

## RESEARCH ARTICLE

# The Parkinson's Disease Genome-Wide Association Study Locus Browser

Francis P. Grenn, BS,<sup>1</sup> Jonggeol J. Kim, BA,<sup>1</sup> Mary B. Makarious, BS,<sup>1</sup> Hiroataka Iwaki, MD, PhD,<sup>1,2</sup> Anastasia Illarionova,<sup>3</sup> Kajsa Brolin,<sup>4</sup> Jillian H. Kluss,<sup>1</sup> Artur F. Schumacher-Schuh, MD, PhD,<sup>5</sup> Hampton Leonard, MSc,<sup>1,2</sup> Faraz Faghri, PhD,<sup>1,2</sup> Kimberley Billingsley, PhD,<sup>1</sup> Lynne Krohn, MSc,<sup>6</sup> Ashley Hall, MRes,<sup>7</sup> Monica Diez-Fairen,<sup>8</sup> Maria Teresa Perrián,<sup>9</sup> Jia Nee Foo, PhD,<sup>10,11</sup> Cynthia Sandor, PhD,<sup>12</sup> Caleb Webber, PhD,<sup>12</sup> Brian K. Fiske, PhD,<sup>13</sup> J. Raphael Gibbs, PhD,<sup>1</sup> Mike A. Nalls, PhD,<sup>1,2</sup> Andrew B. Singleton, PhD,<sup>1</sup> Sara Bandres-Ciga, PhD,<sup>1</sup> Xylena Reed, PhD,<sup>1</sup> and Cornelis Blauwendraat, PhD,<sup>1\*</sup> on behalf of the International Parkinson's Disease Genomics Consortium (IPDGC)

<sup>1</sup>Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA

<sup>2</sup>Data Tecnica International, Glen Echo, Maryland, USA

<sup>3</sup>German Center for Neurodegenerative Diseases, Tubingen, Germany

<sup>4</sup>Lund University, Translational Neurogenetics Unit, Department of Experimental Medical Science, Lund, Sweden

<sup>5</sup>Universidade Federal do Rio Grande do Sul, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

<sup>6</sup>Department of Human Genetics, McGill University, Montreal, Quebec, Canada

<sup>7</sup>Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Liverpool, UK

<sup>8</sup>Fundació Docència i Recerca Mútua Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mútua Terrassa, Barcelona, Spain

<sup>9</sup>Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain

<sup>10</sup>Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore

<sup>11</sup>Human Genetics, Genome Institute of Singapore, A\*STAR, Singapore, Singapore

<sup>12</sup>UK Dementia Research Institute, Cardiff University, Cardiff, UK

<sup>13</sup>The Michael J. Fox Foundation for Parkinson's Research, Grand Central Station, New York, NY, USA

**ABSTRACT: Background:** Parkinson's disease (PD) is a neurodegenerative disease with an often complex component identifiable by genome-wide association studies. The most recent large-scale PD genome-wide association studies have identified more than 90 independent risk variants for PD risk and progression across more than 80 genomic regions. One major challenge in current genomics is the identification of the causal gene(s) and variant(s) at each genome-wide association study locus. The objective of the current study was to create a tool that would display data for relevant PD risk loci and provide guidance with the prioritization of causal genes and potential mechanisms at each locus.

**Methods:** We included all significant genome-wide signals from multiple recent PD genome-wide association studies including the most recent PD risk genome-wide association study, age-at-onset genome-wide association study, progression genome-wide association study, and Asian population PD risk genome-wide association study. We gathered data for all genes 1 Mb up and downstream of each variant to allow users to assess which gene(s) are most associated with the variant of interest based on a set of self-ranked criteria. Multiple databases were queried for each gene to collect additional causal data.

**Results:** We created a PD genome-wide association study browser tool (<https://pdgenetics.shinyapps.io/>)

[Corrections updated on Sep 7, 2020, after first online publication: affiliation updated.]

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

\*Correspondence to: Dr. Cornelis Blauwendraat, Laboratory of Neurogenetics, NIA, NIH, Building 35, 35 Convent Drive, Bethesda, MD 20892, USA; E-mail: [cornelis.blauwendraat@nih.gov](mailto:cornelis.blauwendraat@nih.gov)

**Relevant conflicts of interest/financial disclosures:** Dr. Nalls reported receiving support from a consulting contract between Data Tecnica International and the National Institute on Aging (NIA), National

Institutes of Health (NIH) and consulting for the Michael J. Fox Foundation, Vivid Genomics, Lysosomal Therapeutics Inc., and Neuron23, Inc, among others. No other disclosures are reported.

**Funding agencies:** For a complete overview of members, acknowledgements, and funding, please see the Supplemental data and/or <http://pdgenetics.org/partners>.

**Received:** 3 April 2020; **Revised:** 27 May 2020; **Accepted:** 10 June 2020

**Published online 31 August 2020 in Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com)). DOI: 10.1002/mds.28197**

GWASBrowser/) to assist the PD research community with the prioritization of genes for follow-up functional studies to identify potential therapeutic targets.

**Conclusions:** Our PD genome-wide association study browser tool provides users with a useful method of identifying potential causal genes at all known PD risk loci from large-scale PD genome-wide association studies. We plan to update this tool with new relevant data as

sample sizes increase and new PD risk loci are discovered. © 2020 The Authors. *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society. This article has been contributed to by US Government employees and their work is in the public domain in the USA.

**Key Words:** GWAS; Parkinson's disease; prioritization

Parkinson's disease (PD) is a multifactorial disease for which both genetic and environmental risk factors play a role. In the past decade, approximately 20 genes have been associated with PD or parkinsonism in families.<sup>1</sup> More than 90 common variants have been associated with sporadic PD risk, age at onset, and progression using genome-wide association studies (GWASes).<sup>2-5</sup>

One major challenge remaining after GWAS identification of risk loci is the localization and characterization of specific causal variant(s) and gene(s) at each locus. A common misconception is that the most significant GWAS variant exerts an effect on the nearest gene (as commonly reported in GWAS articles), but this is unlikely to be the case. First, the GWAS variant is not necessarily causative on its own, but is instead likely to tag a functional region or variant in high linkage disequilibrium (LD). Second, variants in noncoding regions containing regulatory sequences may impact distant genes by altering 3-dimensional chromatin conformation, placing these genes outside the predefined LD region.<sup>6,7</sup> Several approaches can be taken when prioritizing genes for each significant variant. For PD, these approaches include using single-cell RNA-seq to determine gene expression in relevant cell populations,<sup>8</sup> transcriptome-wide association studies,<sup>9</sup> and quantitative trait loci (QTL).<sup>3</sup> Others have functionally prioritized a single locus (*SNCA*<sup>10</sup> and *TMEM175*<sup>11</sup>), but these functional single-locus experiments often do not scale up to loci with many genes. Other non-disease-specific pipelines have been developed using epigenetic and chromatin conformation data sets in addition to expression QTL data<sup>12,13</sup>; however, some disease-specific interpretation is required. Therefore, we have aggregated multiple data sets from several sources to create a versatile and user-friendly tool (<https://pdgenetics.shinyapps.io/GWASBrowser/>) to prioritize specific genes and variants for additional PD GWAS and functional studies that aim to identify potential therapeutic targets (Fig. 1).

