### Edited by:

Siddhita D. Mhatre,  
National Aeronautics and Space  
Administration (NASA), United States

### Reviewed by:

Dhivyaa Rajasundaram,  
University of Pittsburgh, United States  
Ganqiang Liu,  
Sun Yat-sen University, China  
Matthew Jensen,  
Yale University, United States

### \*Correspondence:

Andreas W. Kempa-Liehr  
a.kempa-liehr@auckland.ac.nz  
Justin M. O'Sullivan  
justin.osullivan@auckland.ac.nz

### Specialty section:

This article was submitted to  
Genetics of Common and Rare  
Diseases,  
a section of the journal  
Frontiers in Genetics

**Received:** 29 September 2021

**Accepted:** 09 November 2021

**Published:** 03 January 2022

### Citation:

Ho D, Schierding W, Farrow SL,  
Cooper AA, Kempa-Liehr AW and  
O'Sullivan JM (2022) Machine  
Learning Identifies Six Genetic Variants  
and Alterations in the Heart Atrial  
Appendage as Key Contributors to PD  
Risk Predictivity.  
Front. Genet. 12:785436.  
doi: 10.3389/fgene.2021.785436

Parkinson's disease (PD) is a complex neurodegenerative disease with a range of causes and clinical presentations. Over 76 genetic loci (comprising 90 SNPs) have been associated with PD by the most recent GWAS meta-analysis. Most of these PD-associated variants are located in non-coding regions of the genome and it is difficult to understand what they are doing and how they contribute to the aetiology of PD. We hypothesised that PD-associated genetic variants modulate disease risk through tissue-specific expression quantitative trait loci (eQTL) effects. We developed and validated a machine learning approach that integrated tissue-specific eQTL data on known PD-associated genetic variants with PD case and control genotypes from the Wellcome Trust Case Control Consortium. In so doing, our analysis ranked the tissue-specific transcription effects for PD-associated genetic variants and estimated their relative contributions to PD risk. We identified roles for SNPs that are connected with *INPP5P*, *CNTN1*, *GBA* and *SNCA* in PD. Ranking the variants and tissue-specific eQTL effects contributing most to the machine learning model suggested a key role in the risk of developing PD for two variants (rs7617877 and rs6808178) and eQTL associated transcriptional changes of *EAF1-AS1* within the heart atrial appendage. Similarly, effects associated with eQTLs located within the Brain Cerebellum were also recognized to confer major PD risk. These findings were replicated in two additional, independent cohorts (the UK Biobank, and NeuroX) and thus warrant further mechanistic investigations to determine if these transcriptional changes could act as early contributors to PD risk and disease development.

**Keywords:** Parkinson's disease, heart atrial appendage, SNCA, PD-SNPs, tissue specific eQTL, machine learning, GBA, Brain Cerebellum

## INTRODUCTION

Parkinson's disease (PD) is a complex neurodegenerative disease with a range of causes and clinical presentations. The diagnosis of PD is based on the presence of the cardinal motor symptoms (bradykinesia; muscular rigidity; 4–6 Hz resting tremor; postural instability) (Clarke et al., 2016). Genome wide association studies (GWAS) have identified human genetic variants that are associated

with the risk of developing PD (Spencer et al., 2011; Nalls et al., 2019). In the most recent PD GWAS meta-analysis, Nalls et al. (2019) identified 90 independent single nucleotide polymorphisms (SNPs) that are significantly associated with PD risk. There are an additional 290 PD-associated GWAS SNPs (279 in non-coding and 11 in coding regions) listed in the GWAS catalog. However, it is difficult to understand how these variants confer PD risk because the majority of the PD SNPs are located in non-coding regions of the genome (Visscher et al., 2012, 2017; Farrow et al., 2021).

Non-coding SNPs have been shown to be enriched at regulatory loci and can act as expression quantitative trait loci (eQTLs) (Duggal et al., 2014; Fadason et al., 2017, 2018; Delaneau et al., 2019; Yu et al., 2019). eQTLs typically explain a fraction of the variation in mRNA expression levels for target genes, either in *cis* (<1 Mb apart in the linear sequence) or *trans* (>1 Mb apart or located on a different chromosome). Regulatory variants (i.e., eQTLs) can impact different genes in different tissues, making it challenging to determine how SNPs convey risk for a phenotype. Determining the relative contributions of the eQTLs to the risk of developing a disease would help identify the eQTL-gene-tissue combinations that convey the risk associated with the variant (Ho et al., 2021). We have demonstrated that the three-dimensional structure of the genome can be used to help identify eQTL-gene pairs and thus the biological pathways that putatively contribute to disease etiology (Aguet et al., 2017; Schierding et al., 2020). Yet, approaches that calculate relative estimates of the tissue specific contributions that SNPs make to disease development remain elusive.

We reasoned that if PD-associated SNPs contribute to disease development through gene regulatory effects, then the tissue-specificity of these eQTLs may be an important consideration for the aetiology of the disease (Aguet et al., 2017; Ongen et al., 2017; Ho et al., 2021). Therefore, we developed a machine-learning predictor model for PD disease status that utilises and selects SNPs (without eQTLs in GTEx) and tissue-specific eQTL data, for case and control cohorts, to reveal the tissue-specific regulatory effects that are associated with PD risk. Briefly, we used a matrix of: 1) PD-associated SNPs that act as eQTLs, 2) the genes regulated by these eQTLs; 3) the tissues in which the eQTL effects were observed; and 4) SNPs that do not have eQTLs in GTEx to build a logistic predictor that was validated using genotype data from three independent studies (Spencer et al., 2011; Nalls et al., 2014; Bycroft et al., 2018). The logistic predictor model that had the highest PD predictive ability, was trained and selected using the Wellcome Trust Case Control Consortium (WTCCC) cohort. The predictor model was then validated using two datasets derived from the UK Biobank (Bycroft et al., 2018) and NeuroX-dbGap (Nalls et al., 2014). Our predictor ranked the relative contributions that six non-eQTL PD SNPs, and eQTLs that modulated gene regulation specifically within the heart atrial appendage as making the largest contributions to PD risk development.

## METHODS

Workflow for developing the PD predictor model-1 and -2 (Figure 1).

## Generation of Tissue Specific PD eQTL Reference Table

GWAS SNPs associated with PD ( $n = 290$ ,  $p$ -value  $< 1.0 \times 10^{-5}$ ; **Supplementary Table S1**) were obtained from the GWAS catalogue ([www.ebi.ac.uk/gwas](http://www.ebi.ac.uk/gwas), downloaded August 27, 2020). This SNP set included young adult-onset Parkinsonism SNPs (Siitonen et al., 2017) and the 90 SNPs identified by the most recent meta-analysis by Nalls et al. (2019). The PD associated SNPs were analysed by CoDeS3D and mapped to their tissue-specific eQTL effects for creating a PD eQTL reference table (**Supplementary Material**).

## WTCCC Cohort Cleaning and Genotype Imputation

The PD genotype dataset was acquired from the WTCCC (Request ID 10584) and were imputed by Sanger imputation service (<https://imputation.sanger.ac.uk>) (**Supplementary Material**).

## Creation of a Weighted WTCCC PD Genotype eQTL Effect Matrix

We created a matrix that combined individual genotypes with the eQTL effects for the PD-associated SNPs (**Supplementary Material**) which contains three groups of data fields:

1. Individual sample information
2. Individual sample PD-associated SNP genotype (SNP minor-allele count) weighted by GTEx tissue-specific eQTL normalised effect sizes (NES)
3. Individual PD-associated SNP genotype for the SNPs without known eQTL effects

