Published in final edited form as:

Ann Neurol. 2018 July ; 84(1): 117–129. doi:10.1002/ana.25274.

A Comprehensive Analysis of *SNCA*–Related Genetic Risk in Sporadic Parkinson Disease

Lasse Pihlstrøm, MD, PhD¹, Cornelis Blauwendraat, PhD^{2,3}, Chiara Cappelletti, BSc¹, Victoria Berge-Seidl, MD¹, Margrete Langmyhr, MSc¹, Sandra Pilar Henriksen, MSc¹, Wilma D. J. van de Berg, MD, PhD⁴, J. Raphael Gibbs, PhD³, Mark R. Cookson, PhD³, the International Parkinson Disease Genomics Consortium, the North American Brain Expression Consortium, Andrew B. Singleton, PhD³, Mike A. Nalls, PhD^{3,5}, and Mathias Toft, MD, PhD^{1,6}

¹Department of Neurology, Oslo University Hospital, Oslo, Norway; ²Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD; ³Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD; ⁴Department of Anatomy and Neurosciences, Clinical Neuroanatomy Section, Amsterdam Neuroscience, VU Medical Center, Amsterdam, the Netherlands; ⁵Data Tecnica International, Glen Echo, MD; ⁶Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Abstract

Objective: The goal of this study was to refine our understanding of disease risk attributable to common genetic variation in *SNCA*, a major locus in Parkinson disease, with potential implications for clinical trials targeting α -synuclein. We aimed to dissect the multiple independent association signals, stratify individuals by *SNCA*–specific risk profiles, and explore expression quantitative trait loci.

Methods: We analyzed participant-level data from 12,503 patients and 12,502 controls, optimizing a risk model and assessing *SNCA*–specific risk scores and haplotypes as predictors of individual risk. We also explored hypotheses about functional mechanisms and correlated risk variants to gene expression in human brain and protein levels in cerebrospinal fluid.

Results: We report and replicate a novel, third independent association signal at genome-wide significance level downstream of *SNCA* (rs2870004, $p = 3.0 \times 10^{-8}$, odds ratio [OR] = 0.88, 95% confidence interval [CI] = 0.84–0.92). *SNCA* risk score stratification showed a 2-fold difference in

Address correspondence to Dr Pihlstrøm, Department of Neurology, Oslo University Hospital, Box 4950 Nydalen, 0424 Oslo, Norway. lasse.pihlstrom@medisin.uio.no.

Author Contributions

L.P., C.B., M.A.N., A.B.S., and M.T. contributed to conception and design of the study. L.P., C.B., C.C., V.B.-S., M.L., S.P.H., W.D.J.v.d.B., J.R.G., M.R.C., M.A.N., and M.T. contributed to acquisition, analysis, and/or interpretation of data. The IPDGC and North American Brain Expression Consortium provided data. L.P. drafted the manuscript with significant input from C.B. All authors contributed to editing and approval of the final draft.

A full list of IPDGC and NABEC consortium members is available as an online supplementary file.

Potential Conflicts of Interest

Nothing to report.

disease susceptibility between top and bottom quintiles (OR = 1.99, 95% CI = 1.78–2.23). Contrary to previous reports, we provide evidence supporting top variant rs356182 as functional in itself and associated with a specific *SNCA* 5' untranslated region transcript isoform in frontal cortex.

Interpretation: The *SNCA* locus harbors a minimum of 3 independent association signals for Parkinson disease. We demonstrate a fine-grained stratification of α -synuclein-related genetic burden in individual patients of potential future clinical relevance. Further efforts to pinpoint the functional mechanisms are warranted, including studies of the likely causal top variant rs356182 and its role in regulating levels of specific *SNCA* mRNA transcript variants.

SNCA was the first gene to be implicated in Parkinson disease (PD). Missense mutations and genomic multiplications of *SNCA* cause a rare autosomal dominant form of PD,^{1,2} whereas common, noncoding single nucleotide polymorphisms (SNPs) at the same locus have consistently shown the most significant association with disease risk across all large-scale genome-wide association studies (GWASs).^{3–11} The corresponding protein, α -synuclein, plays a pivotal role in the molecular pathogenesis of PD, where it is known to exhibit prionlike properties,¹² forming toxic oligomers and aggregates within neurons¹³ that ultimately manifest as Lewy bodies,¹⁴ the histopathological hallmark lesions of PD. Targeting α -synuclein is currently seen as one of the most promising strategies toward developing disease-modifying therapies for PD, where elucidating the causative mechanisms linking genetic variability to PD risk could provide important clues.¹⁵ Furthermore, the prospect of identifying individuals or patients with a particularly high *SNCA*-related genetic burden will likely have implications for case selection to clinical trials and, ultimately, for tailored treatment.