## Methods

### GWAS Loci and Gene Selection

All GWAS loci and summary statistics were gathered from our recent PD GWAS studies.<sup>2-5,14,15</sup> We selected all genes 1 megabase (Mb) up and downstream of each

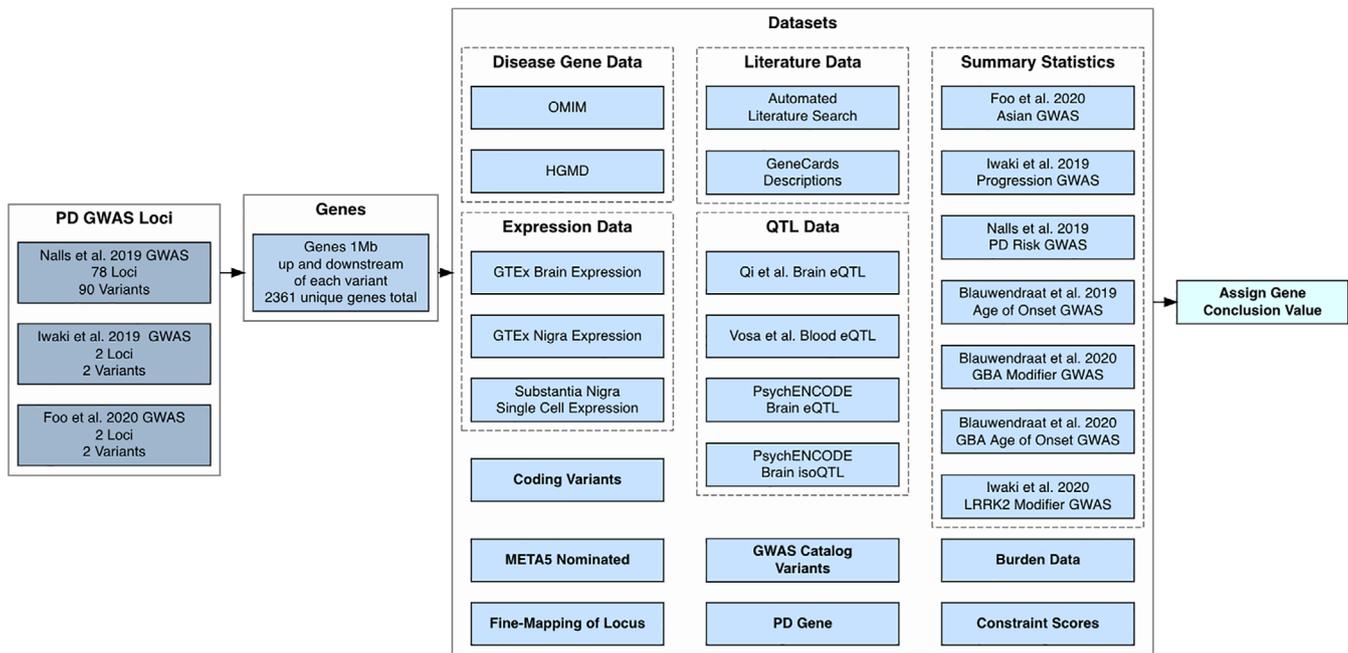
significant variant from the hg19 reference genome.<sup>16</sup> Genes in this 2-Mb range were included in locus zoom plots created for each variant.<sup>17</sup>

### Gene Expression Data

The Genotype-Tissue Expression (GTEx) portal was accessed on February 12, 2020, to obtain v8 gene expression data. Transcript per million (TPM) data was averaged across all available brain and substantia nigra (SN) samples individually for each gene. Average single-cell RNA sequencing expression data were also included for SN astrocytes, SN dopaminergic neurons, SN endothelial cells, SN GABAergic cells, SN microglial cells, SN oligodendrocyte cells, and SN oligodendrocyte progenitor cells.<sup>18</sup> An arbitrary value of 5 TPM was chosen as the cut-off for significance in brain tissue, SN tissue, and SN dopaminergic neuron averages. Genes with greater than 5 TPM in any of these 3 data sets were given a value of 1 in the evidence table in the Brain Expression, Nigra Expression, or SN-Dop. Neuron Expression columns. Genes with no available expression data were set to NA in these columns.

### Expression Quantitative Trait Loci Data

Expression quantitative trait locus (eQTL) data were collected from the summary-data-based Mendelian randomization (SMR) website for brain tissues<sup>19</sup> and processed using SMR software tools.<sup>20</sup> Blood tissue eQTL data were collected from the eQTLGen consortium.<sup>21</sup> Additional brain tissue data were collected from the PsychENCODE project.<sup>22</sup> Locus compare plots were generated using GWAS data and the appropriate eQTL data to compare the distribution of eQTL and GWAS data.<sup>23</sup> Plots were omitted if they contained neither the locus risk variant nor a good proxy variant. Proxy variants were obtained using the LDlinkR library<sup>24,25</sup> in R (<https://www.r-project.org/>) to find variants with  $r^2 > 0.7$  with the risk variant. Genes were given a value of 1 in the "QTL-brain" column in the evidence table if there were sufficient data to create brain locus compare plots using Qi et al eQTL data, PsychENCODE eQTL data, or PsychENCODE isoQTL data. Genes were given a value of 1 in the "QTL-blood" column in the evidence table if there were sufficient data in the blood tissue eQTL data to create a locus compare plot. Pearson correlation



**FIG. 1.** Flowchart of data gathered for the browser. Summary of the variants, genes, and data sets included in the browser to prioritize genes for each locus. Data sets include genome-wide association study (GWAS) summary statistics, known coding variants (Nalls et al, 2019),<sup>3</sup> nominated genes, online Mendelian inheritance in man (OMIM), and human gene mutation database (HGMD) disease genes, expression quantitative trait locus data (eQTL), variants from the GWAS catalogue, known Parkinson’s disease and related disorder genes, PubMed literature data, Genotype-Tissue Expression (GTEx), and single-cell expression data, burden test data, fine-mapping data, and constraint data.

coefficients between GWAS and blood or brain eQTL/ isoQTL *P* values were calculated using R for each gene with sufficient data. The “QTL-correl” column in the evidence table is given a value of 1 if the magnitude of the correlation coefficient is greater than the user assigned cutoff (default cutoff of 0.3) in any of the gene’s locus compare plots. This column was given a value of “NA” if the gene has no plots with a “QTL-brain” or “QTL-blood” score of 1.

### Literature Search

Automated literature searches were performed in PubMed using the search term “GENE\_NAME[Title/Abstract] AND Parkinson’s[Title/Abstract]” and “GENE\_NAME[Title/Abstract]” to search for the respective gene and its occurrence in PD literature using the rentrez package in R.<sup>26</sup> The number of search results was collected for all the genes in each locus to create a bar plot of PubMed hits for each locus. Genes with 5 or more search results in the PD and gene name search were given a score of 1 in the “Literature Search” column in the evidence table. A word cloud was generated for each gene using the PubMedWordcloud package in R.<sup>27</sup> Last, gene descriptions were obtained from GeneCards.<sup>28</sup>

### Constraint Data

Constraint data were downloaded from the gnomAD browser (<https://gnomad.broadinstitute.org/downloads>).<sup>29</sup>

These data were included to predict how resistant genes are to variation. Constraint *z* scores were included for synonymous variants and missense variants. A probability of being loss-of-function intolerant (pLI) score was included for loss-of-function variants. Observed/expected variant values were included for these 3 variant types, along with the 90% confidence interval for each value. Genes were given a score of 1 in the “Variant Intolerant” column in the evidence table if the upper limit of the 90% confidence interval was less than 0.35 for any variant type.

### Burden Data

Burden summary statistics were obtained from the most recent GWAS<sup>3</sup> and an exome sequencing study in PD (article in process). In total, 40 different burden tests on exome sequencing data and 2 different burden tests on imputed GWAS data were performed using only missense and loss-of-function variants with minor allele frequency cutoffs of 0.05 and 0.01. The minimum *P* values of all 40 exome burden tests were included for each gene in the burden table of the browser. The minimum *P* value for the 2 burden tests on imputed data for each gene was included in the same table. These 2 minimum *P* values were Bonferroni-corrected by the number of genes with data in the burden test to determine significance (1480 genes for exome and 1026 genes for imputed). Genes with significant burden

results in either exome or imputed data were given a value of 1 in the “Burden” column in the evidence table. Genes with no available burden data were given a value of “NA” in this column.

