

Genome-wide contribution of common Short-Tandem Repeats to Parkinson's Disease genetic risk

Bernabe I. Bustos^{1*}, Kimberley Billingsley^{2*}, Cornelis Blauwendraat², J. Raphael Gibbs³, Ziv Gan-Or^{4,5,6}, Dimitri Krainc¹, Andrew B. Singleton^{2†} and Steven J. Lubbe^{1†}. For the International Parkinson's Disease Genomics Consortium (IPDGC).

1) Ken and Ruth Davee Department of Neurology and Simpson Querrey Center for Neurogenetics, Northwestern University, Feinberg School of Medicine, Chicago, IL 60611, USA.

2) Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA

3) Computational Biology Group, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA.

4) The Neuro (Montreal Neurological Institute-Hospital), McGill University, Montréal, QC, Canada

5) Department of Human Genetics, McGill University, Montréal, QC, Canada

6) Department of Neurology and neurosurgery, McGill University, Montréal, QC, Canada

*These authors contributed equally to this work

†Correspondence to: Steven J. Lubbe (steven.lubbe@northwestern.edu) and Andrew. B. Singleton (singleta@mail.nih.gov)

ABSTRACT

Parkinson's disease (PD) is a complex neurodegenerative disorder with a strong genetic component, where most known disease-associated variants are single nucleotide polymorphisms (SNPs) and small insertions and deletions (Indels). DNA repetitive elements account for >50% of the human genome, however little is known of their contribution to PD etiology. While select short tandem repeats (STRs) within candidate genes have been studied in PD, their genome-wide contribution remains unknown. Here we present the first genome-wide association study (GWAS) of STRs in PD. Through a meta-analysis of 16 imputed GWAS cohorts from the International Parkinson's Disease Genomic Consortium (IPDGC), totalling 39,087 individuals (16,642 PD cases and 22,445 controls of European ancestry) we identified 34 genome-wide significant STR loci ($p < 5.34 \times 10^{-6}$), with the strongest signal located in *KANSL1* (chr17:44205351:[T]₁₁, $p=3 \times 10^{-39}$, OR=1.31 [CI 95%=1.26-1.36]). Conditional-joint analyses suggested that 4 significant STRs mapping nearby *NDUFAF2*, *TRIML2*, *MIRNA-129-1* and *NCOR1* were independent from known PD risk SNPs. Including STRs in heritability estimates increased the variance explained by SNPs alone. Gene expression analysis of STRs (eSTR) in RNASeq data from 13 brain regions, identified significant associations of STRs influencing the expression of multiple genes, including PD known genes. Further functional annotation of candidate STRs revealed that significant eSTRs within *NDUFAF2* and *ZSWIM7* overlap with regulatory features and are associated with change in the expression levels of nearby genes. Here we show that STRs at known and novel candidate PD loci contribute to PD risk, and have functional effects in disease-relevant tissues and pathways, supporting previously reported disease-associated genes and giving further evidence for their functional prioritization. These data represent a valuable resource for researchers currently dissecting PD risk loci.

INTRODUCTION

Parkinson's disease (PD) is a complex neurodegenerative disease with an established genetic component. Studies over the years have identified several rare variants that cause or significantly increase the risk of disease in carriers, and genome-wide association studies (GWAS) have recently uncovered 90 common variants that influence PD risk¹. It is estimated that common GWAS variants account for 16-36% of the overall genetic heritability of PD^{1,2} highlighting that a large proportion of the missing heritability remains to be identified.

The vast majority of PD genetics studies have focused on the role of single nucleotide polymorphisms (SNPs), meaning that contributions of other genetic elements such as structural variants and repetitive elements have largely been ignored. Repetitive elements represent more than 55% of the human genome³. Short tandem repeat expansions (STRs), are small repetitive units ranging from 1 to 7 base pairs in length that vary among individuals, and account for ~10% of all repetitive elements⁴. STRs are the cause of several neurological diseases and are associated to genes such as in Fragile X syndrome (*FMR-1*)^{5,6}, Huntington's disease (*HTT*)⁷, amyotrophic lateral sclerosis and frontotemporal dementia (*C9ORF72*)^{8,9}, and spinocerebellar ataxia (*SCA1*)¹⁰, and have also been linked to numerous complex neurological and psychiatric traits¹¹. A role for STRs as drivers of GWAS signals have been identified¹², where a risk SNP connected adjacent GGAA repeats by converting an interspaced GGAT motif into a GGAA motif, thereby increasing the number of consecutive GGAA motifs and modifying the activity of its sequence and functional impact. STRs have also been shown to significantly regulate gene expression and contribute to phenotypic plasticity¹³. STRs therefore represent a potential source of unexplored genetic variation that may account for some of the missing heritability of PD. In this regard, other repetitive elements, such as satellite repeats, have been shown to alter gene expression in blood of PD patients¹⁴. However, no genome-wide assessment of STRs in large population studies has yet been performed in this disease. Due to their more complex and highly repetitive structure compared to SNVs,

STRs have been difficult to assess. Despite the recent explosion of genetic data stemming from next generation sequencing, STRs are still difficult to genotype. Recent advances in PCR-free deep sequencing methods and STR genotyping tools now allow for the simultaneous assessment of STRs genome-wide¹⁵. Studies have shown high linkage disequilibrium (LD) between STRs and SNPs across the genome¹⁶. Exploiting this high LD, Saini *et al.* (2018)¹⁷ generated a phased SNP-STR haplotype panel based on the 1000 Genomes Project samples that allows for the accurate genome-wide imputation of common STRs into array-based genotype data. To assess the role of common STRs in PD risk, we imputed and interrogated STRs across 16 independent PD case-control cohorts, totaling 39,087 individuals available through the International Parkinson's disease Genomics Consortium (IPDGC).

RESULTS

Meta-analysis of IPDGC GWAS cohorts imputed with an STR reference panel

The 16 GWAS cohorts used in this study, with a combined sample size of 39,087 individuals composed of 16,642 PD cases and 22,445 controls of self-reported European ancestry (**Supplementary Table 1**). After cohort-wise quality controls (see Methods), we performed genome-wide imputation using the 1000 Genomes STR-SNP reference panel¹⁷, and carried out case-control association analyses with PD status following a meta-analysis of fixed effects across all cohorts. After removing variants with high heterogeneity across meta-analyses ($I^2 > 0.8$), we obtained association p-values for 407,879 STRs, where 214 variants surpassed the threshold for genome-wide significance of 5.34×10^{-6} (**Figure 1, upper side**), which was estimated by permutation procedures for the STR reference panel, as described elsewhere¹⁸. The inflation factor lambda for the association was 1.18 and the rescaled lambda for 1000 cases and controls (λ 1000) was 1.01. To characterize and identify independently associated STRs, we first performed a conditional-joint analysis using GCTA-COJO¹⁹ and identified 34 STR variants mapping to 32 unique nearby genes, with the strongest signal located in *KANSL1* (chr17:44205351:[T]₁₁, $p = 3 \times 10^{-39}$,