However, after more than a decade of *SNCA* association studies in PD, several important questions remain unanswered. A number of GWASs have reported independently significant, secondary *SNCA* association signals, indicating allelic heterogeneity of susceptibility variants for sporadic PD, yet it remains unclear how these signals relate to each other and how many distinct functional variants they represent.^{5,6,8,10} Adding to the complexity, the secondary signal from the largest meta-GWAS, but not the top hit, also shows association with dementia with Lewy bodies (DLB), a closely related neurodegenerative disorder.¹⁶ It is generally believed that common risk variants affect gene regulation, causing higher intracellular levels of α -synuclein, consistent with the pathogenic effect of increased gene dosage observed in patients with genomic multiplications.² However, investigations of *SNCA* susceptibility SNPs as expression quantitative trait loci (eQTLs) in brain have thus far not yielded sufficient evidence to firmly corroborate this hypothesis.^{5,17} The *SNCA* top hit signal for PD risk has consistently been located near the 3' end of the gene, and 2 recent studies have proposed different functional explanations and causal variants underlying this strong disease association, which have yet to be followed up in independent investigations.^{18,19}

The present study aimed to refine our understanding of the *SNCA* association with PD using several different approaches. We comprehensively explored the allelic heterogeneity using individual-level genotype data from >25,000 individuals, where we identified and replicated

a novel, third independent genome-wide significant association signal. Combining *SNCA* risk alleles across 3 independent signals into a locus-specific risk score, we demonstrate a fine-grained stratification of α -synuclein-related genetic burden of potential future clinical relevance in individual PD patients. We argue that the top hit variant rs356182 is in itself causal and drives the top association signal, a hypothesis supported by both association statistics and publicly available functional data. We further report a novel association between this variant and a specific *SNCA* mRNA transcript variant, consistently observed across 3 independent postmortem human frontal cortex datasets.

Subjects and Methods

Genetic Disease Risk Association Analyses

To explore the pattern of association with PD risk at the *SNCA* locus in depth, we analyzed individual-level genotypes from a large sample set of 5,542 PD patients and 5,866 healthy controls generated by the International Parkinson Disease Genomics Consortium (IPDGC). Aiming to replicate the findings from this discovery sample, we subsequently obtained 10 smaller case-control cohorts, which were combined into a replication dataset of 6,961 PD patients and 6,636 controls. Together, the discovery and replication sample sets comprised a total of 12,503 PD patients and 12,502 controls, all from Caucasian populations, which are summarized in Table 1. Recruitment to each of the substudies contributing data has been based on ethical approval by relevant authorities locally, with written informed consent from all participants. For all substudies, healthy controls were free from symptoms or signs of neurological disease and/or parkinsonism, based on either self-report or clinical examination. Following sample level (missingness, excess heterozygosity, ancestry outliers, sex-check failures) and variant level (genotype rate, Hardy–Weinberg equilibrium) quality filtering and removal of samples with known monogenic variants, all datasets were imputed using the Michigan Imputation Server²⁰ with default settings and reference data from the Haplotype Reference Consortium.²¹ Biallelic SNPs with an imputation quality corresponding to $r^2 > 0.3$ and minor allele frequency (MAF) > 0.01 were included for analyses. For all datasets with available genome-wide genotype data (Oslo, Parkinson's Progression Marker Initiative [PPMI], and the 3 dbGAP datasets), identity by descent was analyzed in a merged dataset of pruned, common SNPs to identify and filter out duplicates and related participants across studies. We calculated principal components (PCs) for each dataset based on a subset of common, linkage disequilibrium (LD)-pruned SNPs, using MAF > 0.05 and pairwise $r^2 < 0.5$ as thresholds. Top 5 PCs, sex, and study were used as covariates in all association tests. A random set of 25,000 SNPs outside the *SNCA* region on chromosome 4 was used to assess genomic inflation, showing lambda 1.07 without and 1.02 with covariates. Adding age at onset (cases) or study (controls) to the model did not improve lambda and gave similar association results for the *SNCA* region (identical signals significant, data not shown). Given that age data were not complete for all replication datasets, and not suitable as a covariate for the Finnish substudy, which included particularly old controls, we chose not to include age as a covariate.