### Nalls et al 2019 Nominated Genes and Known Parkinson’s Disease Genes

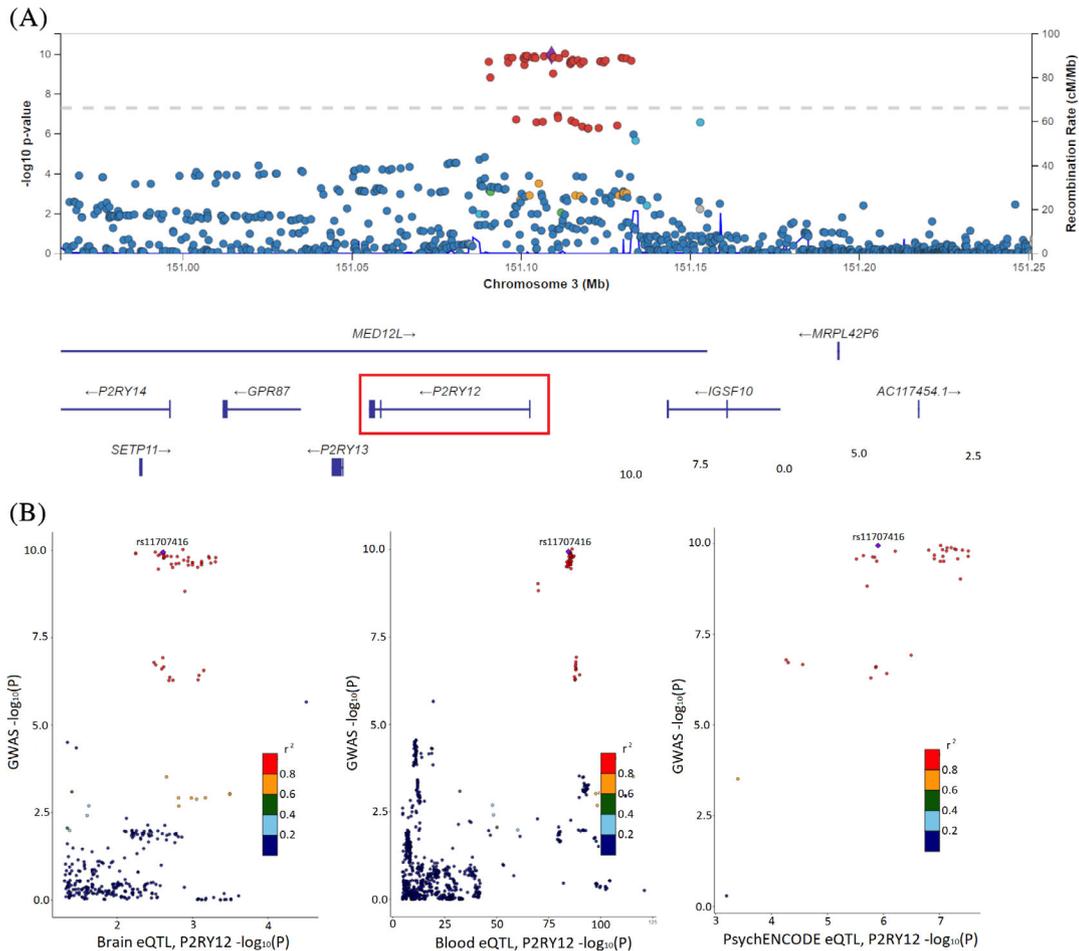
The PD risk GWAS<sup>3</sup> used 4 QTL data sets to determine causal genes for each GWAS signal in the study. These data were obtained from Supplementary Table 1 of the Nalls et al 2019 GWAS paper. Seventy of the 90 variants from this study were found to be associated with a putative causal gene. These genes were given a value of 1 in the “Nominated by META5” column of the evidence table. Genes known to be monogenic for PD, parkinsonism, or other related movement disorders were given a value of 1 in the “PD Gene” column in the evidence table.<sup>1,30</sup>

### Other Disease Genes

Disease gene data were gathered from the Human Gene Mutation Database (HGMD)<sup>31</sup> and Online Mendelian Inheritance in Man (OMIM).<sup>32</sup> For HGMD, only genes with variants classified as “DM” (disease-causing mutations) were included. These genes were given a value of 1 in the “Disease Gene” column in the evidence table.

### Coding Variants

Coding variants in linkage disequilibrium (LD) with risk variants were obtained from internal databases.  $R^2$  and  $D'$  LD scores were calculated using PLINK.<sup>33</sup> Combined annotation-dependent depletion (CADD) scores were obtained from the CADD database using ANNOVAR.<sup>34</sup> Frequencies were obtained from the gnomAD database also using ANNOVAR.<sup>29,34</sup>



**FIG. 2.** Locus zoom and locus compare plots for locus 16 and variant rs11707416 and *P2RY12*. **(A)** Locus zoom Manhattan plot for risk variant rs11707416 on locus 16. The risk variant is uniquely colored purple, and all other variants are colored by their  $r^2$  value. Recombination rate peaks are plotted in blue. Nearby genes are included at the bottom, with *P2RY12* highlighted. **(B)** Locus compare plots for *P2RY12* on locus 16 plot  $-\log_{10}P$  values for Nalls et al (2019) GWAS data (y axis) and for different eQTL data sets (x axis). Data sets available for *P2RY12* are Qi et al brain eQTL, Vosa et al blood eQTL, and PsychENCODE brain eQTL (left to right). Variants are colored by their  $r^2$  value, and the risk variant is labeled and uniquely colored purple.

## Associated Variant Phenotypes

Phenotypes of variants in LD with risk variants were obtained from the GWAS catalog v1.0.2.<sup>35</sup>  $R^2$  and  $D'$  LD scores were calculated using PLINK (v1.9)<sup>36</sup> from a large PD case-control reference set including over 40,000 individuals. Frequencies were obtained from the gnomAD database using ANNOVAR.<sup>29,34</sup>

## Fine-Mapping

Variants in the PD risk meta-analysis summary statistics<sup>3</sup> were reannotated to GRCh38p7 build positions using dbSNP build 151. If a variant's dbSNP rsid was not present in dbSNP build 151, it was excluded from further analysis. The summary statistics were partitioned into risk regions or loci based on physical distance. Per chromosome, these partitions were generated iteratively by finding the variant with the smallest  $P$  value and extracting this variant and those variants within 1 Mb of it. The region of extracted proximal variants was checked against other extracted regions, and if their edges were within 100 kb, the regions were merged. These iterations continued until no variant with a GWAS  $P < 5 \times 10^{-8}$  remained within the PD meta-analysis summary statistics.

The FINEMAP tool<sup>3,37</sup> was used to fine-map these PD-risk locus regions. Finemap uses Shotgun Stochastic Search<sup>38</sup> and Bayesian Model Averaging<sup>39</sup> to identify casual configurations of risk variants. The maximum number of causal variants within the configuration per locus was based on the number of independent risk signals detectable per locus. This number of independent risk signals per locus was estimated using the stepwise model selection procedure implemented in GCTA-COJO. Both the Finemap and GCTA-COJO<sup>40,41</sup> tools require linkage disequilibrium (LD) information for their search and modeling. For this purpose, TOPMed freeze5b samples of European ancestry, available from dbGaP, were used as an LD reference panel. This panel included 16,257 samples.

## Browser Design

The GWAS locus browser is an R shiny application. Data were precompiled and loaded onto the application's server and is not obtained from external databases in real time. The GWAS locus browser is an open-source project. The code is available on our github ([https://github.com/neurogenetics/GWAS\\_locus\\_browser](https://github.com/neurogenetics/GWAS_locus_browser)).