OR=1.31 [CI 95%=1.26-1.36]), and followed by *SNCA* (chr4:90662073:TATTT[GT]₈AT[GT]₇, $p=3.36 \times 10^{-25}$, OR=1.36 [CI 95%=1.28-1.45]) (**Table 1**). Since STRs were imputed by leveraging LD information from SNPs, we carried out a secondary GCTA-COJO analysis including the meta-analysis results from imputed, filtered SNPs (**Figure 1, lower side**), obtaining a total number of 8,179,378 SNPs and STRs in all cohorts. We found eight loci with associations led by STRs (**Supplementary Table 2**), and in order to refine these results, we further investigated their LD patterns with the 90 known PD risk variants from the 2019 PD GWAS meta-analysis¹, and found that four of the eight STRs had LD $r^2 < 0.5$ with any of the known PD variants (**Supplementary Table 3**), indicating that these could be potential new PD risk signals:

- a tetranucleotide repeat within the 3rd intron of *NDUFAF2* (risk allele chr5:60437492:AA[*TGAA*]₇, $p=6.49 \times 10^{-8}$, OR=1.30, CI 95%=1.18-1.43) (**Figure 2A**);
- a mononucleotide repeat downstream of *TRIML2* (risk allele chr4:189000404:TT[A]₁₂, $p=1.44 \times 10^{-7}$, OR=1.31, CI 95%=1.19-1.44) (**Figure 2B**);
- a mononucleotide repeat downstream of *MIR129-1* (risk allele chr7:127793488:[T]₁₅G, $p=2.79 \times 10^{-7}$, OR=1.16, CI 95%=1.09-1.23) (**Figure 2C**); and
- a mononucleotide repeat within the 44th intron of *NCOR1* (risk allele chr17:15941750:[T]₁₁, $p=3.77 \times 10^{-6}$, OR=1.08, CI 95%=1.04-1.12) (**Figure 2D**).

It is important to note here that the independent STR signal at *NDUFAF2* (chr5:60437492:AA[*TGAA*]₇) is within a known PD risk locus (mapping to *ELOVL7*), and was previously identified through Mendelian randomization to be significantly associated with risk of PD¹. Moreover, further LD analysis on this locus showed a high D' statistic with the closest known PD risk SNP at that locus ($D'=0.94$ with rs1867598) indicating that, regardless of frequency disparities, the independency suggested by the GCTA-COJO analysis should be taken with caution.

Quantifying the heritability of STRs in PD

The genetic heritability of PD was recently estimated to be 22%¹. Here, assuming a global disease prevalence of 0.2%², we leveraged the GCTA-LDMS method¹⁹ and estimated that common STRs (MAF >1%) account for 15.2% (SE=0.01) of the additive heritability of the disease on the liability scale. Heritability for imputed SNPs in the same data accounted for 26.9% (SE=0.02), similarly to what was obtained in Keller *et al.*, 2012 using GCTA as well. After including both common STRs and SNPs in the analysis, the heritability estimate increased to 28.8% (SE=0.02). This increase of 1.9% in the heritability estimate due to common STRs corresponds to a 7% increase from the SNP based estimate.

eSTR analysis

We functionally assessed the impact of the 34 significant STR associations through an expression quantitative trait loci analysis (eSTR). We investigated each locus extracting the leading STR and other STRs in high LD ($r^2 > 0.5$) within 1 Mb up- and downstream, obtaining 105 variants for further analysis (**Supplementary Table 4**). We used normalized gene expression data from frontal cortex from the North American Brain Expression Consortium (NABEC)²⁰, and 13 brain tissues from the Genotype-Tissue Expression Consortium (GTEx v.8)²¹, and identified 10,252 STR-gene associations for both datasets and all tissues (**Figure 3A**). Of these, 840 associations showed a False discovery rate (FDR) corrected $p < 0.05$, corresponding to 234 unique eGenes (genes with at least one significant variant), that included 19 of the 78 loci identified in the 2019 PD GWAS meta-analysis (genes nominated from the 90 PD risk variants): *RIT2*, *TMEM163*, *MCCC1*, *LCORL*, *CTSB*, *SETD1A*, *CRHR1*, *GPNMB*, *BIN3*, *TMEM175*, *GAK*, *MAP4K4*, *SNCA*, *SPTSSB*, *WNT3*, *KPNA1*, *ITGA8*, *BST1* and *HIP1R* (**Supplementary Table 5**).

To obtain functional and gene expression insights on the four candidate STRs signals obtained from our STR meta-analysis, we similarly extracted the four variants and their surrounding high LD STRs, obtaining 13 unique variants. We functionally annotated them, using regulatory features for gene expression from the Encyclopedia of DNA Elements (ENCODE)²², and found two STRs overlapping enhancers, transcription

factor binding sites and histone marks for active transcription (*H3K4me3*, *H3K9Ac*, *H3K4me1*, *H3K27Ac*, *H3K36me3* and *H3K79me2*): one nearby *NDUFAF2* (chr5:60408714:TC[T]₁₄GTATC) in high LD with the leading GWAS STR at that locus (chr5:60437492:AA[TGAA]₇, $r^2=0.95$); and one within *ZSWIM7* (chr17:15902070:[T]₁₆GA), similarly, in high LD with the leading GWAS STR for that locus (chr17:15941750:[T]₁₁, $r^2=0.84$) (**Supplementary Table 6**). The PD risk allele for the eSTR in *NDUFAF2* (major allele with 13 T repetitions) is significantly associated with higher expression levels of the gene *PART1* (~624 kb upstream) in the frontal cortex (**Figure 3B**). Interestingly, the significant eSTR in *ZSWIM7* showed associations in more brain tissues, where the risk allele in the STR meta-analysis (minor allele with 15 T repeats) was correlated with lower expression levels of *TRPV2* in the hypothalamus, anterior cingulate cortex, nucleus accumbens and frontal cortex; higher expression levels of *NCOR1* in the hippocampus; higher expression levels of *ADORA2B* in the anterior cingulate cortex; and lower expression levels of a long non-coding RNA gene located nearby *NCOR1* (CTC-529I10.1 or Inc-NCOR1-1) in the spinal cord and substantia nigra (**Figure 3C**).

Gene-wise, gene-set and pathway enrichment analysis of PD associated STRs

MAGMA²³ gene-wise enrichment analysis of the STR meta-analysis results yielded 47 genes surpassing genome-wide significance (Bonferroni $p < 2.99 \times 10^{-6}$, $\alpha = 0.05/16,696$); **Supplementary Table 7**). Of the 47 genes, 12 overlapped with the STR meta-analysis results, 8 overlapped with 78 PD loci nominated from the 90 PD risk variants (2019 PD GWAS meta-analysis), and 27 genes have not previously been identified as enriched genes. Gene-property analysis using gene expression data from GTEx v.8, as described in FUMA²⁴, showed significant enrichment of genes in the pituitary and brain tissues after FDR correction (FDR $p < 0.05$) for the 30 GTEx general tissues, (**Supplementary Figure 2A**) and in the cerebellum, cortex, pituitary, cerebellar hemisphere and frontal cortex for the 54 GTExp specific tissues (**Supplementary Figure 2B**). We further investigated gene connectivity via protein-protein interactions using a list of 445

genes surpassing nominal gene-wise STR enrichment (MAGMA $p < 0.01$) with WebGestalt²⁵, and leveraged the Network-Topology Analysis (NTA), finding 16 subnetworks (**Supplementary Figure 3**), which were significantly enriched in 27 gene ontology categories, such as synaptic vesicle cycle (GO:0099504), presynaptic endocytosis (GO:0140238) and autophagy (GO:0006914) (**Supplementary Table 8**).