We defined our region of interest as the 1Mb window centered on the top hit SNP rs356182, including 2,328 variants with MAF > 0.01 . To determine the number of independent

statistical tests for multiple testing correction, we applied spectral decomposition adjustment using the SNPSpD method (<https://gump.qimr.edu.au/general/daleN/SNPSpD/>),^{22,23} where the most conservative estimate of independent tests was 220, corresponding to a Bonferroni-corrected significance threshold of $p < 0.00023$. Independent association signals were investigated by logistic regression using PLINK 1.9,²⁴ in a stepwise forward approach originally suggested by Cordell and Clayton.²⁵ Starting with unconditional regression analysis assuming a log-additive model, the most significant SNP is identified, which is then included as a covariate in the next round of analysis, repeatedly until no SNP has a significant p value (Fig 1). Having identified 3 independent signals, we fit a model including all SNPs using R3.1.2, and applied the same joint model for replication and combined analysis before exploring alternative models including haplotypes or proposed functional SNPs. We used fastPHASE 1.4.8 for estimation of individual haplotypes.²⁶ Individual genetic risk scores for the replication dataset were calculated according to a log-additive model, adding the β estimates of all risk SNP alleles from the discovery association analysis.

Expression Quantitative Trait Locus Analyses

We reanalyzed postmortem frontal cortex cap analysis gene expression sequencing (CAGEseq) data and imputed genotypes from the *SNCA* locus from 117 neurologically normal individuals, where materials and methods have been described in detail in a previous publication.¹⁷ Samples with complete data for the relevant SNPs included 77 males and 33 females with mean age = 51 years (range = 2–95), mean RNA integrity number (RIN) = 7.6 (range = 5.4–9.1), and mean postmortem interval = 11 hours (range = 2–28). Findings were followed up in available data from an ongoing RNA sequencing (RNAseq) study by the North American Brain Expression Consortium (dbGAP phs001300.v1.p1). In brief, total RNA was extracted from human frontal cortex using Trizol and cDNA libraries synthesized using the TruSeq stranded total RNA library prep kit with Ribo-Zero (Illumina, San Diego, CA). Libraries were pooled and run on a HiSeq2500 sequencer to 200bp sequence length. Samples overlapping with CAGEseq were removed, leaving a total of 150 nonneurological frontal cortex samples with nonmissing genotype data for 55 males and 86 females with mean age = 59 years (range = 15–97), mean RIN = 7.7 (range = 6.0–9.5), and mean postmortem interval = 10.0 hours (range = 1.2–36.0). The use of human brain samples from both of these datasets was approved by the NIH Office for Human Subjects Research. For further replication, we analyzed 82 postmortem frontal cortex samples from the Netherlands Brain Bank with neuropathologically confirmed Lewy body disease. DNA and RNA extraction and cDNA synthesis were performed using standard techniques, and RNA integrity was assessed using the Agilent 2100 Bioanalyzer. Genotyping and gene expression analyses were performed by quantitative polymerase chain reaction (qPCR) using TaqMan assays on a ViiA7 instrument (Life Technologies, Carlsbad, CA). Expression was quantitated by the delta-CT method, using 3 reference genes (*GAPDH*, *HPRT1*, *SDHA*). After excluding 4 samples with RIN < 6 and 1 with missing index SNP genotype calls, remaining samples included 47 males and 39 females of mean age = 78.6 years (range = 54–98), mean RIN = 8.4 (range = 6.1–9.9), and mean postmortem interval = 5.8 hours (range = 2.8–10.2). All brain samples were from subjects of Caucasian ancestry.

We investigated association between mRNA expression and index SNPs representing 3 independent PD susceptibility signals by linear regression using R 3.1.2. Given that these SNPs are in LD, yet carry an independent effect on disease risk, all SNPs were included in the regression model as potential individual predictors for the level of total *SNCA* mRNA, as well as 3 individual 5' untranslated region (UTR) transcript isoforms (NM_000345, NM_001146054, and NM_001146055; Fig 2A). The results prompted us to also investigate the relative proportion of transcript variant NM_001146054 as a fraction of total *SNCA* mRNA as an alternative measure. Based on a total of 15 independent tests (3 SNPs $\times$ 5 expression outcomes), we set a stringent Bonferroni-corrected significance threshold for discovery to $p < 0.0033$. For comparison, we also repeated the analyses with each of the previously proposed functional SNPs rs356168 and rs356165 replacing the top variant rs356182 in CAGEseq data.