## Results

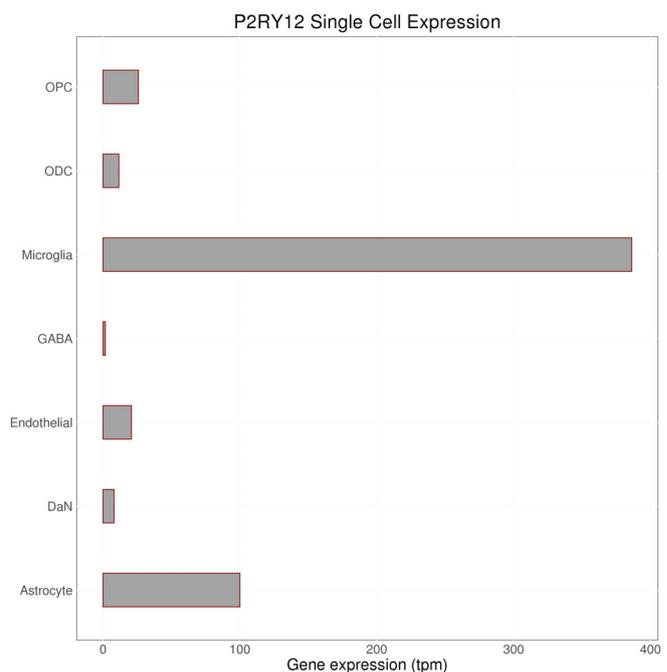
### Data Browser

The PD GWAS locus browser (<https://pdgenetics.shinyapps.io/GWASBrowser/>) is an online platform

designed to assist researchers with the prioritization of genes located within PD GWAS loci. It includes multiple layers of data, including: GWAS statistic, eQTL, burden, expression, constraint, and literature data and a flexible scoring system that users may configure for their own needs. Interestingly, several PD GWAS hits show high correlation and overlap,  $R^2 > 0.8$  and  $D' > 0.9$  (Supplementary Table 1), with other disease risk signals including diseases such as inflammatory bowel disease (locus 2),<sup>42,43</sup> neuroticism (loci 29, 62, 69, and 73),<sup>44-46</sup> body mass index (loci 4, 11, 61),<sup>47,48</sup> and insomnia (locus 73).<sup>49</sup> Below we describe 2 use case scenarios on how this application could be used for prioritizing genes from PD GWAS loci.

### Use Case Scenario 1, Locus 16, rs11707416

A possible use case for the browser exists on locus 16 for the risk variant rs11707416. This variant was discovered in the most recent PD risk GWAS (OR, 0.94; SE, 0.0097;  $P = 1.13 \times 10^{-10}$ ),<sup>3</sup> and the closest gene is *MED12L*. The locus zoom plot showed a clear GWAS signal for this region (Fig. 2A); however, no genes in this locus were prioritized using current methods in the PD risk GWAS. Using the default settings, *P2RY12* (purinergic receptor P2Y) has the highest conclusion score (7) in the evidence per gene table, meaning it has the highest sum of data set scores



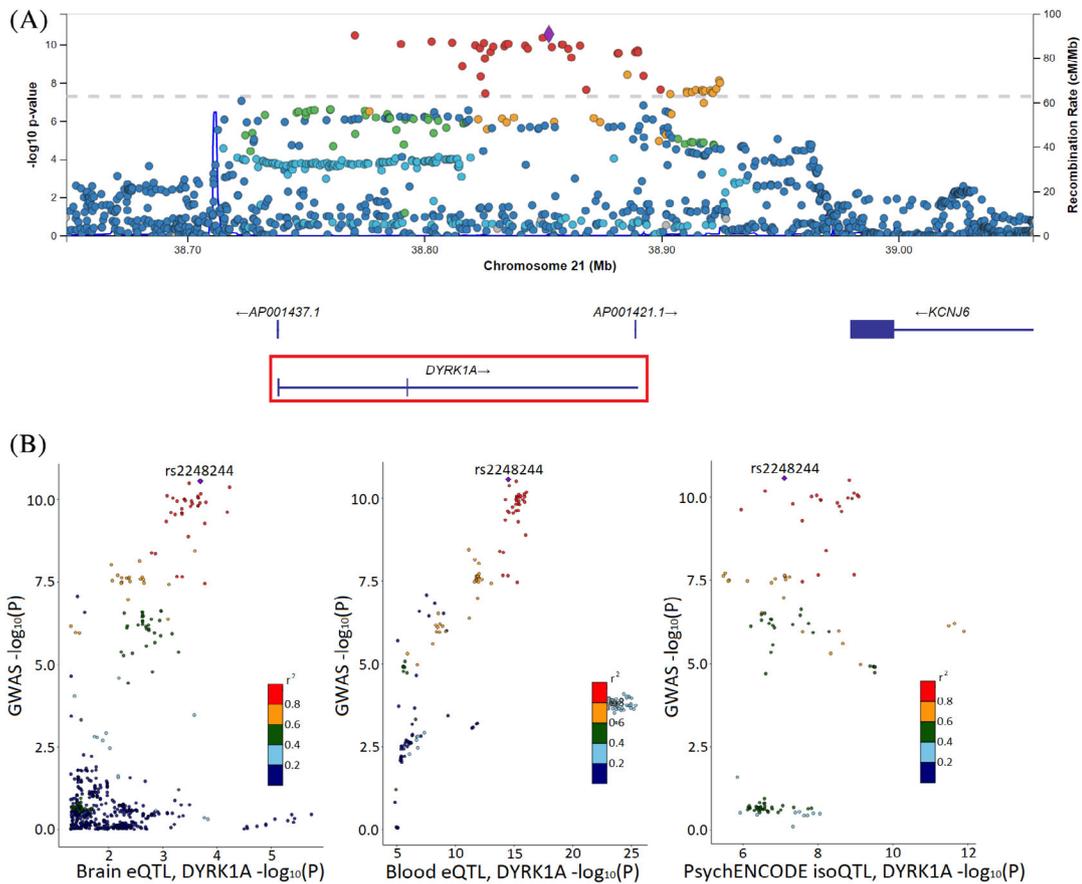
**FIG. 3.** Bar plot for single-cell expression of *P2RY12* on locus 16. Transcript per million (TPM) data for *P2RY12* was averaged across all samples for 7 cell types from the substantia nigra. These include oligodendrocyte progenitor cells (OPCs), oligodendrocyte cells (ODCs), microglia, GABAergic neurons (GABA), endothelial cells, dopaminergic neurons (DaNs), and astrocytes.

for that locus. Locus compare eQTL plots showed some correlation in both brain and blood, indicating some overlap in the distribution of eQTL and GWAS data for this gene (Fig. 2B). There is one common coding variant at this locus, but it is located within *MED12L* (mediator complex subunit 12-like; NM\_053002: exon25:c.G3629A:p.R1210Q) and not *P2RY12*. *MED12L* has a lower conclusion score (3) than *P2RY12*, suggesting that even with this coding variant, *MED12L* is not the primary candidate in this locus. This is a very complex and unusual locus, in that there are multiple genes encoded within an intron of an isoform of *MED12L* (*P2RY12*, *P2RY13*, *P2RY14*, *GPR171*), and even more interesting is that the missense variant that changes an amino acid of *MED12L* is located in an intron of the much smaller gene, *P2RY12* (Fig. 2A). Despite the uniqueness of the locus structure, we will focus on *P2RY12* as the primary candidate nominated by the PD GWAS browser data sets.

Expression of *P2RY12* was significant in all included databases (GTEx brain, GTEx SN, and

single-cell dopaminergic neuron data). However, it appears that *P2RY12* has much higher expression in astrocytes and microglia than neurons (Fig. 3). The disease gene section shows that *P2RY12* has been linked to platelet type 8 bleeding disorder.<sup>50</sup> This disorder appears to be caused by dominant-negative mutations in the *P2RY12* gene that disrupt the homodimerization of the receptor that is required for normal function.<sup>51</sup> No direct links between *P2RY12* and PD have been reported in the literature, but *P2RY12* is a widely studied gene with roles suggested in neuroinflammation, apoptosis, and autophagy, pathways that are relevant to PD and other neurodegenerative diseases.<sup>52-54</sup> Experiments have been done to characterize expression patterns of *P2RY12* in microglia and its role in neuroinflammation.<sup>55</sup> As these experiments focused on Alzheimer's disease, it would be useful to build on them in the PD context in the future.