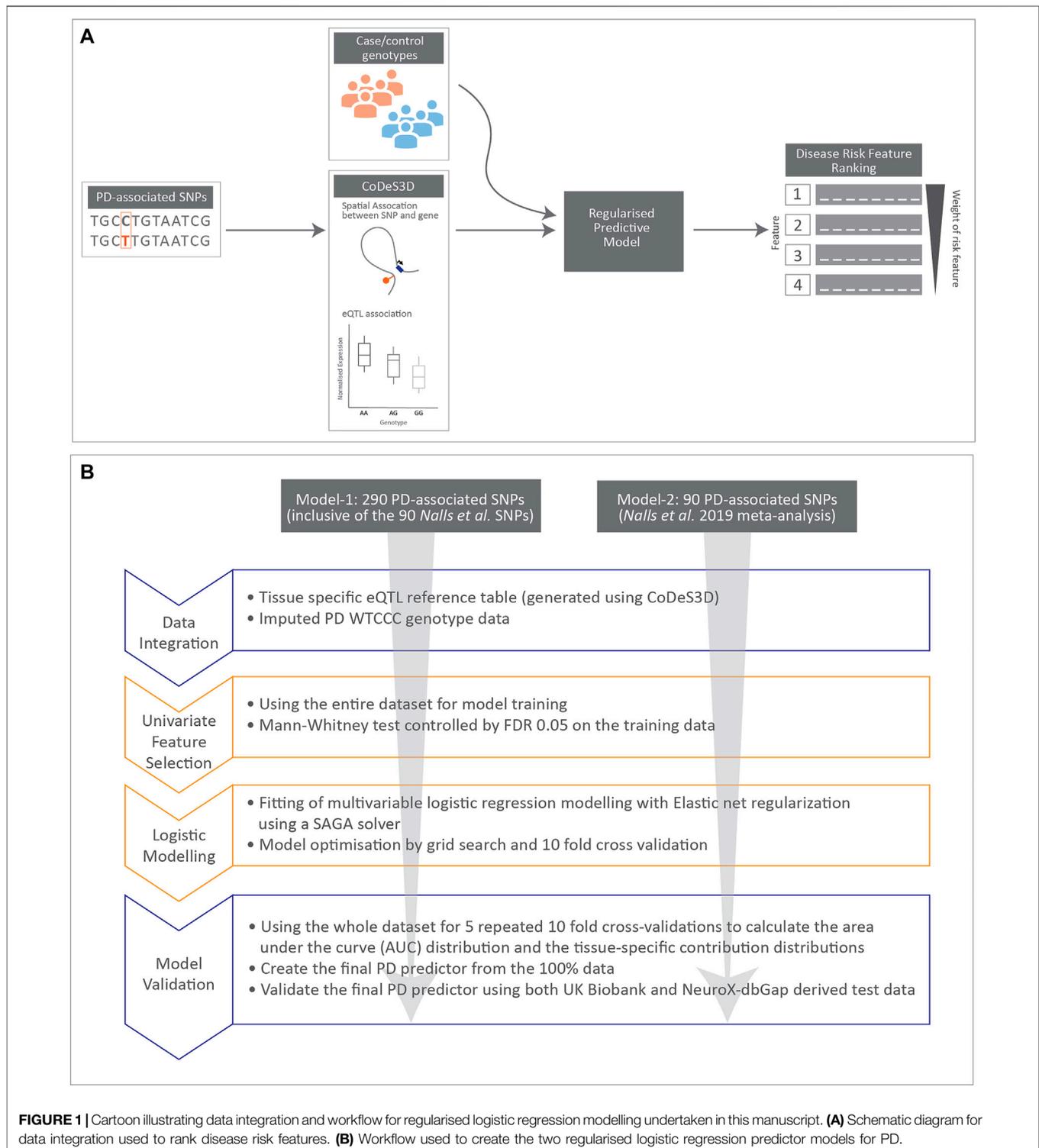
## Generation, Training, and Validation of the Regularised Logistic Regression Models (Model-1 and Model-2)

We created two regularised logistic regression models (see below): for model-1 from a weighted WTCCC PD genotype eQTL matrix for all 290 SNPs (GWAS catalogue) and for model-2 from a weighted WTCCC PD genotype eQTL matrix for the subset of 90 SNPs (Nalls et al., 2019).

We developed a regularised logistic regression predictor that incorporated a: 1) Mann-Whitney  $U$  tests in combination with Benjamini-Yekutieli (BY) procedure for controlling False Discovery Rate (FDR); and 2) multivariate prediction step with regularization that considers all features in context and removes redundant information, to identify the best combination of features for prediction of PD.

## Calculation of Tissue-Specific Contributions to PD Risk

The 50 PD regularised logistic regression predictors created from the five repeats of 10-fold cross-validation were used to test the



predictive power of the models created with the optimised predictor hyperparameters. Tissue-specific contributions to the PD risk were extracted from each of the 50 PD regularised logistic regression predictors as the sum of the absolute values of the model weights associated with each tissue.

### Validation of Model-1 and Model-2

The generalising PD predictive power of models-1 and -2 was validated by testing on two independent test datasets derived from the UK Biobank (30 test cohorts) and NeuroX-dbGap genotype data (**Supplementary Material**).

## Data Analysis

All statistical tests were performed with Scikit-learn (version 0.23.2) (Abraham et al., 2014), and tsfresh (version 0.16.0) (Christ et al., 2018). Polygenic Risk Scores were calculated by R (version 3.2.3) with pROC library (Robin et al., 2011; R Core Team, 2014).

## RESULTS

### PD-Associated SNPs Are Tissue Specific eQTLs for 1,334 eGenes

We hypothesised that PD SNPs modulate disease risk through tissue-specific eQTL effects (i.e., eQTL-eGene) (Aguet et al., 2017; Ongen et al., 2017). We analysed 290 PD-associated GWAS SNPs (Supplementary Table S1) for spatial eQTL interactions (Ramani et al., 2016; Fadason et al., 2017; Pal et al., 2019) across 49 GTEx tissues (Aguet et al., 2017). 231 of the 290 (79.7%) PD SNPs tested were involved in 18,041 tissue-specific eQTL associations (Benjamini–Hochberg FDR < 0.05 (Benjamini and Hochberg, 1995); Supplementary Table S2), regulating 1,334 eGenes across the 49 GTEx tissues. Gene ontology analysis (David Functional Annotation) (Jiao et al., 2012) identified that the regulated genes were significantly enriched for intracellular signal transduction, antigen processing and presentation of peptides, among other pathways (Supplementary Table S3).

### Modelling Genotype Data to Identify the Genetic Risk Associated With Tissue-Specific eQTL Effects for PD Disease Status

Understanding the impacts and complex networks associated with eQTLs is challenging. We hypothesised that regularised logistic regression models could be used to identify and rank the tissue-specific eQTLs that were significant contributors to PD risk.

We integrated the CoDeS3D eQTL analysis of the 290 PD SNPs with the genotype data for individuals within the WTCCC (Burton et al., 2007) PD cohort (4,366 individual samples: 1,698 cases and 2,668 controls; methods) (Spencer et al., 2011). Of the 290 PD SNPs, 281 SNPs were present in the WTCCC data. This resulted in the generation of a PD-SNP derived weighted WTCCC PD genotype eQTL effect matrix containing 17,829 tissue-specific eQTL-eGene pairs (227 SNPs, 1,310 eGenes, 49 tissues) and 54 (of the 281) SNPs that had no known eQTL effects following our CoDeS3D analysis. Uninformative features for PD prediction were removed using a Mann-Whitney *U* test (McKnight and Najab, 2010) (FDR < 0.05) (Methods). After filtering, 11,288 PD SNP derived features (53 SNPs, 245 eGenes, 49 tissues) remained within the relevant attribute subset of the weighted WTCCC PD genotype eQTL effect matrix.

To test the effectiveness of the Mann-Whitney *U* test filter, we generated a PD and type 1 diabetes (T1D) SNP derived eQTL

effect matrix using a mixed set of 290 PD and 313 T1D-associated SNPs and integrating with the WTCCC PD cohort genotypes (Supplementary Table S4). The PD + T1D SNP derived tissue-specific eQTL effect matrix included 25,052 SNP related data fields (556 SNPs, 1927 eGenes, 49 tissues). After the Mann-Whitney *U* test filtering (FDR < 0.05), 11,147 of the data fields (45 SNPs, 209 eGenes, 49 tissues) were selected using PD as the phenotypic outcome. Only one of the 313 (0.32%) T1D-associated SNP, rs1052553, remained following the Mann-Whitney *U* test filtering. Although rs1052553 has not previously been associated with PD in GWA studies, it has been implicated in PD as part of a PD risk haplotype (Tobin et al., 2008; Wider et al., 2010). Therefore, these results confirm that the Mann-Whitney *U* test filters uninformative data while preserving valuable PD information for our modelling.

We created regularised logistic regression models for PD risk using the Mann-Whitney *U* test filtered PD variant derived eQTL effect matrix (11,288 PD-SNP derived features [53 SNPs, 245 eGenes, 49 tissues]). The AUCs of the 50 PD regularised logistic regression predictors had a mean of 0.565 (distributed from 0.516 to 0.637) and a standard deviation of 0.024 (generated with the optimised predictor model hyperparameters by five repeats of 10-fold cross validation). The final PD predictor model (model-1) was trained using the entire WTCCC PD cohort. After the Mann-Whitney *U* test filtered WTCCC PD variant derived eQTL effect matrix contained 17,829 variant derived features. Model-1 selected 827 tissue-specific eQTLs and six SNPs with no eQTL effect (Supplementary Table S5). Model-1 had an enhanced diagnostic ability as represented by an AUC of 0.627 obtained using the training data.

We validated the predictive power of model-1 using two independent PD cohorts (UK Biobank (Bycroft et al., 2018) (30 datasets of 923 cases and 1,456 controls) and NeuroX-dbGap (Nalls et al., 2014, 2015)). Model-1 was validated in both cohorts, producing mean AUCs of 0.572 and 0.571 in the UK BioBank and NeuroX-dbGap cohorts, respectively. These two validation results are highly consistent and within the range of the model AUCs (0.516–0.637) estimated by the 50 optimised logistic regression predictor models.

### eQTLs Specific to the Heart Atrial Appendage Contribute to Genetic Risk in PD

We used the magnitude of the model weights (coefficients) for the genetic features, grouped by tissue-specificity of the effects, in the logistic regression model-1 as proxies for the contribution of the features to PD risk.