DISCUSSION

In the present study we performed a genome-wide meta-analysis of STRs in 16 cohorts from the IPDGC. We have shown that associated STR signals overlap with known PD risk loci, and with candidate novel signals, that represented by STRs independent from current 90 risk variants¹, and are located nearby *TRIML2*, *NDUFAF2*, *MIR129-1* and *NCOR1* (on chromosomes 4, 5, 7 and 17 respectively). We also assessed the functional consequences of the STRs at a gene expression level in brain tissues, which further supports their candidacy for functional studies to further understand the biological mechanisms behind their associations.

The fact that 88% (30/34) of the associated STRs overlap with the current list of PD GWAS risk variants is not surprising as the STRs were imputed based on their existent LD with SNPs. Known PD loci with STR associations could potentially help to explain the current unknown molecular mechanisms underlying those regions, such as in *MAPT* and *SNCA*, where evidence has shown that repetitive elements play a major role in gene expression regulation, splicing, and hence protein structure^{26,27}. This overlap is further reflected in the heritability estimates we obtained which indicated that the contribution of STRs to the genetic variance of PD is largely explained by their high LD with SNPs. However, STRs have shown to increase the contribution to overall SNP-only heritability estimates specifically on gene expression^{13,28}, where STRs explained between 10%–15% of the *cis*-heritability, thereby supporting our observation that STRs contribute to the heritability of PD.

The eSTR colocalization analyses, using available RNAseq datasets, where we analyzed the top 34 STR signals and their surrounding high LD STRs, showed us different distributions of STR associations throughout the various brain regions, and at the gene level, we observed significant associations in 19 known PD risk genes, suggesting that these STRs are likely to be functionally relevant in these loci. Further investigation of the four independent nominated STRs managed to uncover likely functional mechanisms underlying the STR association in genes nearby *NDUFAF2* and *NCOR1*, due to the STR colocalization with regulatory features (epigenetic marks) involved in active transcription. The eSTR near *NDUFAF2* was found to significantly increase the expression *PART1* (**Figure 3B**). *PART1* is a long non-coding RNA that was found to be differentially expressed (downregulated) in a microarray-based analysis of 50 PD patients compared to 22 healthy controls²⁹. The *ZSWIM7* eSTR was associated with significant effects on gene expression in different genes, such as *TRPV2*, a cation channel part of the Transient receptor potential family of proteins (TRPs) that are activated by physical and chemical stimuli³⁰, and that are known to be involved in the regulation of ionic homeostasis, which is disrupted in PD³¹; *ADORA2B* is an adenosine receptor which has been associated with neurodegenerative conditions such as Huntington's disease³², however no link to PD has been established so far; *Inc-NCOR1-1* and *NCOR1* (Nuclear Receptor Corepressor 1) are located within the same chromosomal region (short arm of chromosome 17) and were also influenced by the eSTRs. The former long non-coding gene has not been thoroughly characterized, therefore little is known about its function. The latter encodes a transcriptional inhibitor that has been found to regulate mitochondrial function³³. Moreover, gene expression analyses showed that *NCOR1* is significantly upregulated in the substantia nigra of PD patients³⁴.

This evidence suggests that those genes associated with eSTRs in PD would be good candidates for follow-up analyses.

The functional consequences of STRs captured by gene-wise and pathway analyses demonstrated that STRs are enriched in known PD-relevant pathways such as synaptic vesicle trafficking³⁵ and autophagy³⁶, and in tissues, such as the cortex, cerebellar hemisphere and frontal cortex. Also highlighted is the pituitary gland, that is known to express the dopaminergic receptors D2 and D4³⁷ and is part of the hypothalamic–pituitary–thyroid axis, where alterations in its balance has been shown to increase risk to PD³⁸.

This study marks the first (to our knowledge) PD STR GWAS to date and highlights the importance of incorporating other forms of genetic variation, such as STRs, into routine genetic analyses. Despite this, like with any study profiling repeat-based variants using short-read sequencing data, the analyses presented in this study have several limitations. First, focusing on the STR calls, STRs were imputed using a reference panel that was generated by the STR caller hipSTR using short-read whole-genome sequencing (WGS)¹⁷. There are two main drawbacks to this approach: (1) hipSTR cannot call STRs that are longer than the read length. Given that many of the known pathogenic STRs in neurological diseases are large repeat expansions, we currently lack the power to detect this important and potentially disease-associated class of tandem repeats; (2) As highlighted in the original study, imputation accuracy varies widely across STR loci, with highly variable multi-allelic STR only achieving ~70% concordance. Hence future studies that validate the PD associated STRs with methods such as long-read sequencing will be crucial to confirm these loci and will be key to resolving complex repeat-based PD associated haplotypes. Second, although the majority of the STRs tested were biallelic, multi-allelic variants were split into biallelic for the GWAS and downstream analyses. This approach enabled us to perform commonly used GWAS methods for the different analysis presented in the study, but set aside the consideration of variant length as unit of analysis, an important aspect of repetitive elements, that need to be addressed in future developments with association tools that can incorporate these multi-allelic

variants, which will likely give valuable insight into the specific role of repeat copy number to risk of disease we. Finally, it is important to highlight that, despite the fact that the STR panel used to impute our PD GWAS cohorts showed high levels of concordance (96.7%)¹⁷ with read-based callers such as hipSTR and TREDPARSE, the STRs reported in this study need further experimental validation, in order to discard any potential artifacts that could exist in both cases and controls, and to confirm their association with PD.

Overall, we have performed the first STR GWAS meta-analysis in PD and reported that STRs contribute to its genetic risk. We have characterized another layer of genetic variation, helping us to gain statistical power to nominate novel candidate PD risk variants and genes, and to provide a more complete reference of the genetic variation that contributes to the disease. Hence this data is a valuable resource for researchers currently dissecting the known PD risk loci. Moving forward, a large-scale GWAS which utilizes calls directly from WGS data and validates hits using long-read sequencing methodologies is essential for fully understanding the contribution of STRs to the genetics of PD.

METHODS

A summary diagram for the methodological steps followed in the present study is shown in **Supplementary Figure 1**.