The 2 RNA sequencing datasets were generated as part of projects studying expression genome-wide, and we chose to adopt the original use of covariates as applied in each of these larger studies. For CAGEseq and qPCR data, sex, age at death, postmortem interval, and RIN value were included as covariates, and for the former also the top 6 PCs as calculated from geno-type and gene expression data. The RNAseq quantified expression levels were normalized using quantile normalization and then adjusted for both known and unknown covariates using the probabilistic estimation of expression residuals (PEER)²⁷ tool. PEER was run allowing the tool to infer the suitable adjustment factors based on up to 31 unknown and 19 known covariates. The normalized and covariate adjusted data were used for regression analysis.

In a final exploratory analysis, we investigated association between the 3 index SNPs and cerebrospinal fluid (CSF) α -synuclein levels in data from the PPMI study, including 162 healthy controls and 352 PD patients. Data from CSF collection at baseline generated using a commercially available enzyme-linked immunosorbent assay kit (Covance, Princeton, NJ)⁴⁴ were downloaded from the PPMI site (www.ppmi-info.org). Similar to above, we tested for association in a linear regression model including all index SNPs and sex, age, disease status (PD vs healthy control), and top 6 PCs as covariates. Adjusting for 3 SNPs investigated, significance threshold was set to $p < 0.016$. As a post hoc analysis, we also ran the same analysis separately in PD patients and controls.

Results

Stepwise Conditional Regression Analyses Identify a Novel Risk Signal

Replication of consistent association findings across independent, large datasets is crucial to corroborate a proposed multiallelic model for locus-specific risk. We used a single-study dataset of individual-level genotypes including 5,542 patients versus 5,866 controls from the IPDGC for stepwise conditional logistic regression analysis in the discovery phase of our study.²⁸ Three signals reached our multitest significance threshold, the top 2 identical to or in high LD with previously reported independent GWAS hits (rs356182 and rs763443),⁵ followed by a novel independent signal downstream of *SNCA* (rs2870004; Table 2, Fig 1). The independent association with each of these 3 variants was reproduced in the replication

dataset. In joint logistic regression analyses, all 3 signals passed the genome-wide significance threshold. Results from association analyses are presented in Table 2.

No Evidence for SNP × SNP Interaction or Haplotypic Effects

Concluding that the data support the presence of at least 3 independently significant association signals in the *SNCA* region, we went on to explore whether these SNPs collectively affect PD risk in a simple additive manner or whether they show evidence of interaction or haplotypic effects. We tested logistic regression models that included all combinations of pairwise interactions between the 3 SNPs, but no interaction passed a multitest significance threshold in the discovery dataset (smallest $p = 0.04$ for rs356182*rs763443).

The combined effect of multiple risk variants could hypothetically be sensitive to phase, as *cis* acting interactions would occur only between alleles on the same chromosome. Under such a scenario, risk profiling based on haplotypes should capture more of the effect than a standard multi-SNP approach. We therefore estimated individual haplotypes for all patients and evaluated the effect of haplotype counts on risk by logistic regression. Expectedly, several haplotypes were strongly associated with PD, yet there was no overall improvement in model fit compared to the 3-SNP model in a likelihood ratio test ($p = 0.75$, Akaike information criterion 33580 vs 33574).

Individual Stratification of SNCA-Specific Genetic Burden

Genetic risk scores (GRSs) are frequently calculated as a measure of cumulative genetic burden, for example, based on allele counts and log odds ratios (ORs) across all GWAS top hits. The identification of 3 significant independent *SNCA* signals, with 6 alleles modulating risk in each individual, suggested that a similar approach could be applied on a single-locus level to estimate an individual's total *SNCA*-specific genetic burden. We therefore estimated allele effect sizes in the discovery data and applied these as weights to calculate *SNCA*-specific GRSs in the independent replication dataset. We thus avoided the common problem of overfitting that arises when the same dataset is used both to generate and to evaluate the GRSs. Stratifying the data into GRS quintiles demonstrates a sliding scale of *SNCA*-specific risk in the replication data (Fig 3). In logistic regression analysis including sex and top 5 PCs as covariates, the difference in risk between the bottom and top 20% of individuals corresponded to about 2-fold (OR = 1.99, 95% confidence interval [CI] = 1.78–2.23). Area under the curve (AUC) for the 3-SNP score was 0.57, significantly higher than a score based on only the 2 known top signals (DeLong test $p = 0.00068$), and the proportions of patients across quintiles 1 to 5 were 0.42, 0.48, 0.51, 0.55, and 0.59.