Six SNPs that had no identified eQTL effects (from CoDeS3D analysis of GTEx) made the most significant group contribution (18% of the total model weight) to the risk of PD development (Table 1; Figure 2). The six non-eQTL SNPs are: rs117896735, rs144210190, rs35749011, rs12726330, rs356220 and rs5019538 (Table 1). Note that the GTEx study (Aguet et al., 2017) removed rs356220 and rs5019538 from the tissue-specific eQTL data as part of their QC processing. Therefore, we were unable to test if rs356220 and rs5019538 were eQTLs. rs117896735 also has no

**TABLE 1 |** SNPs identified as being the main contributors to model-1. a) SNPs with no detected eQTL effects, and b) eQTL effects within the Heart Atrial Appendage. The model weight is the coefficient assigned to each variant or eQTL in the logistic regression predictor model-1. “\*” indicates the non eQTL SNP is in the 90 SNPs of Nalls et al.

a)			
SNP (no detected eQTLs)	Model weight		
*rs117896735_A	0.42436		
rs1442190_A	0.40106		
*rs35749011_A	0.24949		
rs12726330_A	0.18701		
rs356220_T	0.17507		
*rs5019538_G	-0.08418		
b)			
Tissue	eQTL (rsID_major allele)	Gene	Model weight
Heart_Atrial_Appendage	rs7617877_A	EAF1-AS1	0.28996
Heart_Atrial_Appendage	rs6808178_T	EAF1-AS1	0.25261
Heart_Atrial_Appendage	rs6808178_T	TMEM161B-AS1	0.16339
Heart_Atrial_Appendage	rs11707416_A	P2RY12	0.01163
Heart_Atrial_Appendage	rs26431_G	EIF3KP1	0.0089
Heart_Atrial_Appendage	rs17577094_G	RP11-259G18.3	0.00703
Heart_Atrial_Appendage	rs365825_G	RP11-259G18.3	-0.00467
Heart_Atrial_Appendage	rs17577094_G	LRRC37A4P	-0.00434
Heart_Atrial_Appendage	rs17577094_G	KANSL1-AS1	0.0036
Heart_Atrial_Appendage	rs8070723_G	RP11-259G18.3	-0.00294
Heart_Atrial_Appendage	rs365825_G	LRRC37A4P	0.00237
Heart_Atrial_Appendage	rs365825_G	KANSL1-AS1	-0.00128
Heart_Atrial_Appendage	rs17577094_G	DND1P1	0.00116
Heart_Atrial_Appendage	rs17577094_G	MAPK8IP1P2	0.00058
Heart_Atrial_Appendage	rs199515_G	RP11-259G18.3	-0.00011

eQTL effect information found in GTEx database. The other three SNPs (rs144210190, rs35749011 and rs12726330) were not detected by CoDeS3D to have spatial eQTL and eGene interactions within the Hi-C libraries used in this study.

For the top six contributing SNPs to the model, our analyses did not identify any spatial eQTL interactions. The SNPs that are in high linkage disequilibrium ( $r^2 > 0.8$ ) with these six SNPs also did not have significant spatial eQTLs. However, previous research has shown connections between these SNPs and three well-known PD-associated genes (*INPP5F*, *GBA*, *SNCA*) (Siddiqui et al., 2016; Berge-Seidl et al., 2017; Riboldi and Di Fonzo, 2019; Cao et al., 2020), and an additional gene (*CNTN1*). rs117896735, the top contributor to model-1, is an intronic variant of *INPP5F* and has previously been identified as eQTL for *INPP5F* transcript levels (the IPDGC locus browser (Grenn et al., 2020)).

The next most significant contributions to the risk of PD development involved eQTLs that affected the Heart Atrial Appendage (9%) and Brain Cerebellum (4%; **Figure 2**). The substantia nigra is viewed as a central brain region in PD yet eQTL gene regulation specific to the substantia nigra contributed

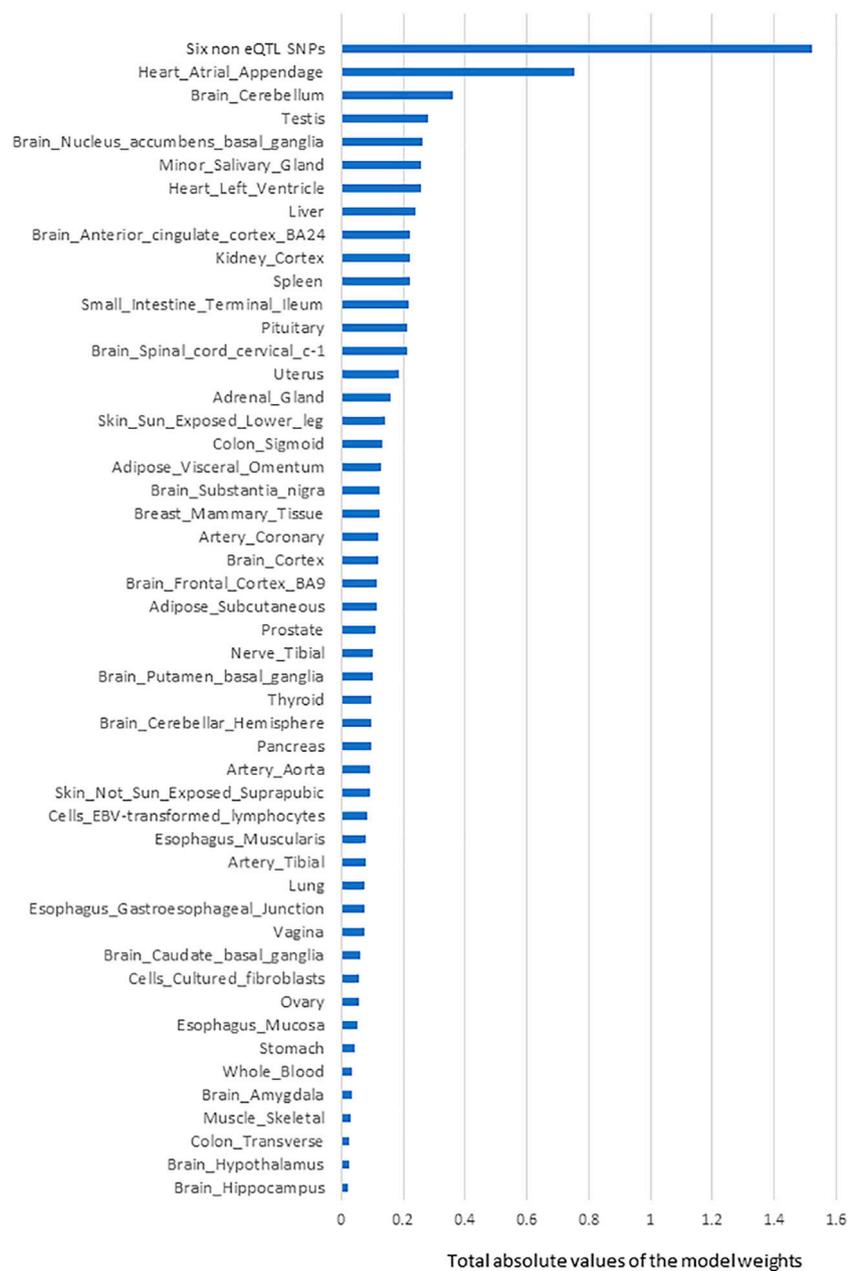
~1.5% of the risk of PD development. We repeated the calculation of the tissue-specific contribution ranking using data from the 50 optimised predictor models, generated with model-1's hyperparameters by five repeats of 10-fold cross validation (randomizing the full Mann-Whitney *U* test filtered PD variant derived eQTL effect matrix). Again, SNPs lacking known eQTL effects, Heart Atrial Appendage, and Brain Cerebellum were identified as the top three genetic contributors to the risk of PD development (**Figure 3**).

Fifteen eQTLs contributed to the Heart Atrial Appendages contribution to the risk of developing PD measured in model-1 (**Table 1**). Notably, the two biggest eQTL contributors, rs7617877 and rs6808178, each accounted for approximately 3% of the total model weight. rs7617877 and rs6808178 are in high linkage disequilibrium ( $R^2 = 0.86$ ) (Machiela and Chanock, 2015) within European populations. rs7617877 and rs6808178 do not show detectable spatial regulatory associations with their nearest genes and instead both act as eQTLs for a gene >13 Mb downstream, *EAF1-AS1*, in the Heart Atrial Appendage. *EAF1-AS1* is a long antisense non-coding RNA gene transcribed in antisense to *EAF1*, that undergoes an isoform switch, and has a significantly different transcript usage in the brains of patients with Parkinson's disease (Dick et al., 2020). Interestingly, rs6808178 also acts as a Heart Atrial Appendage eQTL for *TMEM161B-AS1* (**Table 1**), which has also been implicated in neurodegeneration (Boros et al., 2020).

## Creating a PD Logistic Regression Predictor Model Using the 90 SNPs From the PRS Calculated by Nalls et al.

In the latest PD GWAS meta-analysis, Nalls et al. (2019) identified 90 SNPs that contribute to a PRS model for PD risk. We therefore sought to understand the PD risk contribution that was specific to these 90 SNPs and created a logistic regression predictor model using only this subset. 88 of the 90 variants passed quality control (post-imputation data cleaning and quality checking). The 88 SNPs were integrated with the WTCCC PD genotype data to create a PD SNP derived eQTL effect matrix of WTCCC individual samples (4,366 individual samples: 1,698 cases and 2,668 controls). The PD SNP derived eQTL effect matrix contained 3,206 features consisting of related tissue-specific eQTL-eGene pairs (76 SNPs, 518 genes, 49 tissue types) and 12 SNPs that lacked CoDeS3D detectable eQTL effects. Mann-Whitney *U* test filtering (FDR < 0.05) left 920 features (12 SNPs, 95 genes, 49 tissue types) that were used in the subsequent logistic regression modelling (Abraham et al., 2014). Model training was repeated using the optimised predictor hyperparameters and the eQTL effect matrix for the full WTCCC cohort to create predictor model-2. Model-2 achieved in-sample PD prediction with an AUC = 0.604 using 311 features (12 SNPs, 46 genes, 49 tissue types) (**Supplementary Table S6**) that included 308 tissue-specific eQTLs and three SNPs without known eQTL effects.

We determined the tissue-specific distribution for the 50 predictors that were created with model-2's hyperparameters.