Samples and quality control

All genotyping data was obtained from previously generated IPDGC datasets, consisting of 39,087 individuals (16,642 cases and 22,445 controls) of European ancestry¹. All individuals provided informed consent for participation in genetics studies, which was approved by the relevant local ethics committee for each of the datasets used. Detailed demographic, sample sizes and PD status are given in

Supplementary Table 1. Further information along with detailed quality control (QC) methods have been previously published^{1,39}. Briefly, for sample QC prior to imputation, individuals with low call rate, discordance between genetic and reported sex, heterozygosity outliers and ancestry outliers were removed. For genotype QC, variants with a missingness rate of >5%, minor allele frequency (MAF) < 0.01, exhibiting deviations from Hardy–Weinberg Equilibrium (HWE) < 1×10^{-5} and palindromic SNPs were excluded.

STR imputation and filtering

STR genotypes were imputed into the IPDGC SNP unimputed genotyping datasets using Beagle v.5.1⁴⁰ with the 1000 Genomes SNP-STR Haplotype reference panel¹⁷. In brief, STR genotypes in the reference panel were imputed from STRs called from the catalog-based STR caller hipSTR¹⁵ and supplemented using a second STR caller, TREDPARSE⁴¹. STRs were phased with corresponding SNPs creating a final panel in the 1000 genomes project data that contained 27,185,239 SNP and 445,725 STR markers. Once STRs were imputed into all IPDGC SNP genotype datasets, the STR calls were filtered to facilitate downstream association analysis. First STRs were split from multi-allelic variants to single biallelic variants using the vt variant tool⁴². Finally SNPs and STRs with a dosage R-squared (DR2) < 0.3 were removed to filter out low quality imputed variants.

Study-level STR analysis and meta-analysis

To estimate PD risk, imputed dosages (*i.e.* genotype probabilities for a variant to be A/A, A/B, or B/B from 0 to 2) were analyzed using a logistic regression model adjusted for sex, age at onset (AAO) for cases or examination for controls, and the first 10 principal components (PCs). To note, AAO could not be included as a covariate for the Myers-Faroud⁴³ and Vance (dbGap phs000394) studies as no AAO information was available. Summary statistics were generated using the RVTESTS package⁴⁴ and filtered

for a MAF >1%. Meta-analysis was conducted based on the fixed-effect model as implemented in METAL⁴⁵ by combining summary statistics across all 16 IPDGC datasets. All variants with a meta-analysis heterogeneity value of less than 80% ($I^2 < 0.80$) were kept for further analysis.

Conditional-joint and linkage disequilibrium analyses

To select candidate variants, we used the Genome-wide Complex Trait Analysis software (GCTA)¹⁹ to perform conditional and joint analysis (COJO) STRs, from the meta-analysis summary statistics. In order to differentiate associations between STRs and SNPs, we performed two COJO analyses, first with STRs only and second, with STRs and SNPs together. As an LD reference for GCTA we used a sample subset of merged imputed genotypes (hard call threshold of 0.8) from the IPDGC GWAS cohorts⁴⁶, totaling 4,397 PD cases and 9,137 controls. Additionally, we performed LD calculations between top STRs with the previously reported list of 90 PD variants¹ using PLINK v.1.9⁴⁷ to determine highly linked STRs to known PD risk variants. Hudson plot showing the genome-wide association results for STRs and SNPs separately was done with the *hudson* R package (<https://github.com/anastasia-lucas/hudson>). Regional plots for the GCTA-nominated independent STRs were done with Locuszoom standalone version⁴⁸.

eSTR analysis

Using sample level genotypes and gene expression data from the North American Brain Expression Consortium (NABEC)²⁰ and the Genotype-Tissue Expression project²¹, we carried out an eQTL analysis with imputed STRs (eSTR). The NABEC data was composed of 343 individuals with genotypes obtained from high-coverage Illumina WGS. Corresponding gene expression data was generated from frontal cortex tissue by RNASeq and normalized gene counts were used. The GTEx v.8 data (dbGaP: phs000424.v7.p2) comprises high-coverage (30X) Illumina WGS data from 838 unrelated samples. We downloaded the fully processed, filtered and normalized gene expression matrices (in BED format) for

each of the 13 brain tissues including: amygdala, Anterior cingulate cortex (BA24), caudate (basal ganglia), cerebellar hemisphere, cerebellum, cortex, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens (basal ganglia), putamen (basal ganglia), spinal cord cervical (c-1) and substantia nigra (<https://gtexportal.org/home/datasets>). WGS genotypes from GTEx and gene start-end coordinates for expression data for GTEx and NABEC were converted from hg38 reference to hg19 using UCSC liftover tool⁴⁹. STRs were imputed as described above. eSTR analysis was performed using the FastQTL software⁵⁰ correcting for PCs 1-10, sample age, sex (if available) and probabilistic estimations of expression residuals factors (PEER) generated using the PEER software⁵¹: 45 factors for NABEC and 15 for GTEx (as indicated in the GTEx documentation). The 34 top STRs from the meta-analysis along with variants with LD >0.5 (105 STRs total) were used to conduct the eSTR analysis. QQ-plots and box plots were done using ggplot2 R package⁵².

Heritability estimation

We used the GCTA-LDMS method^{19,53} to estimate the heritability of STRs only, both SNPs and STRs together and SNPs only. The method corrects for LD bias in the estimated variant-based heritability from WGS or imputed data. Heritability estimates and their corresponding standard errors are shown in the liability scale.

Gene-set, network and pathway enrichment analyses of significant STR loci

To functionally characterize the top associated STRs, we carried out loci connectivity analyses across gene-ontologies and gene-expression datasets using FUMA²⁴ and protein-protein interaction networks using Webgestalt²⁵. We ran MAGMA gene-wise analysis²³ using the meta-analysis summary statistics for all STRs, and used the 1000 Genomes SNP-STR dataset as our reference panel¹⁷. We selected 445 genes with a gene-wise $p < 0.01$ for further analyses (**Supplementary Table 7**). Gene lists were analyzed for

functional enrichments using (i) FUMA gene2func tool, (ii) Biogrid PPI Network Topology-based Analysis (NTA) in Webgestalt and (iii) gene property analysis for tissue specificity, using 23,675 genes from Genotype-Tissue Expression (GTEx) RNASeq data²¹ across the 30 general and 54 specific tissues. Data preprocessing and gene expression normalization methods are presented in the FUMA tutorial section (<https://fuma.ctglab.nl/tutorial>). Bonferroni and Benjamini-Hochberg FDR corrections for multiple testing were performed for MAGMA gene-wise results and functional enrichment analyses, respectively.

Data Availability

Full STR GWAS summary statistics for the 16 datasets meta-analysed are available at <https://drive.google.com/file/d/1kD1i6tHdYC5w0xvxWLD4B-bSPqpnwzNV/view?usp=sharing>

Code Availability

The STR imputation, study level GWAS and meta-analysis: https://github.com/neurogenetics/PD_STR_imputation. Downstream analyses: https://github.com/bibb/STR_GWAS_downstream_analysis

Acknowledgements

We would like to thank all of the subjects who donated their time and biological samples to be a part of this study. We also would like to thank all members of the International Parkinson's Disease Genomics Consortium (IPDGC). For a complete overview of members, acknowledgements and funding, please see <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. In addition, this work was supported by the Department of Defense (award W81XWH-09-2-0128), and The Michael J Fox Foundation for Parkinson's Research. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). The access to part of the participants for this research has been made possible thanks to the Quebec Parkinson's Network (<http://rpg-qpn.ca/en/>).