However, this is not conclusive evidence, and other potential candidates exist on this locus. For example,

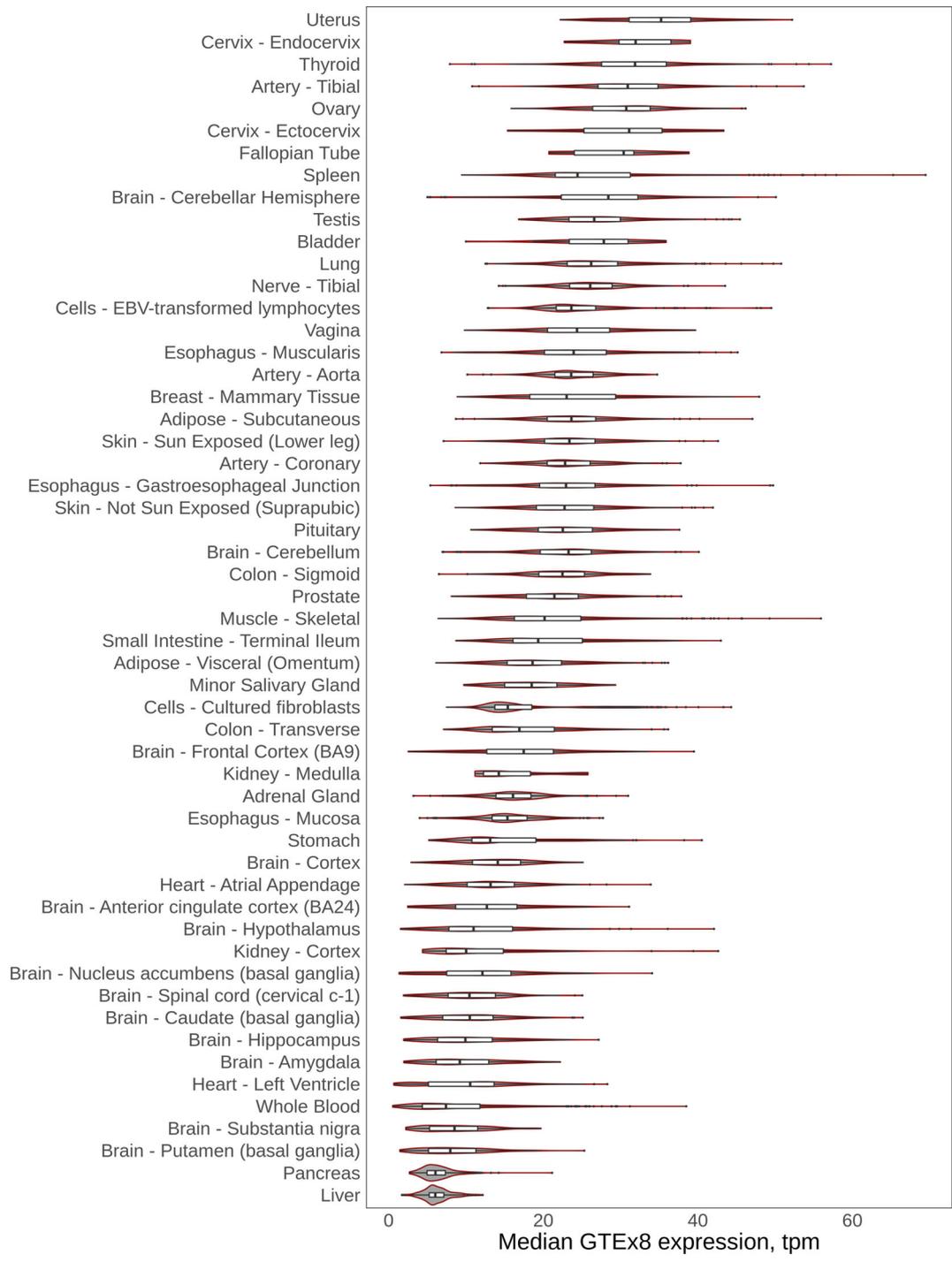


**FIG. 4.** Locus zoom and locus compare plots for locus 78 and variant rs2248244 and *DYRK1A*. **(A)** Locus zoom Manhattan plot for risk variant rs2248244 at locus 78. The risk variant is uniquely colored purple, and all other variants are colored by their  $r^2$  value. Recombination rate peaks are plotted in blue. Nearby genes are included at the bottom, with *DYRK1A* highlighted. **(B)** Locus compare plots for *DYRK1A* on locus 78 plot  $-\log_{10}P$  values for Nalls et al (2019) GWAS data (y axis) and for different eQTL data sets (x axis). Data sets available for *DYRK1A* are Qi et al brain eQTL, Vosa et al blood eQTL, and PsychENCODE brain isoQTL (left to right). Variants are colored by their  $r^2$  value, and the risk variant is labeled and uniquely colored purple.

*SLENOT* has a low conclusion score (2), partly because of lack of data for the gene, but is suggested to play a protective role in PD in the literature.<sup>56</sup> Overall, the possible role of *P2RY12* in PD should be further analyzed with functional experiments.

**Use Case Scenario 2: Locus 78, rs2248244**

We chose a second use case for the browser at locus 78 for the risk variant rs2248244. This variant was discovered in the most recent PD risk GWAS (OR, 1.074; SE, 0.0107;  $P = 2.74 \times 10^{-11}$ ),<sup>3</sup> and its nearest gene,



**FIG. 5.** GTEx violin plot for *DYRK1A* on locus 78. *DYRK1A* transcript per million (TPM) data were averaged across all available samples from GTEx v8 data for different tissues. Distribution and probability of expression level are included for each tissue.

*DYRK1A*, was nominated as the causal gene in that study. The locus zoom plot showed a clear GWAS signal for this region (Fig. 4A). The default settings in the evidence per gene table nominated *DYRK1A* as the top candidate with a conclusion score of 10, which is higher than any of other genes within the locus (second highest is 4).

The brain and blood eQTL plots and the isoQTL plot showed good correlation between GWAS and QTL values (Fig. 4B). No coding variants or other known associated disease variants exist for this locus. *DYRK1A* showed significant gene expression in all databases (GTEx brain, GTEx SN, and single-cell dopaminergic neuron data; Fig. 5). Constraint data for *DYRK1A* showed a low 90% CI for loss-of-function variation and a pLI of 1, suggesting significant intolerance to loss-of-function variation. However, burden test results showed no significant change in variants for *DYRK1A* after Bonferroni correction. FINEMAP results of this locus nominated several variants with rs2248244 and rs11701722, both intronic with the highest probability score.

*DYRK1A* is located in the Down syndrome critical region on chromosome 21, and various deletions and single-nucleotide variants have been linked to autosomal-dominant mental retardation-7 (MRD7).<sup>57-59</sup> These data are not enough to conclude that *DYRK1A* is the relevant risk gene at this locus. However, *DYRK1A* also has significant representation in the PD literature. Previous studies have suggested that *DYRK1A* encodes a kinase that can phosphorylate  $\alpha$ -synuclein and *Pink1* in mammalian cells.<sup>60-63</sup> Studies in mouse models have indicated that haploinsufficiency of *DYRK1A* leads to a reduction in dopaminergic neurons while increasing the dosage results in more dopaminergic neurons by altering apoptosis.<sup>60-62</sup> All these studies combined indicate that loss of *DYRK1A* expression may influence the number of dopaminergic neurons and thereby the development of PD.

Again, this is not conclusive, and other potential candidates exist at this locus. For example, *TTC3* has significant expression data and is suggested to play a role in Alzheimer's disease<sup>64</sup> and Down syndrome<sup>65</sup> in the literature. Overall, the PD GWAS browser predicts *DYRK1A* is a good candidate for more functional PD experiments and high-powered human genetic studies aimed at characterizing the molecular mechanism underlying risk at this locus.

## Discussion

GWASes have identified numerous risk loci for a large number of diseases (<https://www.ebi.ac.uk/gwas/>).<sup>35</sup> The

current bottleneck after completing these studies is identifying the causal genes and variants underlying the GWAS signals. There is a large discrepancy between the number of new GWAS loci identified and the number of studies that molecularly characterize these loci to identify a causative variant or gene. Currently, the number of PD loci that are functionally validated is very low and mostly includes genes that are known to cause monogenic forms of PD. PD GWASes have identified a number of these pleiotropic genes at various loci such as *SNCA* (locus 23), *GBA* (locus 1), *LRRK2* (locus 49), and *VPS13C* (locus 59). Current evidence suggests that *TMEM175* (locus 19<sup>11</sup>), *CTSB* (locus 37<sup>2,14</sup>) and *GCH1* (locus 56<sup>66</sup>) are also pleiotropic and lead to PD by multiple mechanisms.