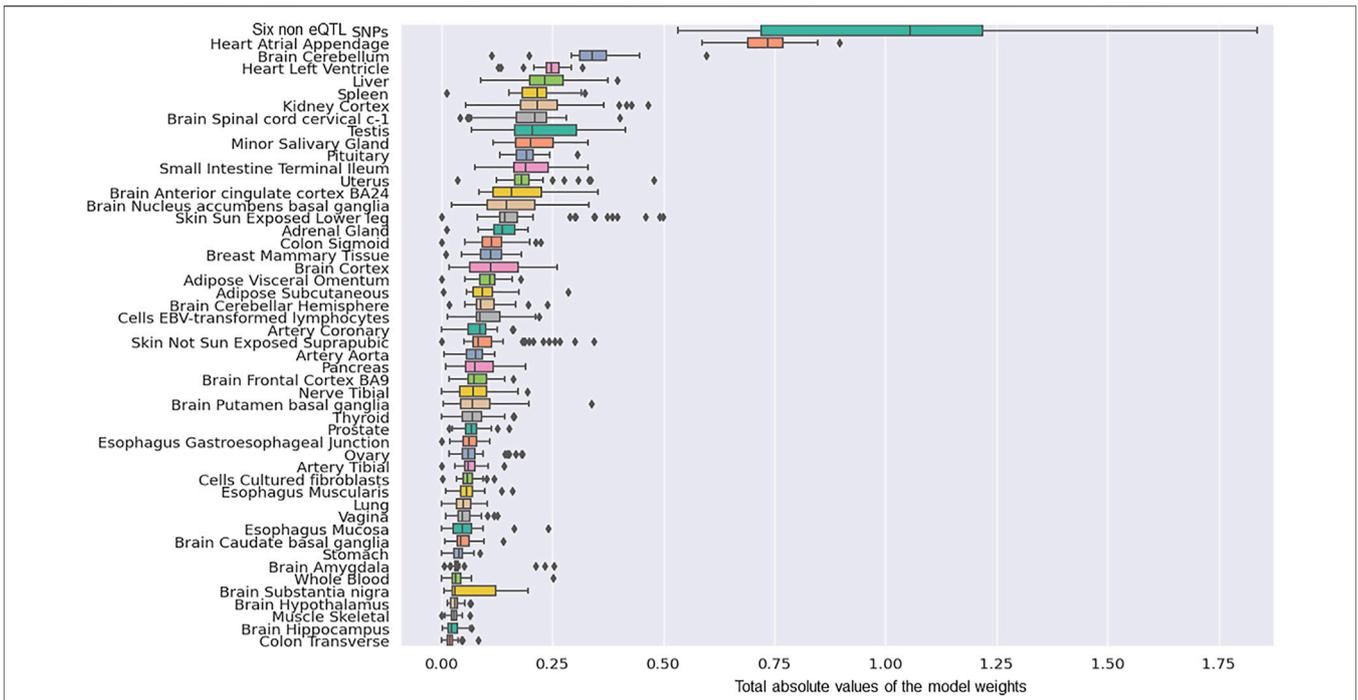


**FIGURE 2** | The rank order of tissue-specific risk contributions to risk of developing PD calculated using model-1. Tissue PD risk contributions were the sum of the absolute values of the model weights (coefficients) of the features used in the logistic regression predictor (model-1) according to their tissues. The SNPs/eQTLs that contributed to each category are listed (**Supplementary Table S5**).

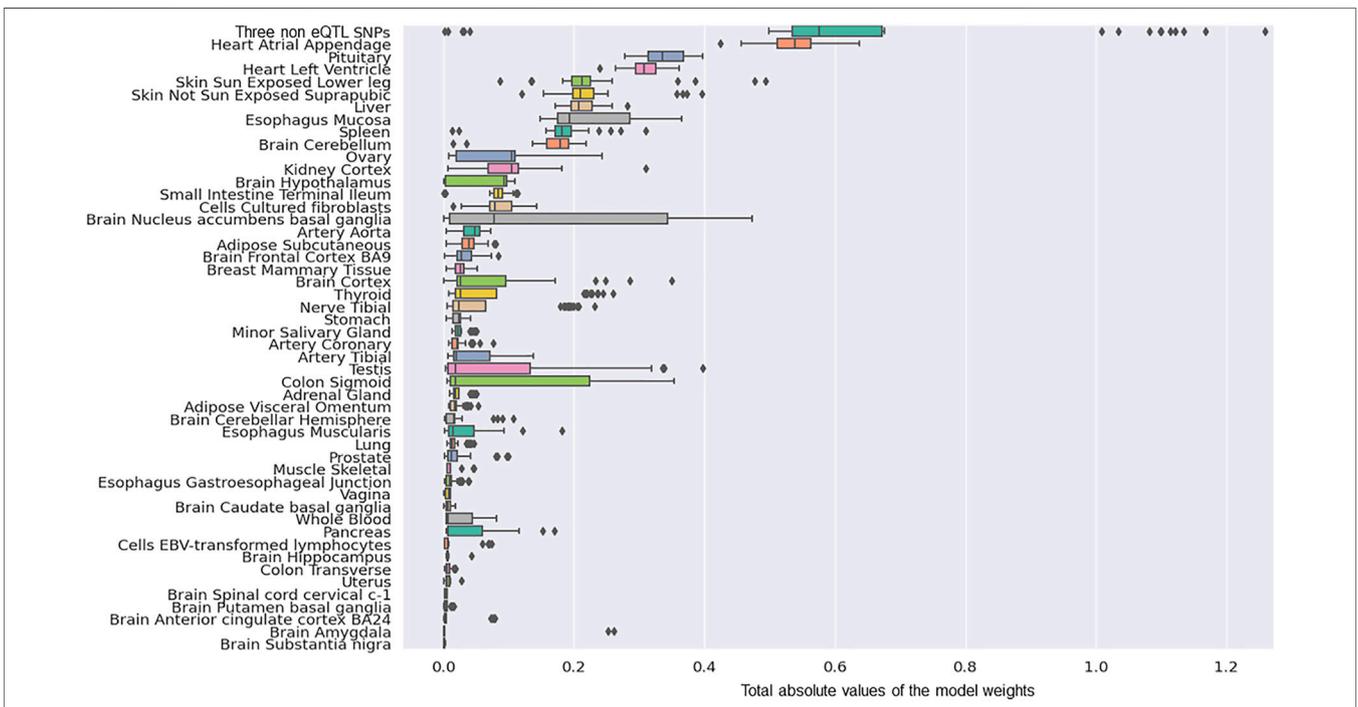
The results we observed were consistent with what we observed using model-1 (**Figure 4**). Specifically, three SNPs (rs117896735, rs35749011 and rs5019538) with no identifiable eQTL effects (**Table 2**) and the eQTLs within the Heart Atrial Appendage were the top contributors to the risk of developing PD (**Figure 4** and **Table 2**). The three non-eQTL SNPs appeared in both Model-1 and Model-2 and were observed to have similar effect sizes (both magnitude and direction) across both models. Also consistent with model-1,

model-2 identified rs6808178 as the top eQTL contributing to the Heart Atrial Appendage signal.

The PRS using the 290 PD SNPs calculated for the WTCCC cohort (AUC = 0.634) was within the range of those calculated for model-1 (AUC = 0.516–0.637) using the weighted genotype eQTL matrix. Greater variation was observed for the 90 PD SNP PRS (AUC = 0.667) when compared to that calculated by model-2 (AUC range 0.504–0.631) using the weighted genotype eQTL matrix for the WTCCC cohort.



**FIGURE 3 |** The rank order of tissue-specific risk contributions calculated across 50 predictor models created from randomised modelling and model-1's hyperparameters. The tissue ranking was consistent with that observed for model-1.



**FIGURE 4 |** The group contributions of 50 predictors created with model 2 hyperparameters by five repeats of 10 fold cross-validation.

**TABLE 2 |** SNPs identified as being the main contributors to model-2. a) SNPs with no detected eQTL effects, and b) eQTL effects within the Heart Atrial Appendage. The model weight is the coefficient assigned to each variant or eQTL in the logistic regression predictor model-2.

a)			
SNPs (no detected eQTLs)	Model weight		
rs117896735_A	0.54172		
rs35749011_A	0.47224		
rs5019538_G	-0.04028		
b)			
Tissue	eQTL (rsID_major allele)	Gene	Model weight
Heart_Atrial_Appendage	rs6808178_T	EAF1-AS1	0.53154
Heart_Atrial_Appendage	rs26431_G	EIF3KP1	0.0079
Heart_Atrial_Appendage	rs504594_A	HLA-DQA2	-0.00333
Heart_Atrial_Appendage	rs62053943_T	RP11-259G18.3	-0.0014
Heart_Atrial_Appendage	rs62053943_T	DND1P1	-0.00092
Heart_Atrial_Appendage	rs62053943_T	KANSL1-AS1	-0.00064
Heart_Atrial_Appendage	rs62053943_T	LRRC37A4P	0.00061

## DISCUSSION

The mechanisms by which PD-associated genetic variants (Nalls et al., 2014; Escott-Price et al., 2015; Lill et al., 2015; Visscher et al., 2017) contribute to disease risk and development have not been fully elucidated. Yet, it is critical that we identify the mechanisms by which they impact on PD because this will allow patient stratification and the development of therapeutics that target disease progression and not just pathology. We used machine learning to understand the genetic architecture of PD risk, by identifying and ranking the pivotal risk variants and tissue-specific eQTL effects that contribute to such risk. Curated PD-associated SNPs from the GWAS catalogue (MacArthur et al., 2017) were analysed to identify their tissue-specific eQTL effects. Regularised logistic regression predictor models that evaluated PD risk were built and validated across three independent case: control cohorts (Spencer et al., 2011; Nalls et al., 2014; Bycroft et al., 2018). Model-1 (generated from 290 SNPs) identified six SNPs without known eQTL effects and the SNP modulated gene regulation within the Heart Atrial Appendage as being the major contributors to the predicted risk of developing PD. A second model (Model-2) that was generated using only 90 SNPs (Nalls et al., 2019) (which were previously identified to have the greatest predictive power with a PRS analysis) confirmed a subset of the top predictors we observed with model-1. Collectively, our results confirm roles for SNPs that are significantly connected with *INPP5P*, *CNTN1*, *GBA* and *SNCA* in PD and separately suggest a key role for transcriptional changes within the heart atrial appendage in the risk of developing PD. Effects associated with eQTLs located within the Brain Cerebellum were also recognized to confer major PD risk in the more extensive model (model-1) consistent with current hypotheses suggesting the Brain Cerebellum plays a role in PD development (Wu and Hallett, 2013; Seidel et al., 2017; Riou et al., 2021).