This work was supported by the Simpson Querrey Center for Neurogenetics (to D.K.)

Conflicts of Interest

D.K. is the Founder and Scientific Advisory Board Chair of Lysosomal Therapeutics Inc. and Vanqua Bio. D.K. serves on the scientific advisory boards of The Silverstein Foundation, Intellia Therapeutics, AcureX and Prevail Therapeutics and is a Venture Partner at OrbiMed. Z.GO. has received consulting fees from Lysosomal Therapeutics Inc., Idorsia, Prevail Therapeutics, Denali, Ono Therapeutics, Neuron23, Handl Therapeutics, Bial Biotech Inc., Deerfield, Lighthouse and Inception Sciences (now Ventus). B.I.B., K.B., C.B., J.R.G., A.B.S. and S.J.L. declare that they have no competing interests.

REFERENCES

1. Nalls, M. A. *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.* **18**, 1091–1102 (2019).
2. Keller, M. F. *et al.* Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum. Mol. Genet.* **21**, 4996–5009 (2012).
3. Consortium, I. H. G. S. & International Human Genome Sequencing Consortium. Correction: Initial sequencing and analysis of the human genome. *Nature* vol. 412 565–566 (2001).
4. Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M. & Neretti, N. Transcriptional landscape of

- repetitive elements in normal and cancer human cells. *BMC Genomics* **15**, 583 (2014).
5. Fu, Y. H. *et al.* Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**, 1047–1058 (1991).
 6. Verkerk, A. J. *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914 (1991).
 7. MacDonald, M. E. *et al.* A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington’s disease chromosomes. *Cell* **72**, 971–983 (1993).
 8. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
 9. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
 10. Orr, H. T. *et al.* Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* **4**, 221–226 (1993).
 11. Hannan, A. J. Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.* **19**, 286–298 (2018).
 12. Grünewald, T. G. P. *et al.* Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. *Nat. Genet.* **47**, 1073–1078 (2015).
 13. Gymrek, M. *et al.* Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat. Genet.* **48**, 22–29 (2016).
 14. Billingsley, K. J. *et al.* Analysis of repetitive element expression in the blood and skin of patients with Parkinson’s disease identifies differential expression of satellite elements. *Scientific Reports* vol. 9 (2019).
 15. Willems, T. *et al.* Genome-wide profiling of heritable and de novo STR variations. *Nat. Methods* **14**, 590–592 (2017).

16. Payseur, B. A., Place, M. & Weber, J. L. Linkage disequilibrium between STRPs and SNPs across the human genome. *Am. J. Hum. Genet.* **82**, 1039–1050 (2008).
17. Saini, S., Mitra, I., Mousavi, N., Fotsing, S. F. & Gymrek, M. A reference haplotype panel for genome-wide imputation of short tandem repeats. *Nat. Commun.* **9**, 4397 (2018).
18. S. Saini, M. Gymrek, PGC Schizophrenia Working Group. Studying the role of short tandem repeat variants in schizophrenia risk. (2019).
19. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
20. Gibbs, J. R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
21. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
22. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
23. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
24. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
25. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
26. Caillet-Boudin, M.-L., Buée, L., Sergeant, N. & Lefebvre, B. Regulation of human MAPT gene expression. *Mol. Neurodegener.* **10**, 28 (2015).
27. Afek, A. *et al.* Toward deciphering the mechanistic role of variations in the Rep1 repeat site in the transcription regulation of SNCA gene. *Neurogenetics* **19**, 135–144 (2018).

28. Quilez, J. *et al.* Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. *Nucleic Acids Research* vol. 44 3750–3762 (2016).
29. Chi, L.-M., Wang, L.-P. & Jiao, D. Identification of Differentially Expressed Genes and Long Noncoding RNAs Associated with Parkinson's Disease. *Parkinson's Disease* vol. 2019 1–7 (2019).
30. Duitama, M. *et al.* TRP Channels Role in Pain Associated With Neurodegenerative Diseases. *Front. Neurosci.* **14**, 782 (2020).
31. Vaidya, B. & Sharma, S. S. Transient Receptor Potential Channels as an Emerging Target for the Treatment of Parkinson's Disease: An Insight Into Role of Pharmacological Interventions. *Front Cell Dev Biol* **8**, 584513 (2020).
32. Liu, J. *et al.* Genetics Modulate Gray Matter Variation Beyond Disease Burden in Prodromal Huntington's Disease. *Frontiers in Neurology* vol. 9 (2018).
33. Fan, W. & Evans, R. PPARs and ERRs: molecular mediators of mitochondrial metabolism. *Curr. Opin. Cell Biol.* **33**, 49–54 (2015).
34. Cowherd, M. & Lee, I. Transcriptional regulators are upregulated in the substantia nigra of Parkinson's disease patients. *J. Emerg. Invest.*, 1–7. (2015).
35. Esposito, G., Clara, F. A. & Verstreken, P. Synaptic vesicle trafficking and Parkinson's disease. *Developmental Neurobiology* vol. 72 134–144 (2012).
36. Lim, G. G. Y. *Role of Autophagy in Parkinson's Disease*. (IntechOpen, 2013).
37. Jaber, M., Robinson, S. W., Missale, C. & Caron, M. G. Dopamine receptors and brain function. *Neuropharmacology* **35**, 1503–1519 (1996).
38. Mohammadi, S., Dolatshahi, M. & Rahmani, F. Shedding light on thyroid hormone disorders and Parkinson disease pathology: mechanisms and risk factors. *J. Endocrinol. Invest.* **44**, 1–13 (2021).
39. Blauwendraat, C. *et al.* Parkinson disease age of onset GWAS: defining heritability, genetic loci and a-synuclein mechanisms. (2019) doi:10.1101/424010.

40. Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
41. Tang, H. *et al.* Profiling of Short-Tandem-Repeat Disease Alleles in 12,632 Human Whole Genomes. *Am. J. Hum. Genet.* **101**, (2017).
42. Tan, A., Abecasis, G. R. & Kang, H. M. Unified representation of genetic variants. *Bioinformatics* **31**, 2202–2204 (2015).
43. Pankratz, N. *et al.* Meta-analysis of Parkinson’s disease: identification of a novel locus, RIT2. *Ann. Neurol.* **71**, 370–384 (2012).
44. Zhan, X., Hu, Y., Li, B., Abecasis, G. R. & Liu, D. J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423–1426 (2016).
45. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* vol. 26 2190–2191 (2010).
46. Simon-Sanchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson’s disease. *Nat. Genet.* **41**, 1308 (2009).
47. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
48. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
49. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–8 (2006).
50. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
51. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat.*

Protoc. **7**, 500–507 (2012).

52. Wickham, H. *GGPLOT2: Elegant Graphics for Data Analysis 2016 Springer-Verlag, New York.* (2016).
53. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).