Our goal is to provide the PD research community with a tool that catalogs all significant PD GWAS signals and helps to prioritize the genes at each locus for functional studies. The overall significance score for each gene is displayed in the "Conclusion" column of the evidence table. This score is the sum of all other numerical values from each data set for the respective gene. The inclusion of genes 1 Mb up and downstream of each variant is an arbitrary cutoff and may limit the ability of the browser to accurately prioritize genes. This limitation will come into effect when variants affect genes outside this range; however, prior studies have found that functional noncoding variants are often located within this 2 Mb window. In addition, some loci contain more than 1 significant variant (eg, locus 1, *GBA* with 3 reported independent signals), and until a way to detect precise causative variants is developed, it is assumed that all significant variants within the same locus will impact the same gene.

Although all the data sets included in the PD browser may contribute to the prioritization of PD genes, each data set comes with its own limitations that should be taken into account when considering the conclusion value. Blood and brain eQTL data were included to identify genes with similar GWAS and eQTL data distribution. Although similar distributions may suggest causality, the power of these data sets may reduce their importance. Blood eQTL data will have more statistical power than brain data because of their larger sample size, but may be less relevant to PD. Another important note is that we just tested for cis eQTLs because we simply did not have the power to detect robust trans-QTLs. Scoring for the "QTL-brain" and "QTL-blood" columns is not indicative of the causality of a gene because it relies on the existence of eQTL data, not the information provided by the data. Although the "QTL-correl" column gives more insight into the eQTL data, the default Pearson correlation coefficient cutoff significance values of  $-0.3$  and  $0.3$  are quite broad, so we

have included an option to allow users to modify this cutoff. In addition, eQTL data are not available for all genes of interest. For these reasons, the correlation between eQTL and GWAS data does not guarantee causality for a gene, but is still good evidence for prioritization. Gene expression data were included to account for possible increased expression of genes associated with GWAS variants in relevant cell types and tissues. GTEx transcript per million (TPM) v8 data was used to measure gene expression. We focused on brain tissue, SN tissue, and SN dopaminergic neuron data because of their established role in PD and enrichment in PD GWAS loci.<sup>3</sup> However, genes do not necessarily need to be expressed in the substantia nigra or other brain tissues and cell types to increase the risk of PD.

Constraint data were included to identify genes that are intolerant to specific types of variation, suggesting conservation. Therefore, we suspect genes with significant intolerance may be causal because normal variation in these genes is unlikely. We used a significance cutoff of 0.35 for the upper limit of the 90% confidence interval of the observed/expected values, as suggested by gnomAD.<sup>29</sup> It is possible that variations within genes may not be associated with PD, suggesting that low intolerance/constraint scores are not necessarily causal. Previously published burden test results were included to account for genes thought to have a significant burden of rare variants in PD.<sup>3</sup> However, this does not guarantee causality because causal genes may be tagged by common variants instead. The literature count was included to quickly measure the significance of genes in previously published research. However, existing studies can easily be biased, and our automated search of the literature does not account for this. In addition, some genes are difficult to identify in automated literature searches because of nomenclature changing over time or their similarity to common names. Examples of these name complications include *SHE* and *MAL* (sometimes used for “mean axonal length” instead of the gene encoding “myelin and lymphocyte protein”).

Overall, for approximately half of all PD GWAS loci, an educated guess can be made based on the data included in our browser to choose the most likely candidate gene underlying the GWAS risk signal. Some loci have no obvious candidate (eg, loci 34, 50, and 74), but our browser may still help to prioritize candidate genes with the help of additional specific reference data. It should also be noted that there is not a “one-fits-all” efficient scoring system for the prioritization of genes under GWAS peaks. In addition, each included data set has clear limitations as discussed above. Therefore, we included a weight option for each data column in the evidence table. This will allow users to assign weights (0–4, with 0 for no points and 4 for 4 points) to different columns to alter the significance of the data in the

final conclusion summation based on the data sets they deem most important.

In summary, we have presented an online platform that allows for prioritization of genes within PD GWAS loci. We have highlighted 2 examples (*P2RY12* and *DYRK1A*), but many other interesting gene candidates can be identified using this application. The platform is designed to be versatile, flexible, and easily expandable when more loci or data sets of interest become available. By using this platform, GWAS follow-up studies can systematically prioritize genes based on publicly available data sets that may help to improve the design of functional experiments. In turn, this workflow could help to nominate these genes as potential therapeutic targets worthy of translating to the clinic. ■

**Acknowledgments:** We thank all members of the International Parkinson Disease Genomics Consortium (IPDGC). For a complete overview of members, acknowledgements, and funding, please see the Supplemental data and/or <http://pdgenetics.org/partners>. This work was supported in part by the Intramural Research Programs of the National Institute of Neurological Disorders and Stroke (NINDS), the National Institute on Aging (NIA), and the National Institute of Environmental Health Sciences, both part of the National Institutes of Health, Department of Health and Human Services (project numbers 1ZIA-NS003154, Z01-AG000949-02, and Z01-ES101986). We thank the research participants and employees of 23andMe for making this work possible. C.W. is supported by the UK Dementia Research Institute funded by the Medical Research Council (MRC), Alzheimer’s Society and Alzheimer’s Research UK. C.S. is supported by the Ser Cymru II program, which is partly funded by Cardiff University and the European Regional Development Fund through the Welsh Government. Data were generated as part of the PsychENCODE Consortium supported by: U01MH103339, U01MH103365, U01MH103392, U01MH103340, U01MH103346, R01MH105472, R01MH094714, R01MH105898, R21MH102791, R21MH105881, R21MH103877, and P50MH106934 awarded to Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (USC), Mark Gerstein (Yale), Daniel Geschwind (UCLA), Thomas M. Hyde (LIBD), Andrew Jaffe (LIBD), James A. Knowles (USC), Chunyu Liu (UIC), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Nenad Sestan (Yale), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (UCSF), Patrick Sullivan (UNC), Flora Vaccarino (Yale), Sherman Weissman (Yale), Kevin White (UChicago), and Peter Zandi (JHU). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this article were obtained from the GTEx Portal on February 12, 2020. Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung, and Blood Institute (NHLBI). Genome sequencing for “NHLBI TOPMed: Atherosclerosis Risk in Communities (ARIC)” (phs001211.v2.p2) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1) and at the Baylor Human Genome Sequencing Center (3U54HG003273-12S2, HHSN268201500015C). Genome sequencing for the “NHLBI TOPMed: Cleveland Clinic Atrial Fibrillation (CCAF) Study” (phs001189.v1.p1) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). Genome sequencing for “NHLBI TOPMed: Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Project: Cardiovascular Health Study (phs001368.v1.p1) was performed at the Baylor Human Genome Sequencing Center (3U54HG003273-12S2, HHSN268201500015C). Genome sequencing for “NHLBI TOPMed: Partners HealthCare Biobank” (phs001024.v3.p1) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). Genome sequencing for “NHLBI TOPMed: Whole Genome Sequencing of Venous Thromboembolism (WGS of VTE)” (phs001402.v1.p1) was performed at the Baylor Human Genome Sequencing Center (3U54HG003273-12S2, HHSN268201500015C). Genome sequencing for “NHLBI TOPMed: Novel Risk Factors for the Development of Atrial Fibrillation in Women” (phs001040.v3.p1) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). Genome sequencing for “NHLBI TOPMed: The Genetics and Epidemiology of Asthma in Barbados” (phs001143.v2.p1) was performed by Illumina Genomic Services (3R01HL104608-04S1).