*INPP5P* is a known risk gene for PD (Cao et al., 2020) that regulates STAT3 intracellular signalling pathways (Kim et al., 2014) and has functional roles in cardiac myocytes and axons (Zhu et al., 2009; Zou et al., 2015). rs1442190 is an intronic variant within *CNTN1*, a known risk gene for dementia with Lewy bodies (Guerreiro et al., 2018; Chatterjee et al., 2020) that encodes a cell adhesion protein, which is important for axon connections and nervous system development (Anderson et al., 2018). rs35749011 and rs12726330 are linked to the well-known PD-associated gene *GBA* (Berge-Seidl et al., 2017) through strong linkage disequilibrium connections ( $R^2 = 0.77$  (Machiela and Chanock, 2015)) with rs2230288 (Berge-Seidl et al., 2017; Mata et al., 2017), a missense coding variant located within *GBA*. rs35749011 has eQTL effects on *GBA* gene identified by the IPDGC database (Grenn et al., 2020). The final two SNPs, rs356220 and rs5019538, are located downstream of *SNCA*. *SNCA* encodes  $\alpha$ -synuclein, which is central to PD pathogenesis (Siddiqui et al., 2016). The IPDGC database (Grenn et al., 2020) indicates that rs5019538 has eQTL effects on *SNCA*. Notably, rs356220 had the strongest association to PD in the original WTCCC GWAS (Spencer et al., 2011). Therefore, there is sufficient evidence that has previously associated these six variants with PD through connections to PD risk genes.

## Allele Specific Regulatory Changes in the Heart Atrial Appendage Confer PD Risk

We identified that eQTLs specific to the heart atrial appendage make a reproducible and substantial (second highest) contribution to the risk of developing PD. The heart atrial appendage is a trigger site of atrial fibrillation (AF) (Di Biase et al., 2010) and highly associated with hypertension and stroke (Hart and Halperin, 2001; Stöllberger et al., 2003; Turagam et al., 2018; Du et al., 2020). Notably, none of the heart atrial appendage eQTLs we identified have been previously associated with cardiac health or atrial fibrillation by GWAS (GWAS catalog, November 2, 2021). However, the genes on the opposite strands to the two antisense genes (i.e., *TMEM161B-AS1* and *KANSL1-AS1*) have been previously implicated in regulating cardiac rhythm with zebrafish model (i.e., *TMEM161B* (Koopman et al., 2021)) and congenital heart defects in humans (i.e., *KANSL1* (Koolen et al., 2016; León et al., 2017)). However, there is a growing body of research indicating a close relationship between cardiovascular health and PD development (Awerbuch and Sandyk, 1994; Ascherio and Tanner, 2009; Fang et al., 2018; Scorza et al., 2018; Hong et al., 2019; Potashkin et al., 2020). The eQTL rs11707416 and its regulated eGene P2RY12 have been implicated in the brain blood barrier maintenance functions of microglial cells (Andersen et al., 2021). Moreover, AF has been strongly related to early-stage PD (Hong et al., 2019). Moreover, Moon et al. identified that patients with PD have an increased risk of AF, with a threefold increased risk (HR: 3.06, 95% CI: 1.20–7.77) of AF in younger PD patients (age: 40–49 years) (Han et al., 2021). Observations of a cross-sectional PD patient cohort have identified abnormal blood flow patterns in brains (Teune et al., 2014) and it is argued that AF-associated perturbation of the brain blood supply networks promotes tissue inflammation and damage leading to PD pathogenesis (Junejo et al., 2020).

Amongst the 15 eQTL features that combined to make the Heart Atrial Appendage's contribution to the risk of developing PD (Tables 1, 2), eQTL up-regulation of *EAF1-ASI* (a long non-coding mRNA) made the greatest contribution. *EAF1-ASI* has different isoforms some of which overlap *EAF1* and *COLQ* (collagen like tail subunit of asymmetric acetylcholinesterase). Elevated *EAF1-ASI* transcript levels have previously been identified by differential gene expression analyses in brain tissue samples from PD patients (Dick et al., 2020). It is interesting to speculate that the impact of this change is mediated through the interaction of *EAF1-ASI* with *EAF1*. Notably, *EAF1* has been associated with both neural development (Liu et al., 2013) and TGF- $\beta$  signalling (Liu et al., 2017), which is a key pathway in many cardiac physiological processes (Yousefi et al., 2020). As such, the deregulation of *EAF1-ASI* might impact on cardiac health. However, the anti-sense overlap is limited to the 3' UTR of *EAF1* (UCSC Genome browser GRCh38/hg38). Therefore, we propose that future studies should investigate the regulatory impacts of *EAF1-ASI* on *EAF1* and the consequences of alterations in expression levels on heart function and PD disease. We contend that understanding this relationship may help to decipher the complex interactions connecting cardiovascular fitness and PD pathogenesis.

Similar to our work, Li et al. (2019) used linkage disequilibrium score regression (LDSC) analysis (Finucane et al., 2015, 2018) to identify enrichments of PD risk signals in six GTEx (Aguet et al., 2017) central nervous system tissues. However, three subsequent studies using LDSC have failed to reproduce Li et al.'s results (Gagliano et al., 2016; Reynolds et al., 2019; Bryois et al., 2021). LDSC focuses on measuring the risk enrichment of genes uniquely expressed in each GTEx tissue (Finucane et al., 2015, 2018). By contrast, our model does not assume unique tissue expression. Rather, it identifies the risk associated with the PD-SNP, or the expression of all genes modulated specifically by PD SNPs in different or multiple GTEx tissues. We therefore hypothesise that the fact that Li et al. did not identify any signals in heart tissues is likely due to the differences in the assumptions underlying the methodologies.

## What are the Functions of the Six SNPs for Which We Identified No eQTLs?

It should not be ignored that Model-1 (generated from 290 SNPs) identified six SNPs without known eQTL effects as making the greatest contribution to PD risk. A subset of these SNPs (rs117896735, rs35749011 and rs5019538) were confirmed in Model-2. Given the contribution of these SNPs to the models, it is interesting to speculate on the function(s) of these SNPs with respect to PD risk. As noted earlier, several of the SNPs are connected to well-known PD-associated genes (*INPP5F*, *GBA*, *SNCA*) (Siddiqui et al., 2016; Berge-Seidl et al., 2017; Riboldi and Di Fonzo, 2019; Cao et al., 2020). It remains possible that these SNPs may be eQTLs for these genes at different developmental stages, or in tissues or cell types that are not represented in the datasets we used in this study. Consistent with this, the top contributor to model-1 is an

intronic variant of *INPP5F* that has previously been identified as an eQTL for *INPP5F* transcript levels (the IPDGC locus browser (Grenn et al., 2020)). However, the inclusion of these SNPs in the models did not assume a functional impact on transcription. As such, the SNPs may impact on PD risk through processes or functions that: 1) do not require the formation of spatially constrained eQTLs; 2) affect transcript levels by another mechanism (e.g., DNA methylation and protein-protein interactions) (Ryan and Matthews, 2005; Volkov et al., 2016); or 3) function through another as yet uncharacterized mechanism. While we are currently unable to further expand on the function(s) of these SNPs, the significant contributions they make to PD require further experimental investigation.

## What Are the Limitations of Our Study?

We acknowledge several limitations within our work. Firstly, our models were not generated for use in clinical screening and the predictivity is clearly insufficient for such applications. Rather, our objective was to construct models that enabled the determination of the relative SNP-gene-tissue contributions to PD risk in individuals, using recognized PD-associated SNPs identified in population level association studies. We also acknowledge that the individuals in the included datasets are predominantly of European descent, and thus the significance of our findings are limited to this ethnicity. One limitation that impacts the vast majority of PD research is the lack of consistency in diagnostic criteria from one cohort to the other, and our study is not exempt from this.

The limitations within our study do not detract from the strengths of our model which included the fact that contributing features were: 1) validated across three independent cohorts; 2) easily identifiable; and 3) consistently identified genomic regions that are unanimously recognised as being associated with PD (e.g., *SNCA*).