Figure Legends

Figure 1. Genome-wide association results for imputed STRs and SNPs in 16 PD GWAS cohorts from the International Parkinson’s disease genomics consortium. Hudson plot representing the association analysis results for STRs (upper) and SNPs (lower) across the human genome, showing the 34 genome-wide significant STR loci ($p < 5.34 \times 10^{-6}$) after the conditional-joint analysis with GCTA. Genes in bold represent the loci influenced by STRs after including SNPs associations. Genes with a black dot at the top represent STR loci independent from the current 90 PD risk variants from the 2019 PD GWAS meta-analysis.

Figure 2. Regional association plots for the 4 candidate independent STR loci. Locus zoom plots were generated for the 4 GCTA-nominated independent STR loci from SNPs and 90 risk PD loci in **(A)** chromosome 5 within *NDUFAF2*, **(B)** chromosome 4 nearby *TRIML2*, **(C)** chromosome 7 nearby *MIR129-1* and **(D)** chromosome 17 within *NCOR1*. Lead STR variant is depicted as a purple diamond and nearby variants (STRs and SNPs) in circles colored by their LD r^2 value to the lead STR variant. Gene annotations for each region are displayed in the bottom part of each panel, showing gene strand orientation with arrows.

Figure 3. eSTR analysis of top STR loci in gene expression data from brain tissues. **(A)** Quantile-quantile plot for eSTR analysis of all 34 top loci from meta-analysis and high LD STRs, totalling 105 variants, across

brain tissues from the NABEC and GTEx datasets. Colors for each dot and lines were added to enhance resolution. **(B)** Box plots showing gene expression changes associated with STR variant chr5:60408714:TC[T]₁₄GTATC, across the tissues where the variant was analyzed. **(C)** Box plots showing gene expression changes associated with STR variant chr17:15902070:[T]₁₆GA.

Locus representation is shown at the top of each box plot, representing the STR location and the distance (in kilobases) and orientation of the target gene. At the bottom, allele dosages, tissue, sample size and eSTR FDR-adjusted p-value is shown, with p-values in red representing significant associations.

NA=Nucleus accumbens (basal ganglia); CT=Caudate (basal ganglia); CB=Cerebellum; CTX=Cortex; FC (NABEC)=Frontal cortex; FC=Frontal cortex (BA9); HT=Hypothalamus;CBH=Cerebellar hemisphere; HC=Hippocampus; PM=Putamen (basal ganglia); ACC=Anterior cingulate cortex (BA24); SC=Spinal cord (cervical c-1); AM= Amygdala; SN=Substantia nigra.