Genome sequencing for “NHLBI TOPMed: The Vanderbilt Genetic Basis of Atrial Fibrillation” (phs001032.v4.p2) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1). Genome sequencing for “NHLBI TOPMed: Heart and Vascular Health Study (HVH)” (phs000993.v3.p2) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1) and at the Baylor Human Genome Sequencing Center (3U54HG003273-12S2, HHSN268201500015C). Genome sequencing for “NHLBI TOPMed: Genetic Epidemiology of COPD (COPDGene)” (phs000951.v3.p3) was performed at the University of Washington Northwest Genomics Center (3R01HL089856-08S1) and at the Broad Institute of MIT and Harvard (HHSN268201500014C). Genome sequencing for “NHLBI TOPMed: The Vanderbilt Atrial Fibrillation Ablation Registry” (phs000997.v3.p2) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-12S2, 3U54HG003067-13S1). Genome sequencing for “NHLBI TOPMed: The Jackson Heart Study” (phs000964.v3.p1) was performed at the University of Washington Northwest Genomics Center (HHSN268201100037C). Genome sequencing for “NHLBI TOPMed: Genetics of Cardiometabolic Health in the Amish” (phs000956.v3.p1) was performed at the Broad Institute of MIT and Harvard (3R01HL121007-01S1). Genome sequencing for “NHLBI TOPMed: Massachusetts General Hospital Atrial Fibrillation (MGH AF) Study” (phs001062.v3.p2) was performed at the Broad Institute of MIT and Harvard (3R01HL092577-06S1, 3U54HG003067-12S2, 3U54HG003067-13S1, 3UM1HG008895-01S2). Genome sequencing for “NHLBI TOPMed: The Framingham Heart Study” (phs000974.v3.p2) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-12S2). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering, were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. The Atherosclerosis Risk in Communities study has been funded in whole or in part with federal funds from the National Heart, Lung, and Blood Institute, National Institute of Health, Department of Health and Human Services, under contract numbers (HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I, and HHSN268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions. The research reported in this article was supported by grants from the National Institutes of Health (NIH) National Heart, Lung, and Blood Institute grants R01 HL090620 and R01 HL111314, the NIH National Center for Research Resources for Case Western Reserve University and Cleveland Clinic Clinical and Translational Science Award (CTSA) UL1-RR024989, the Department of Cardiovascular Medicine philanthropic research fund, Heart and Vascular Institute, Cleveland Clinic, the Fondation Leducq grant 07-CVD 03, and the Atrial Fibrillation Innovation Center, state of Ohio. This research was supported by contracts HHSN268201200036C, HHSN268200800007C, N01-HC85079, N01-HC-85080, N01-HC-85081, N01-HC-85082, N01-HC-85083, N01-HC-85084, N01-HC-85085, N01-HC-85086, N01-HC-35129, N01-HC-15103, N01-HC-55222, N01-HC-75150, N01-HC-45133, and N01-HC-85239; grant numbers U01 HL080295 and U01 HL130014 from the National Heart, Lung, and Blood Institute, and R01 AG023629 from the National Institute on Aging, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at <https://chs-nhlbi.org/pi>. This article was not prepared in collaboration with CHS investigators and does not necessarily reflect the opinions or views of CHS or the NHLBI. We thank the Broad Institute for generating high-quality sequence data supported by NHLBI grant 3R01HL092577-06S1 to Dr. Patrick Ellinor. Funded in part by grants from the National Institutes of Health, National Heart, Lung, and Blood Institute (HL66216 and HL83141), and the National Human Genome Research Institute (HG04735). The Women's Genome Health Study (WGHS) is supported by HL 043851 and HL099355 from the National Heart, Lung, and Blood Institute and CA 047988 from the National Cancer Institute, the Donald W. Reynolds Foundation with collaborative scientific support and funding for genotyping provided by Amgen. AF end-point confirmation was supported by HL-093613 and a grant from the Harris Family Foundation and Watkin's Foundation. The Genetics and Epidemiology of Asthma in Barbados is supported by National Institutes of Health (NIH) National Heart, Lung, and Blood Institute TOPMed (R01 HL104608-S1), and R01 AI20059, K23 HL076322, and RC2 HL101651. The research reported in this article was supported by grants from the American Heart Association to Dr. Darbar (EIA 0940116N), and grants from the National Institutes of Health (NIH) to Dr. Darbar (HL092217), and Dr. Roden (U19 HL65962, and UL1 RR024975). This project was also supported by a CTSA award (UL1TR000445) from the National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do

not necessarily represent official views of the National Center for Advancing Translational Sciences of the NIH. The research reported in this article was supported by grants HL068986, HL085251, HL095080, and HL073410 from the National Heart, Lung, and Blood Institute. This article was not prepared in collaboration with Heart and Vascular Health (HVH) Study investigators and does not necessarily reflect the opinions or views of the HVH Study or the NHLBI. This research used data generated by the COPDGene study, which was supported by NIH grants U01 HL089856 and U01 HL089897. The COPDGene project is also supported by the COPD Foundation through contributions made by an Industry Advisory Board composed of Pfizer, AstraZeneca, Boehringer Ingelheim, Novartis, and Sunovion. Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. This study is part of the Centers for Common Disease Genomics (CCDG) program, a large-scale genome sequencing effort to identify rare risk and protective alleles that contribute to a range of common disease phenotypes. The CCDG program is funded by the National Human Genome Research Institute (NHGRI) and the National Heart, Lung, and Blood Institute (NHLBI). Sequencing was completed at the Human Genome Sequencing Center at Baylor College of Medicine under NHGRI grant UM1 HG008898. The research reported in this article was supported by grants from the American Heart Association to Dr. Shoemaker (11CRP742009) and Dr. Darbar (EIA 0940116N), and grants from the National Institutes of Health (NIH) to Dr. Darbar (R01 HL092217) and Dr. Roden (U19 HL65962 and UL1 RR024975). The project was also supported by a CTSA award (UL1 TR000445) from the National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences or the NIH. The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I/HHSN26800001I), and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I, and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The authors also thank the staffs and participants of the JHS. The Amish studies on which these data are based were supported by NIH grants R01 AG18728, U01 HL072515, R01 HL088119, R01 HL121007, and P30 DK072488. See publication PMID: 18440328. The research reported in this article was supported by NIH grants K23HL071632, K23HL114724, R21DA027021, R01HL092577, R01HL092577S1, R01HL104156, K24HL105780, and U01HL65962. The research has also been supported by an Established Investigator Award from the American Heart Association (13EIA14220013) and by support from the Fondation Leducq (14CVD01). This article was not prepared in collaboration with MGH AF Study investigators and does not necessarily reflect the opinions or views of the MGH AF Study investigators or the NHLBI. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (contract nos. N01-HC-25195, HHSN268201500001I, and 75N92019D00031). This article was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI.

## References

1. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol* 2020;19:170–178.
2. Iwaki H, Blauwendraat C, Leonard HL, et al. Genomewide association study of Parkinson's disease clinical biomarkers in 12 longitudinal patients' cohorts. *Mov Disord* 2019;34(12):1839–1850.
3. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* 2019;18:1091–1102.
4. Blauwendraat C, Heilbron K, Vallerga CL, et al. Parkinson's disease age at onset genome-wide association study: defining heritability, genetic loci, and  $\alpha$ -synuclein mechanisms. *Mov Disord* 2019;34: 866–875.