Our approach provides a significant advance over other previously reported methods. The novelty revolves around the ability of our method to: 1) rank the contributions that SNPs make to a phenotype through regulatory changes; 2) identify the tissues in which these changes are occurring; and 3) include effects from variants that do not have detectable eQTLs in the reference library that is used in the assay. Finally, the consistency between models and ability to filter extraneous SNPs (e.g., T1D eQTLs) out of the final predictor is another strength of this study. The higher predictive power observed for model-1 (Supplementary Table S7) may be explained by the observation that the final model included more features (827 vs 308). However, given that model-1 leveraged 290 PD-associated SNPs, the result also suggests that the 90 SNPs, originally identified as part of the Nalls et al. PRS analysis (Nalls et al., 2019), do in fact contain the major genetic components that are associated with the risk of developing PD. Therefore, while other genetic signals clearly remain to be identified, the finding that both models consistently identified the same SNPs and heart atrial appendage eQTLs as the top contributors to the risk of developing PD further confirms the significance of these observations.

## Conclusion

In conclusion, we applied machine learning algorithms to rank the pivotal variants and tissue-specific eQTL effects that may contribute to the risk of developing PD by integrating PD-associated SNPs with information on genome organisation, tissue-specific eQTLs and the genotypes of PD cases and controls. Across our two models we consistently identified the same SNPs and heart atrial appendage eQTLs, linked to *EAF1-ASI* regulation, as the top contributors to the risk of developing PD. It could be argued that the lack of significant findings in established PD tissues (e.g., the *substantia nigra*) indicates that our models did not identify the biologically significant variants. However, studies show that disease associated SNPs are enriched in enhancer elements (GTEx Consortium, 2021) and it is widely recognized that pathology does not necessarily equate to the root cause of the PD. Rather the etiology of PD, and other movement disorders, is consistent with life-long contributions from early developmental changes. As such, we contend that our results, which replicate across three independent biological cohorts, provide insights into the non-motor multi-tissue features/processes (non-motor PD) that collectively, or singularly contribute to an individual's progression to motor symptoms (motor PD) with age (Mhyre et al., 2012; Schapira et al., 2017). Future experiments should test the putative tissue specific enhancer activities we have identified using luciferase enhancer assays within edited human cell-lines that are isogenic except for the change of interest. Validation of the biological significance of the tissue level processes could then be addressed using tissue organoids and humanized animal models. These analyses should be performed in parallel with prospective studies that include analyses of the ever-expanding datasets (pulse oxygen levels, heart rate and blood pressure) that are being collected by wearable BioActive devices (e.g., Galaxy watch, Fitbits) Validation of our findings will provide insights into high value therapies for the prevention or delay of PD development.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

The CoDeS3D pipeline is available at: <https://github.com/Genome3d/codes3d-v2>. The Python scripts and machine learning code used in this analysis are available at: [https://github.com/Genome3d/PD\\_lg\\_predictor\\_analysis](https://github.com/Genome3d/PD_lg_predictor_analysis). Python version 3.7.3 was used for all the Python scripts. eQTL information from 49 human tissues were obtained from the Genotype-Tissue Expression database [GTEx] v8; [www.gtexportal.org](http://www.gtexportal.org). PD genotype datasets were acquired from the: Wellcome Trust Case Control Consortium (Request ID 10584); UKBioBank (Application Number 61507); NeuroX-dbGap (dbGap project#98581-1).

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and

institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

DH was responsible for research project execution (1A), statistical analysis, design, execution (2A & B); manuscript preparation, writing of the first draft (3A). WS, AK-L, and JO'S co-supervised DH and were responsible for statistical analysis review and critique (2C). WS, AK-L, AC, SF, and JO'S were responsible for manuscript review and critique (3B). JO'S was responsible for the research project conception, organization and execution (1A, B and C).

## FUNDING

WS, SF, AC, and JO'S 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 JO'S. SF was funded by a Liggins Institute Doctoral Scholarship and Dines Family Trust. AC received grant funding from the Australian government. DH was funded by an MBIE Catalyst grant (The New Zealand-Australia Life-Course Collaboration on Genes, Environment, Nutrition and Obesity (GENO); UOAX1611; to JO'S).

## ACKNOWLEDGMENTS

The authors would like to thank Tayaza Fadason and the Genomics and Systems Biology Group at the Liggins Institute for their helpful discussions. The eQTL data used for the analyses were obtained from Genotype-Tissue Expression (GTEx) Portal. We would like to acknowledge the funders of GTEx Project—common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from [www.wtccc.org.uk](http://www.wtccc.org.uk). Funding for the project was provided by the Wellcome Trust under award 076113, 085475 and 090355. This research has been conducted using the UK Biobank Resource under Application Number 61507. This study has been previously appeared online as preprint in medRxiv (<https://www.medrxiv.org/content/10.1101/2021.06.29.21259734v1>).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.785436/full#supplementary-material>