Figure 1

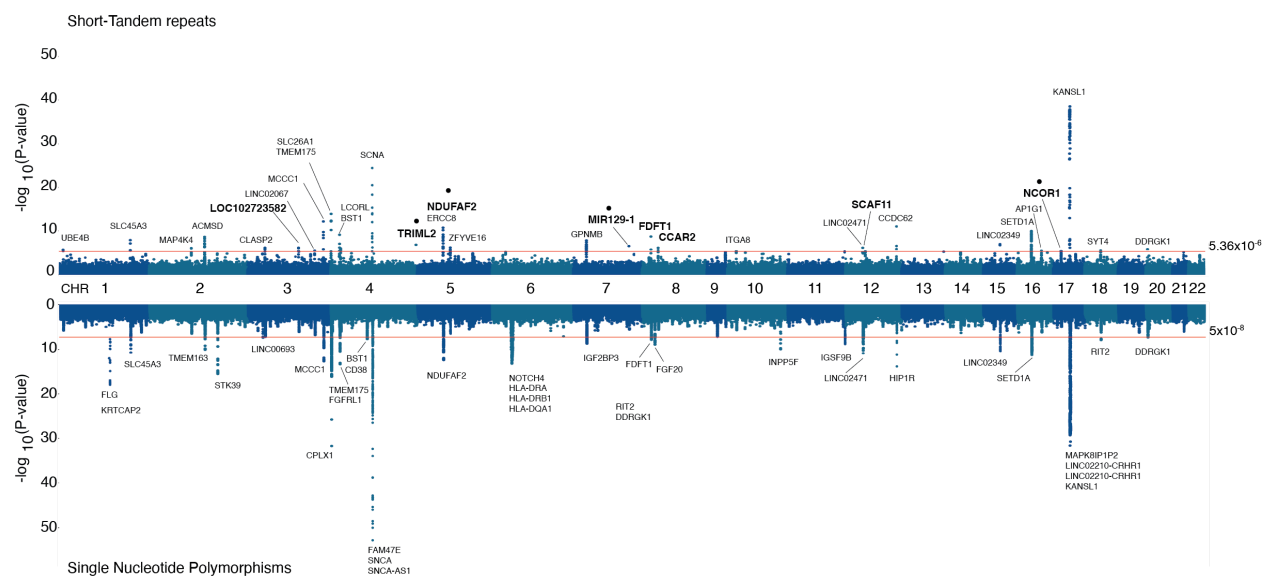


Figure 2

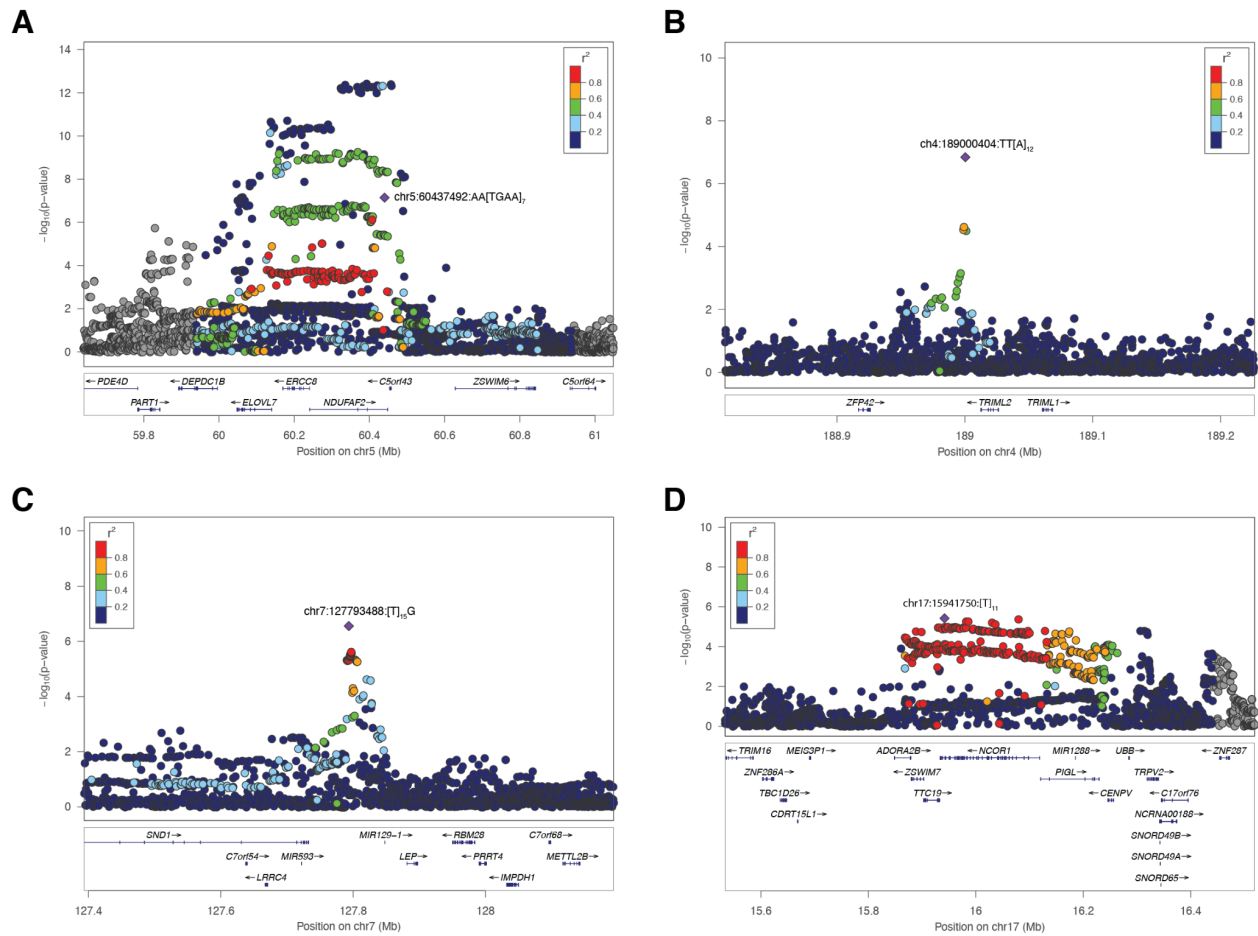


Figure 3

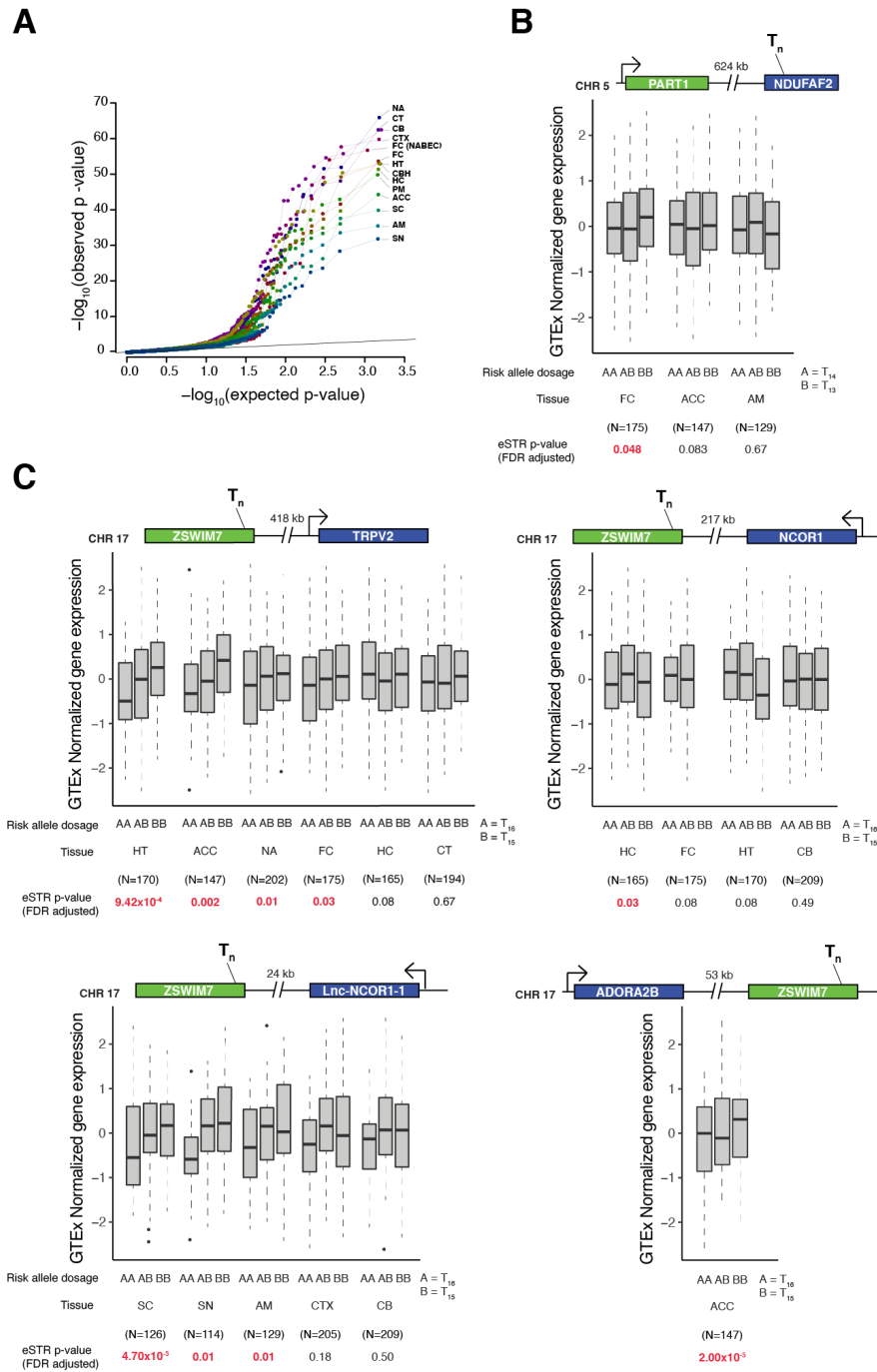


Table 1. Genome-wide significant STR loci from meta-analysis of 16 case-control Parkinson's Disease GWAS cohorts.