An alternative to stratification based on risk scores is haplotype analysis. The maximum genetic burden would be expected for a haplotype carrying the high-risk allele for all 3 associated SNPs (rs2870004-rs356182-rs763443 = TGC). Interestingly, this haplotype is very infrequent, occurring on an estimated 0.34% of PD and 0.18% of control chromosomes in our data. Assuming a log-additive 3-SNP model, the combined effect of 3 risk alleles adds up to an estimated OR of 1.94, corresponding well with the effect size of the haplotype against all others in logistic regression (OR = 2.05, 95% CI = 1.23–3.41). Only 1 individual,

a PD patient, was homozygous for this haplo-type in our dataset. This combination would be exceedingly rare in the population but carries an estimated PD risk of approximately 4-fold, comparable to the effect of protein-coding high-risk variants in *GBA*.²⁹

Evidence Favors Top Hit rs356182 as Causal

We further aimed to clarify the potential contribution of recently proposed causal variants to PD risk. Rhinn et al linked the pathogenic mechanism to a transcript isoform of *SNCA* with an extended 3'UTR, suggesting that the risk allele of rs356165 affects secondary mRNA structure, miRNA interaction, and translation.¹⁸ In a subsequent paper by Soldner et al, gene expression in induced pluripotent stem cell–derived neurons was significantly associated with rs356168 alleles, introduced by the CRISPR/Cas9 gene editing technique and claimed by the authors to alter the binding affinity of an enhancer element in *SNCA* intron 4.¹⁹

Neither of these proposed functional SNPs (rs356165¹⁸ and rs356168¹⁹) emerged as an independent signal in our discovery analysis, yet the variants are in moderate LD with the top hit rs356182 (see Fig 1 and Table 3). We note, however, that rs356165 was not present in the quality-filtered imputed dataset, and we considered the proxy rs356219 (perfect LD $r^2 = 1$ with functional candidate rs356165, 1000 Genomes EUR) to represent this variant. Conditioning on rs356219 and rs356168 respectively in logistic regression left the same 3 SNPs significant, strongly suggesting that neither of the proposed functional SNPs can fully account for the top association signal.

We went on to explore the possible basis for a causal effect of rs356182, using prediction tools and publicly available genomic data. Two recently proposed computational prediction methods assign scores to noncoding variants based on the likelihood of functional consequences. The Combined Annotation and Depletion algorithm (CADD)³⁰ gives rs356182 a high score for a noncoding variant (Phred-scaled CADD = 10.68), compatible with a functional role (see Table 3). Furthermore, a LINSIGHT score of 10.0 is in line with the value expected for an enhancer region.³¹ In comparison, both algorithms give low scores to the proposed functional variants rs356165 and rs356168. rs356182 is located approximately 19kb downstream of *SNCA*. The genomic region surrounding the variant is not annotated with features indicating regulatory function such as histone modifications or DNase hypersensitivity in any of the major ENCODE cell lines, possibly explaining why its potential causal effects have thus far not been investigated in functional studies. Given that a high proportion of regulatory elements in the genome are tissue or cell type specific, we explored publicly available data from human brain generated by Assay for Transposase Accessible Chromatin with sequencing³² and Chromatin Immunoprecipitation sequencing for regulatory histone modification H3K27Ac.³³ Both datasets show a clear signal in the region surrounding rs356182, indicating a likely brain-specific regulatory element (Fig 2B). Furthermore, the Transfac database of predicted transcription factor binding motifs shows that the SNP disrupts a number of putative binding sites, including several that are conserved across species.

rs356182 Is Associated with a Specific 5'UTR Transcript Isoform of *SNCA*

Previously published brain eQTL data have failed to show robust associations between GWAS SNPs and gene-level *SNCA* expression.^{5,17,34} However, the 5' region of *SNCA* harbors multiple alternative transcription start sites, giving rise to distinct mRNA isoforms with variable 5'UTR. Hypothesizing that risk-relevant gene regulation, either pre- or posttranscriptional, may be specific to transcript variant, we performed eQTL analysis individually for each of the major 5'UTR isoforms in 2 existing large sequencing datasets from postmortem human frontal cortex and a third, independent sample set analyzed by qPCR.

CAGE peaks were created using the FANTOM5 annotated peaks phase 1 and 2,³⁵ which resulted in a total of 11 peaks corresponding to alternative transcription start sites. The majority of these are present at very low levels, however, and the 3 most abundant transcripts account for >95% of the total expression, corresponding to each of the *SNCA* 5'UTR transcript variants listed in the RefSeq database (accession IDs NM_000345, NM_001146054, and NM_001146055; see Fig 2A). In further analyses, we considered only these 3 transcripts, which also allowed for direct comparison with RNAseq and qPCR data.