## REFERENCES

- Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine Learning for Neuroimaging with Scikit-Learn. *Front. Neuroinform.* 8, 14. doi:10.3389/fninf.2014.00014
- Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., Jo, B., et al. (2017). Genetic Effects on Gene Expression across Human Tissues. *Nature* 550, 204–213. doi:10.1038/nature24277
- Anderson, C., Gerding, W. M., Fraenz, C., Schlüter, C., Friedrich, P., Raane, M., et al. (2018). PLP1 and CNTN1 Gene Variation Modulates the Microstructure of Human white Matter in the Corpus Callosum. *Brain Struct. Funct.* 223, 3875–3887. doi:10.1007/s00429-018-1729-7
- Ascherio, A., and Tanner, C. M. (2009). Use of Antihypertensives and the Risk of Parkinson Disease. *Neurology* 72, 578–579. doi:10.1212/01.wnl.0000344171.22760.24
- Awerbuch, G. L., and Sandyk, R. (1994). Autonomic Functions in the Early Stages of Parkinson's Disease. *Int. J. Neurosci.* 74, 9–16. doi:10.3109/00207459408987224
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57, 289–300. doi:10.2307/234610110.1111/j.2517-6161.1995.tb02031.x
- Berge-Seidl, V., Pihlström, L., Maple-Grødem, J., Forsgren, L., Linder, J., Larsen, J. P., et al. (2017). The GBA Variant E326K Is Associated with Parkinson's Disease and Explains a Genome-wide Association Signal. *Neurosci. Lett.* 658, 48–52. doi:10.1016/j.neulet.2017.08.040
- Boros, F. A., Maszlag-Török, R., Vécsei, L., and Klivényi, P. (2020). Increased Level of NEAT1 Long Non-coding RNA Is Detectable in Peripheral Blood Cells of Patients with Parkinson's Disease. *Brain Res.* 1730, 146672. doi:10.1016/j.brainres.2020.146672
- Bryois, J., Skene, N. G., Hansen, T. F., Kogelman, L. J. A., Watson, H. J., Liu, Z., et al. (2021). Genetic Identification of Cell Types Underlying Brain Complex Traits Yields Insights into the Etiology of Parkinson's Disease. *Nat. Genet.* 52, 482–493. doi:10.1038/s41588-020-0610-9
- Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., et al. (2007). Genome-wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls. *Nature* 447, 661–678. doi:10.1038/nature05911
- Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank Resource with Deep Phenotyping and Genomic Data. *Nature* 562, 203–209. doi:10.1038/s41586-018-0579-z
- Cao, M., Park, D., Wu, Y., and De Camilli, P. (2020). Absence of Sac2/INPP5F Enhances the Phenotype of a Parkinson's Disease Mutation of Synaptotagmin 1. *Proc. Natl. Acad. Sci. USA* 117, 12428–12434. doi:10.1073/pnas.2004335117
- Chatterjee, M., van Steenoven, I., Huisman, E., Oosterveld, L., Berendse, H., van der Flier, W. M., et al. (2020). Contactin-1 Is Reduced in Cerebrospinal Fluid of Parkinson's Disease Patients and Is Present within Lewy Bodies. *Biomolecules* 10, 1177. doi:10.3390/biom10081177
- Christ, M., Braun, N., Neuffer, J., and Kempa-Liehr, A. W. (2018). Time Series FeatuRe Extraction on Basis of Scalable Hypothesis Tests (Tsfresh - A Python Package). *Neurocomputing* 307, 72–77. doi:10.1016/j.neucom.2018.03.067
- Clarke, C. E., Patel, S., Ives, N., Rick, C. E., Woolley, R., Wheatley, K., et al. (2016). *UK Parkinson's Disease Society Brain Bank Diagnostic Criteria*. Southampton: NIHR Journals Library.
- Delaneau, O., Zazhytska, M., Borel, C., Giannuzzi, G., Rey, G., Howald, C., et al. (2019). Chromatin Three-Dimensional Interactions Mediate Genetic Effects on Gene Expression. *Science* 364, 364. doi:10.1126/science.aat8266
- Di Biase, L., Burkhardt, J. D., Mohanty, P., Sanchez, J., Mohanty, S., Horton, R., et al. (2010). Left Atrial Appendage. *Circulation* 122, 109–118. doi:10.1161/CIRCULATIONAHA.109.928903
- Dick, F., Nido, G. S., Alves, G. W., Tysnes, O.-B., Nilsen, G. H., Dölle, C., et al. (2020). Differential Transcript Usage in the Parkinson's Disease Brain. *PLoS Genet.* 16, e1009182–24. doi:10.1371/journal.pgen.1009182
- Du, W., Dai, M., Wang, M., Gong, Q., Ye, T.-Q., Wang, H., et al. (2020). Large Left Atrial Appendage Predicts the Ablation Outcome in Hypertensive Patients with Atrial Fibrillation. *J. Electrocardiol.* 63, 139–144. doi:10.1016/j.jelectrocard.2020.07.017
- Duggal, G., Wang, H., and Kingsford, C. (2014). Higher-order Chromatin Domains Link eQTLs with the Expression of Far-Away Genes. *Nucleic Acids Res.* 42, 87–96. doi:10.1093/nar/gkt857
- Escott-Price, V., Nalls, M. A., Morris, H. R., Lubbe, S., Brice, A., Gasser, T., et al. (2015). Polygenic Risk of P Arkinson Disease Is Correlated with Disease Age at Onset. *Ann. Neurol.* 77, 582–591. doi:10.1002/ana.24335
- Fadason, T., Ekblad, C., Ingram, J. R., Schierding, W. S., and O'Sullivan, J. M. (2017). Physical Interactions and Expression Quantitative Traits Loci Identify Regulatory Connections for Obesity and Type 2 Diabetes Associated SNPs. *Front. Genet.* 8. doi:10.3389/fgene.2017.00150
- Fadason, T., Schierding, W., Lumley, T., and O'Sullivan, J. M. (2018). Chromatin Interactions and Expression Quantitative Trait Loci Reveal Genetic Drivers of Multimorbidities. *Nat. Commun.* 9, 1–13. doi:10.1038/s41467-018-07692-y
- Fang, X., Han, D., Cheng, Q., Zhang, P., Zhao, C., Min, J., et al. (2018). Association of Levels of Physical Activity with Risk of Parkinson Disease. *JAMA Netw. Open* 1, e182421. doi:10.1001/jamanetworkopen.2018.2421
- Farrow, S. L., Schierding, W., Gokuladhas, S., Golovina, E., Fadason, T., Cooper, A. A., et al. (2021). Establishing Gene Regulatory Networks from Parkinson's Disease Risk Loci. bioRxiv. doi:10.1101/2021.04.08.439080
- Finucane, H. K., Bulik-Sullivan, B., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., et al. (2015). Partitioning Heritability by Functional Annotation Using Genome-wide Association Summary Statistics. *Nat. Genet.* 47, 1228–1235. doi:10.1038/ng.3404
- Finucane, H. K., Reshef, Y. A., Reshef, Y. A., Anttila, V., Slowikowski, K., Gusev, A., et al. (2018). Heritability Enrichment of Specifically Expressed Genes Identifies Disease-Relevant Tissues and Cell Types. *Nat. Genet.* 50, 621–629. doi:10.1038/s41588-018-0081-4
- Gagliano, S. A., Pouget, J. G., Hardy, J., Knight, J., Barnes, M. R., Ryten, M., et al. (2016). Genomics Implicates Adaptive and Innate Immunity in Alzheimer's and Parkinson's Diseases. *Ann. Clin. Transl. Neurol.* 3, 924–933. doi:10.1002/acn3.369
- Grenn, F. P., Kim, J. J., Makarious, M. B., Iwaki, H., Illarionova, A., Brodin, K., et al. (2020). The Parkinson's Disease Genome-Wide Association Study Locus Browser. *Mov. Disord.* 35, 2056–2067. doi:10.1002/mds.28197
- GTEX Consortium (2021). The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues. *Science* 369, 1318–1330. doi:10.1126/science.aaz1776
- Guerreiro, R., Ross, O. A., Kun-Rodrigues, C., Hernandez, D. G., Orme, T., Eicher, J. D., et al. (2018). Investigating the Genetic Architecture of Dementia with Lewy Bodies: a Two-Stage Genome-wide Association Study. *Lancet Neurol.* 17, 64–74. doi:10.1016/s1474-4422(17)30400-3
- Han, S., Moon, I., Choi, E. K., Han, K. D., Cho, H. C., Lee, S. Y., et al. (2021). Increased Atrial Fibrillation Risk in Parkinson's Disease: A Nationwide Population-based Study. *Ann. Clin. Transl. Neurol.* 8, 238–246. doi:10.1002/acn3.51279
- Hart, R. G., and Halperin, J. L. (2001). Atrial Fibrillation and Stroke. *Stroke* 32, 803–808. doi:10.1161/01.str.32.3.803
- Ho, D., Nyaga, D. M., Schierding, W., Saffery, R., Perry, J. K., Taylor, J. A., et al. (2021). Identifying the Lungs as a Susceptible Site for Allele-specific Regulatory Changes Associated with Type 1 Diabetes Risk. *Commun. Biol.* 4, 1072. doi:10.1038/s42003-021-02594-0
- Hong, C.-T., Chan, L., Wu, D., Chen, W.-T., and Chien, L.-N. (2019). Association between Parkinson's Disease and Atrial Fibrillation: A Population-Based Study. *Front. Neurol.* 10. doi:10.3389/fneur.2019.00022
- Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., et al. (2012). DAVID-WS: A Stateful Web Service to Facilitate Gene/protein List Analysis. *Bioinformatics* 28, 1805–1806. doi:10.1093/bioinformatics/bts251
- Junejo, R. T., Lip, G. Y. H., and Fisher, J. P. (2020). Cerebrovascular Dysfunction in Atrial Fibrillation. *Front. Physiol.* 11. doi:10.3389/fphys.2020.01066
- Kim, H. S., Li, A., Ahn, S., Song, H., and Zhang, W. (2014). Inositol Polyphosphate-5-Phosphatase F (INPP5F) Inhibits STAT3 Activity and Suppresses Gliomas Tumorigenicity. *Sci. Rep.* 4, 7330. doi:10.1038/srep07330
- Koolen, D. A., Pfundt, R., Pfundt, R., Linda, K., Beunders, G., Veenstra-Knol, H. E., et al. (2016). The Koolen-De Vries Syndrome: A Phenotypic Comparison of Patients with a 17q21.31 Microdeletion versus a KANSL1 Sequence Variant. *Eur. J. Hum. Genet.* 24, 652–659. doi:10.1038/ejhg.2015.178
- Koopman, C. D., de Angelis, J., Iyer, S. P., Verkerk, A. O., Da Silva, J., Berecki, G., et al. (2021). The Zebrafish Grp1 Mutant Uncovers an Evolutionarily Conserved Role for Tmem161b in the Control of Cardiac Rhythm. *Proc. Natl. Acad. Sci. USA* 118, e2018220118–10. doi:10.1073/pnas.2018220118