5. Foo JN, Chew EGY, Chung SJ, et al. Identification of risk loci for Parkinson disease in Asians and comparison of risk between Asians and Europeans: a Genome-Wide Association Study. *JAMA Neurol* 2020:e200428.
6. Lettice LA, Heaney SJH, Purdie LA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 2003;12:1725–735.
7. Claussnitzer M, Dankel SN, Kim K-H, et al. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med* 2015;373:895–907.
8. Hook PW, McClymont SA, Cannon GH, et al. Single-Cell RNA-Seq of mouse dopaminergic neurons informs candidate gene selection for sporadic Parkinson disease. *Am J Hum Genet* 2018;102:427–446.
9. Li YI, Wong G, Humphrey J, Raj T. Prioritizing Parkinson's disease genes using population-scale transcriptomic data. *Nat Commun* 2019;10:994.
10. Soldner F, Stelzer Y, Shivalila CS, et al. Parkinson-associated risk variant in distal enhancer of  $\alpha$ -synuclein modulates target gene expression. *Nature* 2016;533:95–99.
11. Jinn S, Blauwendraat C, Toolan D, et al. Functionalization of the TMEM175 p.M393T variant as a risk factor for Parkinson disease. *Hum Mol Genet* 2019;28:3244–3254.
12. Peat G, Jones W, Nuhn M, et al. The open targets Post-GWAS analysis pipeline. *Bioinformatics* 2020;36(8):2936–2937.
13. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017;8:1826.
14. Blauwendraat C, Reed X, Krohn L, et al. Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. *Brain* 2020;143:234–248.
15. Iwaki H, Blauwendraat C, Makarios MB, et al. Penetrance of Parkinson's disease in LRRK2 p.G2019S carriers is modified by a polygenic risk score. *Mov Disord* 2020;35(5):774–780.
16. Haussler M, Zweig AS, Tyner C, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res* 2019;47:D853–D858.
17. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336–2337.
18. Agarwal D, Sandor C, Volpato V, et al. A human single-cell atlas of the substantia nigra reveals novel cell-specific pathways associated with the genetic risk of Parkinson's disease and neuropsychiatric disorders. *bioRxiv* 2020; doi.org/https://doi.org/10.1101/2020.04.29.067587.
19. Qi T, Wu Y, Zeng J, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun* 2018;9:2282.
20. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016;48:481–487.
21. Vösa U. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv* 2018; https://doi.org/10.1101/447367.
22. Wang D, Liu S, Warrell J, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* 2018;362. https://doi.org/10.1126/science.aat8464.
23. Liu B, Gloudemans M, Montgomery S. LocusCompare: A Tool to Visualize Pairs of Association [Internet]. 2018. http://locuscompare.com/.
24. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015;31:3555–3557.
25. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R package for rapidly calculating linkage disequilibrium statistics in diverse populations. *Front Genet* 2020. https://doi.org/10.3389/fgene.2020.00157.
26. Winter D. rentrez: An R package for the NCBI eUtils API [Internet]. *R J* 2017;520–526. https://doi.org/10.32614/rj-2017-058.
27. Fan Y. Pubmedwordcloud [Internet]. 2014. http://felixfan.github.io/PubMedWordcloud/
28. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics* 2016;54:1.30.1–1.30.33.
29. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 2019:531210. Accessed February 11, 2020.
30. Klein C, Hattori N, Marras C. MDSGene: Closing data gaps in genotype-phenotype correlations of monogenic Parkinson's disease. *J Parkinsons Dis* 2018;8:S25–S30.
31. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* 2017;136:665–677.
32. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). Online Mendelian Inheritance in Man, OMIM® [Internet]. Accessed February 15, 2020.
33. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575.
34. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
35. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47:D1005–D1012.
36. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7.
37. Benner C, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. Refining fine-mapping: effect sizes and regional heritability. *Genetics bioRxiv* 2018:R111–R119.
38. Hans C, Dobra A, West M. Shotgun stochastic search for “Largep” regression. *J Am Stat Assoc* 2007;507–156.
39. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995;773–795.
40. Yang J, Ferreira T, Morris AP, et al; Genetic Investigation of ANthropometric Traits (GIANT) Consortium. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012;44(4):369–375.
41. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76–82.
42. de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* 2017;49:256–261.
43. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119–124.
44. Luciano M, Hagenaars SP, Davies G, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat Genet* 2018;50:6–11.
45. Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat Commun* 2018;9(1):905.
46. Baselmans BML, Jansen R, Ip HF, et al. Multivariate genome-wide analyses of the well-being spectrum. *Nat Genet* 2019;51:445–451.
47. Wang H, Zhang F, Zeng J, et al. Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci Adv* 2019;5(8):eaaw3538.
48. Akiyama M, Okada Y, Kanai M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet* 2017;49:1458–1467.
49. Jansen PR, Watanabe K, Stringer S, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat Genet* 2019;51:394–403.

50. Cattaneo M, Zighetti ML, Lombardi R, et al. Molecular bases of defective signal transduction in the platelet P2Y12 receptor of a patient with congenital bleeding. *Proc Natl Acad Sci U S A* 2003;100:1978–83.
51. Mundell SJ, Rabbolini D, Gabrielli S, et al. Receptor homodimerization plays a critical role in a novel dominant negative P2RY12 variant identified in a family with severe bleeding. *J Thromb Haemost* 2018;16:44–53.
52. Pi S, Mao L, Chen J, et al. The P2RY12 receptor promotes VSMC-derived foam cell formation by inhibiting autophagy in advanced atherosclerosis. *Autophagy* [Internet]. 2020; doi.org/https://doi.org/10.1080/15548627.2020.1741202.
53. Blume ZI, Lambert JM, Lovel AG, Mitchell DM. Microglia in the developing retina couple phagocytosis with the progression of apoptosis via P2RY12 signaling. *Dev Dyn* 2020;48(6):723–740.
54. van Wageningen TA, Vlaar E, Kooij G, Jongenelen CAM, Geurts JJG, van Dam A-M. Regulation of microglial TMEM119 and P2RY12 immunoreactivity in multiple sclerosis white and grey matter lesions is dependent on their inflammatory environment. *Acta Neuropathol Commun* 2019;7:206.
55. Walker DG, Tang TM, Mendsaikhana A, et al. Patterns of Expression of purinergic receptor P2RY12, a putative marker for non-activated microglia, in aged and Alzheimer's disease brains [Internet]. *Int J Mol Sci* 2020;21(2):678.
56. Shao Z-Q, Zhang X, Fan H-H, et al. Selenoprotein T promotes proliferation and G1-to-S transition in SK-N-SH cells: implications in Parkinson's disease. *J Nutrition* 2019;149(12):2110–2119.
57. van Bon BWM, Hoischen A, Hehir-Kwa J, et al. Intragenic deletion in DYRK1A leads to mental retardation and primary microcephaly. *Clin Genet* 2011;79:296–299.
58. Courcet J-B, Faivre L, Malzac P, et al. The DYRK1A gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. *J Med Genet* 2012;49:731–736.
59. O'Roak BJ, Vives L, Fu W, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012;338:1619–1622.
60. Chiu C-C, Yeh T-H, Chen R-S, et al. Upregulated expression of MicroRNA-204-5p leads to the death of dopaminergic cells by targeting DYRK1A-mediated apoptotic signaling cascade. *Front Cell Neurosci* 2019;13:399.
61. Barallobre MJ, Perier C, Bové J, Laguna A, Delabar JM, Vila M, et al. DYRK1A promotes dopaminergic neuron survival in the developing brain and in a mouse model of Parkinson's disease [Internet]. *Cell Death Dis* 2014;e1289–e1289. <https://doi.org/10.1038/cddis.2014.253>
62. Kim EJ, Sung JY, Lee HJ, et al. Dyrk1A phosphorylates alpha-synuclein and enhances intracellular inclusion formation. *J Biol Chem* 2006;281:33250–33257.
63. Im E, Chung KC. Dyrk1A phosphorylates parkin at Ser-131 and negatively regulates its ubiquitin E3 ligase activity. *J Neurochem* 2015;134:756–768.
64. Kohli MA, Cukier HN, Hamilton-Nelson KL, et al. Segregation of a rareTTC3variant in an extended family with late-onset Alzheimer disease [Internet]. *Neurol Genet* 20162;(1):e41.
65. Berto G, Camera P, Fusco C, Imarisio S, Ambrogio C, Chiarle R, et al. The Down syndrome critical region protein TTC3 inhibits neuronal differentiation via RhoA and Citron kinase. *J Cell Sci* 2007; 120:1859–1867.
66. Mencacci NE, Isaia IU, Reich MM, et al. Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. *Brain* 2014;137: 2480–2492.

## Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.