Chr	BP	Effect allele	Other allele	Nearest Gene	P-value			I ² (%)	P value conditional-joint analysis
					(Fixed effects)	Effect allele freq	OR (95% CI)		
1	10234448	ATC[A] ₁₄	ATC[A] ₁₅	UBE4B	1.89x10 ⁻⁶	0.971	1.33 (1.18-1.5)	0	1.86x10 ⁻⁶
1	205654915	TG[T] ₇ [GTTT] ₄ TTTG	TG[T] ₇ [GTTT] ₃ TTTG	SLC45A3	1.25x10 ⁻⁸	0.826	0.85 (0.81-0.9)	44.8	1.28x10 ⁻⁸
2	102312361	AAA[CAAA] ₃ AAACAA	AAA[CAAA] ₄ AAACAA	MAP4K4	7.98x10 ⁻⁷	0.330	1.08 (1.05-1.12)	0	7.77x10 ⁻⁷
2	135558779	[AT] ₆	[AT] ₈	ACMSD	2.36x10 ⁻⁹	0.352	0.89 (0.86-0.93)	14.8	2.53x10 ⁻⁹
3	33592979	[T] ₁₁	[T] ₁₂	CLASP2	6.57x10 ⁻⁷	0.594	0.9 (0.87-0.94)	0	6.77x10 ⁻⁷
3	122146661	[A] ₁₂ C[A] ₆	[A] ₁₁ C[A] ₆	LOC102723582	6.52x10 ⁻⁷	0.168	1.11 (1.06-1.16)	0	6.76x10 ⁻⁷
3	161125165	TATTG[T] ₁₄ C	TATTG[T] ₁₅ C	LINC02067	2.88x10 ⁻⁶	0.371	1.08 (1.04-1.12)	26.9	2.85x10 ⁻⁶
3	182753293	[T] ₁₇ GGA	[T] ₁₈ GGA	MCCC1	5.98x10 ⁻¹³	0.744	1.15 (1.11-1.2)	0	5.79x10 ⁻¹³
4	935846	[T] ₉ C[T] ₁₀ G	[T] ₉ [T] ₁₁ G	TMEM175	6.59x10 ⁻¹¹	0.644	1.11 (1.08-1.15)	25	5.17x10 ⁻⁷
4	975260	[CAAT] ₅	[CAAT] ₆	SLC26A1	1.19x10 ⁻¹⁴	0.873	0.82 (0.78-0.86)	58.1	1.01x10 ⁻¹⁰
4	15732538	[T] ₁₂	[T] ₁₃	BST1	7.82x10 ⁻¹⁰	0.116	0.85 (0.81-0.89)	0	1.52x10 ⁻⁹
4	17966489	CC[T] ₁₄ C	C[T] ₁₄ C	LCORL	9.38x10 ⁻⁷	0.153	0.89 (0.85-0.93)	0	1.72x10 ⁻⁶
4	90487729	[T] ₁₁	[T] ₁₂	SNCA	5.00x10 ⁻⁶	0.435	0.92 (0.89-0.95)	42.5	9.64x10 ⁻¹⁰
4	90662073	TATTT[GT] ₈ AT[GT] ₇	TATTT[GT] ₉ AT[GT] ₇	SNCA	3.36x10 ⁻²⁵	0.076	1.36 (1.28-1.45)	0	1.31x10 ⁻³⁹
4	90662941	[A] ₁₃	[A] ₁₄	SNCA	4.41x10 ⁻¹⁹	0.683	0.85 (0.82-0.88)	0	4.94x10 ⁻²⁸
4	189000404	TT[A] ₁₂	TT[A] ₁₃	TRIML2	1.44x10 ⁻⁷	0.942	1.3 (1.18-1.44)	16.1	1.46x10 ⁻⁷
5	60181876	[TA] ₂ TG[TA] ₅ TG[TA] ₅	[TA] ₂ TG[TA] ₅ TG[TA] ₆	ERCC8	1.96x10 ⁻¹¹	0.901	0.83 (0.78-0.87)	0	4.04x10 ⁻¹⁰
5	60437492	AA[TGAA] ₆	AA[TGAA] ₇	NDUFAF2	6.94x10 ⁻⁸	0.466	0.76 (0.69-0.84)	0	1.39x10 ⁻⁶

5	79741582	[T] ₁₂	[T] ₁₃	ZFYVE16	6.01x10 ⁻⁷	0.941	1.19 (1.11-1.27)	6	6.10x10 ⁻⁷
7	23290184	TTC[A] ₁₃	TTC[A] ₁₄	GNPMB	1.47x10 ⁻⁸	0.624	1.09 (1.06-1.13)	38	1.41x10 ⁻⁸
7	127793488	[T] ₁₄ G	[T] ₁₅ G	MIR129-1	2.79x10 ⁻⁷	0.742	0.85 (0.81-0.91)	70.6	2.89x10 ⁻⁷
8	11696990	TCTACT[A] ₁₃	TCTACT[A] ₁₄	FDFT1	1.96x10 ⁻⁹	0.259	0.89 (0.86-0.92)	11.8	1.98x10 ⁻⁹
8	22464976	TAGGG[T] ₂₀ GATG	TAGGG[T] ₂₁ GATG	CCAR2	6.91x10 ⁻⁷	0.677	0.91 (0.87-0.94)	0	6.98x10 ⁻⁷
10	15563184	[T] ₁₁	[T] ₁₂	ITGA8	4.73x10 ⁻⁶	0.874	1.11 (1.06-1.17)	49.8	4.87x10 ⁻⁶
12	40587472	CG[T] ₉ [GTT] ₃ [T] ₄ G[T] ₇	C[T] ₈ [GTT] ₃ [T] ₄ G[T] ₇	LINC02471	6.81x10 ⁻⁷	0.913	0.86 (0.81-0.91)	0	6.94x10 ⁻⁷
12	46452915	AAGCAAGCA	AA[GCAA] ₃ GCA	SCAF11	5.19x10 ⁻⁶	0.439	0.92 (0.89-0.95)	0	5.32x10 ⁻⁶
12	123299993	[T] ₁₁ C[T] ₆ G	[T] ₁₂ C[T] ₆ G	CCDC62	1.03x10 ⁻¹¹	0.565	0.88 (0.85-0.91)	33.3	9.68x10 ⁻¹²
15	61988505	[AC] ₇	[AC] ₈	LINC02349	1.01x10 ⁻⁷	0.257	0.9 (0.86-0.93)	30.3	1.05x10 ⁻⁷
16	30981587	GA[T] ₁₇ GAGA	GA[T] ₁₈ GAGA	SETD1A	1.11x10 ⁻¹⁰	0.650	1.11 (1.07-1.15)	22.7	1.08x10 ⁻¹⁰
16	71855316	CTC[A] ₁₃ [GAAAA]3A	CTC[A] ₁₂ [GAAAA]3AA	AP1G1	3.07x10 ⁻⁶	0.668	1.1 (1.06-1.15)	0	3.07x10 ⁻⁶
17	15941750	[T] ₁₀	[T] ₁₁	NCOR1	3.77x10 ⁻⁶	0.568	0.92 (0.89-0.95)	19	3.67x10 ⁻⁶
17	44205351	[T] ₁₁	[T] ₁₃	KANSL1	3.00x10 ⁻³⁹	0.789	1.31 (1.26-1.36)	0	3.65x10 ⁻³⁹
		[AT] ₇ A[CA] ₄ [AT] ₃ GC[AT] ₄ AACAATAAA[TA] ₄ C[AT] ₃ GTGT		[AT] ₈ [AC] ₄ [AT] ₄ GC[AT] ₄ AACAATAAA[AT] ₄ AC[AT] ₃ GTGT[AT]					
18	40777631	AT] ₄ A	₄ A	SYT4	2.64x10 ⁻⁶	0.424	1.08 (1.04-1.11)	33.8	2.52x10 ⁻⁶
20	3163846	TC[A] ₁₃	TC[A] ₁₄	DDRGK1	1.43x10 ⁻⁶	0.705	0.91 (0.88-0.94)	15.8	1.35x10 ⁻⁶

Chr=chromosome. BP=base pair position. O=Odds Ratio. I²=meta-analysis heterogeneity.