- León, L. E., Benavides, F., Espinoza, K., Vial, C., Alvarez, P., Palomares, M., et al. (2017). Partial Microduplication in the Histone Acetyltransferase Complex Member KANSL1 Is Associated with Congenital Heart Defects in 22q11.2 Microdeletion Syndrome Patients. *Sci. Rep.* 7, 1–8. doi:10.1038/s41598-017-01896-w
- Li, Y. I., Wong, G., Humphrey, J., and Raj, T. (2019). Prioritizing Parkinson's Disease Genes Using Population-Scale Transcriptomic Data. *Nat. Commun.* 10, 1–10. doi:10.1038/s41467-019-08912-9
- Lill, C. M., Hansen, J., Olsen, J. H., Binder, H., Ritz, B., and Bertram, L. (2015). Impact of Parkinson's Disease Risk Loci on Age at Onset. *Mov. Disord.* 30, 847–850. doi:10.1002/mds.26237
- Liu, J.-X., Xu, Q.-H., Li, S., Yu, X., Liu, W., Ouyang, G., et al. (2017). Transcriptional Factors Eaf1/2 Inhibit Endoderm and Mesoderm Formation via Suppressing TGF- $\beta$  Signaling. *Biochim. Biophys. Acta (BBA) - Gene Regul. Mech.* 1860, 1103–1116. doi:10.1016/j.bbagr.2017.09.001
- Liu, J.-X., Zhang, D., Xie, X., Ouyang, G., Liu, X., Sun, Y., et al. (2013). Eaf1 and Eaf2 Negatively Regulate Canonical Wnt/ $\beta$ -Catenin Signaling. *Dev.* 140, 1067–1078. doi:10.1242/dev.086157
- MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., et al. (2017). The New NHGRI-EBI Catalog of Published Genome-wide Association Studies (GWAS Catalog). *Nucleic Acids Res.* 45, D896–D901. doi:10.1093/nar/gkw1133
- Machiela, M. J., and Chanock, S. J. (2015). LDlink: a Web-Based Application for Exploring Population-specific Haplotype Structure and Linking Correlated Alleles of Possible Functional Variants: Fig. 1. *Bioinformatics* 31, 3555–3557. doi:10.1093/bioinformatics/btv402
- Mata, I. F., Johnson, C. O., Leverenz, J. B., Weintraub, D., Trojanowski, J. Q., Van Deerlin, V. M., et al. (2017). Large-scale Exploratory Genetic Analysis of Cognitive Impairment in Parkinson's Disease. *Neurobiol. Aging* 56, e1. doi:10.1016/j.neurobiolaging.2017.04.009
- McKnight, P. E., and Najab, J. (2010). "Mann Whitney U Test," in *Corsini Encycl. Psychol.* doi:10.1002/9780470479216.corpsy0524
- Mhyre, T. R., Boyd, J. T., Hamill, R. W., Maguire-Zeiss, K. A., and Room, C. (2012). Parkinson's Disease. *Subcell Biochem.* 65, 389–455. doi:10.1007/978-94-007-5416-4\_16
- Nalls, M. A., Blauwendraat, C., Vallerga, C. L., Heilbron, K., Bandres-Ciga, S., Chang, D., et al. (2019). Identification of Novel Risk Loci, Causal Insights, and Heritable Risk for Parkinson's Disease: a Meta-Analysis of Genome-wide Association Studies. *Lancet Neurol.* 18, 1091–1102. doi:10.1016/S1474-4422(19)30320-5
- Nalls, M. A., Pankratz, N., Lill, C. M., Do, C. B., Hernandez, D. G., Saad, M., et al. (2014). Large-scale Meta-Analysis of Genome-wide Association Data Identifies Six New Risk Loci for Parkinson's Disease. *Nat. Genet.* 46, 989–993. doi:10.1038/ng.3043.Large-scale
- Nalls, M. A., Bras, J., Hernandez, D. G., Keller, M. F., Majounie, E., Renton, A. E., et al. (2015). NeuroX, a Fast and Efficient Genotyping Platform for Investigation of Neurodegenerative Diseases. *Neurobiol. Aging* 36, e7–1605. doi:10.1016/j.neurobiolaging.2014.07.028
- Ongen, H., Brown, A. A., Delaneau, O., Panousis, N. I., Nica, A. C., and Dermitzakis, E. T. (2017). Estimating the Causal Tissues for Complex Traits and Diseases. *Nat. Genet.* 49, doi:10.1038/ng.3981
- Pal, K., Forcato, M., and Ferrari, F. (2019). Hi-C Analysis: from Data Generation to Integration. *Biophys. Rev.* 11, 67–78. doi:10.1007/s12551-018-0489-1
- Potashkin, J., Huang, X., Becker, C., Chen, H., Foltynie, T., and Marras, C. (2020). Understanding the Links between Cardiovascular Disease and Parkinson's Disease. *Mov. Disord.* 35, 55–74. doi:10.1002/mds.27836
- R Core Team (2014). R: A Language and Environment for Statistical Computing. *R. Found. Stat. Comput.* 739, 1–2630.
- Ramani, V., Cusanovich, D. A., Hause, R. J., Ma, W., Qiu, R., Deng, X., et al. (2016). Mapping Three-Dimensional Genome Architecture through *In Situ* DNase Hi-C. *Nat. Protoc.* 11, 2104–2121. doi:10.1038/nprot.2016.126
- Reynolds, R. H., Botía, J., Nalls, M. A., Noyce, A. J., Nicolas, A., Cookson, M. R., et al. (2019). Moving beyond Neurons: the Role of Cell Type-specific Gene Regulation in Parkinson's Disease Heritability. *Npj Park. Dis.* 5. doi:10.1038/s41513-019-0076-6
- Riboldi, G. M., and Di Fonzo, A. B. (2019). GBA, Gaucher Disease, and Parkinson's Disease: from Genetic to Clinic to New Therapeutic Approaches. *Cells* 8, 364. doi:10.3390/cells8040364
- Riou, A., Houvenaghel, J.-F., Dondaine, T., Drapier, S., Sauleau, P., Drapier, D., et al. (2021). Functional Role of the Cerebellum in Parkinson Disease: A PET Study. *Neurology* 96, e2874–e2884. doi:10.1212/WNL.0000000000012036
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* 12, 77. doi:10.1186/1471-2105-12-77
- Ryan, D. P., and Matthews, J. M. (2005). Protein-protein Interactions in Human Disease. *Curr. Opin. Struct. Biol.* 15, 441–446. doi:10.1016/j.sbi.2005.06.001
- Schapira, A. H. V., Chaudhuri, K. R., and Jenner, P. (2017). Non-motor Features of Parkinson Disease. *Nat. Rev. Neurosci.* 18, 435–450. doi:10.1038/nrn.2017.62
- Schierding, W., Farrow, S., Fadason, T., Graham, O. E. E., Pitcher, T. L., Qubisi, S., et al. (2020). Common Variants Coregulate Expression of GBA and Modifier Genes to Delay Parkinson's Disease Onset. *Mov. Disord.* 35, 1346–1356. doi:10.1002/mds.28144
- Scorza, F. A., Fiorini, A. C., Scorza, C. A., and Finsterer, J. (2018). Cardiac Abnormalities in Parkinson's Disease and Parkinsonism. *J. Clin. Neurosci.* 53, 1–5. doi:10.1016/j.jocn.2018.04.031
- Seidel, K., Bouzrou, M., Heidemann, N., Krüger, R., Schöls, L., den Dunnen, W. F. A., et al. (2017). Involvement of the Cerebellum in Parkinson Disease and Dementia with Lewy Bodies. *Ann. Neurol.* 81, 898–903. doi:10.1002/ana.24937
- Siddiqui, I. J., Pervaiz, N., and Abbasi, A. A. (2016). The Parkinson Disease Gene SNCA: Evolutionary and Structural Insights with Pathological Implication. *Sci. Rep.* 6, 1–11. doi:10.1038/srep24475
- Siitonen, A., Nalls, M. A., Hernández, D., Gibbs, J. R., Ding, J., Ylikotila, P., et al. (2017). Genetics of Early-Onset Parkinson's Disease in Finland: Exome Sequencing and Genome-wide Association Study. *Neurobiol. Aging* 53, 195–e7. doi:10.1016/j.neurobiolaging.2017.01.019
- Spencer, C. C. A., Plagnol, V., Strange, A., Gardner, M., Paisan-Ruiz, C., Band, G., et al. (2011). Dissection of the Genetics of Parkinson's Disease Identifies an Additional Association 5' of SNCA and Multiple Associated Haplotypes at 17q21. *Hum. Mol. Genet.* 20, 345–353. doi:10.1093/hmg/ddq469
- Stöllberger, C., Schneider, B., and Finsterer, J. (2003). Elimination of the Left Atrial Appendage to Prevent Stroke or Embolism?: Anatomic, Physiologic, and Pathophysiologic Considerations. *Chest* 124, 2356–2362. doi:10.1378/chest.124.6.2356
- Teune, L. K., Renken, R. J., De Jong, B. M., Willemsen, A. T., Van Osch, M. J., Roerdink, J. B. T. M., et al. (2014). Parkinson's Disease-Related Perfusion and Glucose Metabolic Brain Patterns Identified with PCASL-MRI and FDG-PET Imaging. *Neuroimage Clin.* 5, 240–244. doi:10.1016/j.nicl.2014.06.007
- Tobin, J. E., Latourelle, J. C., Lew, M. F., Klein, C., Suchowersky, O., Shill, H. A., et al. (2008). Haplotypes and Gene Expression Implicate the MAPT Region for Parkinson Disease: The GenePD Study. *Neurology* 71, 28–34. doi:10.1212/01.wnl.0000304051.01650.23
- Turagam, M. K., Vuddanda, V., Verberkmoes, N., Ohtsuka, T., Akca, F., Atkins, D., et al. (2018). Epicardial Left Atrial Appendage Exclusion Reduces Blood Pressure in Patients with Atrial Fibrillation and Hypertension. *J. Am. Coll. Cardiol.* 72, 1346–1353. doi:10.1016/j.jacc.2018.06.066
- Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012). Five Years of GWAS Discovery. *Am. J. Hum. Genet.* 90, doi:10.1016/j.ajhg.2011.11.029
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* 101, 5–22. doi:10.1016/j.ajhg.2017.06.005
- Volkov, P., Olsson, A. H., Gillberg, L., Jørgensen, S. W., Brøns, C., Eriksson, K.-F., et al. (2016). A Genome-wide mQTL Analysis in Human Adipose Tissue Identifies Genetic Variants Associated with DNA Methylation, Gene Expression and Metabolic Traits. *PLoS One* 11, e0157776. doi:10.1371/journal.pone.0157776
- Wider, C., Vilariño-Güell, C., Jasinska-Myga, B., Heckman, M. G., Soto-Ortolaza, A. I., Cobb, S. A., et al. (2010). Association of the MAPT Locus with Parkinson's Disease. *Eur. J. Neurol.* 17, 483–486. doi:10.1111/j.1468-1331.2009.02847.x
- Wu, T., and Hallett, M. (2013). The Cerebellum in Parkinson's Disease. *Brain* 136, 696–709. doi:10.1093/brain/awt360
- Yousefi, F., Shabaninejad, Z., Vakili, S., Derakhshan, M., Movahedpour, A., Dabiri, H., et al. (2020). TGF- $\beta$  and WNT Signaling Pathways in Cardiac Fibrosis: Non-coding RNAs Come into Focus. *Cell Commun. Signal.* 18, 1–16. doi:10.1186/s12964-020-00555-4
- Yu, J., Hu, M., and Li, C. (2019). Joint Analyses of Multi-Tissue Hi-C and eQTL Data Demonstrate Close Spatial Proximity between eQTLs and Their Target Genes. *BMC Genet.* 20, 43. doi:10.1186/s12863-019-0744-x

- Zhu, W., Trivedi, C. M., Zhou, D., Yuan, L., Lu, M. M., and Epstein, J. A. (2009). Inpp5f Is a Polyphosphoinositide Phosphatase that Regulates Cardiac Hypertrophic Responsiveness. *Circ. Res.* 105, 1240–1247. doi:10.1161/circresaha.109.208785
- Zou, Y., Stagi, M., Wang, X., Yigitkanli, K., Siegel, C. S., Nakatsu, F., et al. (2015). Gene-silencing Screen for Mammalian Axon Regeneration Identifies Inpp5f (Sac2) as an Endogenous Suppressor of Repair after Spinal Cord Injury. *J. Neurosci.* 35, 10429–10439. doi:10.1523/jneurosci.1718-15.2015

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ho, Schierding, Farrow, Cooper, Kempa-Liehr and O'Sullivan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.