Results from eQTL analyses are shown in Table 4. An association between rs356182 and NM_001146054 passed our multitest significance threshold in the CAGE-seq data, and was replicated with the same direction of effect at nominal *p* values in both the RNAseq and qPCR datasets (see Fig 4). The same SNP also showed a trend toward association with higher levels of total *SNCA* in CAGEseq data, but this effect was not observed in the other datasets. This led us to consider whether the variant might affect the relative proportion of NM_001146054 as a fraction of total *SNCA* rather than independently increasing the absolute level. Even stronger association was seen between rs356182 and the relative fraction of NM_001146054, with *p* values passing the multitest threshold in all 3 datasets (see Table 4). The other 2 index SNPs showed suggestive eQTL associations in single datasets, but no consistent signals across all 3 experiments. Substituting rs356182 with the proposed functional variants rs356168 and rs356165 gave similar results, as expected based on LD, yet with weaker β and *p* values for the relevant associations (data not shown). We note that whereas NM_000345 is the most abundant transcript in all datasets, the relative proportion of the 2 others vary, with NM_001146054 fraction ranging from 17% in RNAseq to 41% in qPCR data, most likely reflecting methodological differences.

Suggestive Evidence for an Association between rs356182 and CSF α -Synuclein Levels

In PPMI data, the high-risk G-allele of rs356182 was associated with higher CSF levels of α -synuclein at baseline when controlling for disease status (see Table 4). Repeating analysis separately for PD patients and controls, we found that this association was almost entirely driven by the control group ($\beta = 0.14$, *p* = 0.018), although the trend went in the same direction for patients ($\beta = 0.05$, *p* = 0.19).

Discussion

Genetic studies of sporadic PD and other complex neurological disorders have in recent years been dominated by GWASs, primarily aimed toward detecting the largest possible number of susceptibility loci. Studies involving profiling of cumulative genetic risk have typically used genome-wide risk scores designed to capture the total genetic load. Moving toward an era of targeted therapeutics, we anticipate that more sophisticated approaches will be warranted in the efforts to leverage genomic data for tailored, personalized therapy. In the present study, we performed the largest dedicated genetic study to date of *SNCA*, a pivotal gene in PD pathogenesis, refining our understanding of the allelic heterogeneity and exploring the potential for locus-specific individual risk stratification.

Although the existence of 2 independent *SNCA* association signals for PD has been established through several GWASs, we now report a third signal of genome-wide significance, with potential implications both for future efforts to understand and potentially modify gene regulation, and for the ability to accurately stratify individuals based on *SNCA*-related genetic burden. It is highly plausible that the great clinical heterogeneity observed across individual PD patients to some extent reflects differences in underlying genetic architecture, with potential relevance for treatment.^{36–38} As an example, a recent study reported that the β 2-adrenoreceptor is a regulator of *SNCA*, and that β 2 agonist use is negatively associated with PD, suggesting clinical trials of these drugs as a potential neuroprotective therapy.³⁹ Showing that *SNCA*-related risk varies by more than 2-fold across quintiles, and that rare homozygotes for the maximum-risk haplo-types may carry 4-fold PD susceptibility compared to the general population, we suggest that patients in the high-risk end of this spectrum would be the most suitable for such clinical trials.

A strength in the present study is that the risk score was validated in an independent sample set, separate from the data used to estimate allele effect sizes. This approach avoids the problem of overfitting and indicates that such a score may be applicable also in the context of clinical trials. With growing insight into the pathogenic mechanisms of PD and the exact genes implicated at different GWAS loci, we envisage that the concept of a locus-specific risk score could be refined and expanded to capture genetic burden across a number of genes in pathway-specific scores, aiming to profile specific aspects of neuronal vulnerability in the individual patient. We note however, that the locus-specific score is not designed to optimize AUC for disease prediction, where a genetic risk score capturing all relevant variability genome wide is more appropriate.

Interestingly, the pattern of *SNCA* association across PD and DLB shows that the effect of each independent signal is highly phenotype specific, as the strongest signal in PD, represented by rs356182 and located 3' of the gene, is absent in DLB, whereas the secondary 5' signal is significant in both disorders. This suggests that risk alleles may affect regulatory mechanisms that are highly specific to certain biological contexts or distinct populations of vulnerable neurons.

Of all common noncoding SNPs in the genome, the *SNCA* top hit rs356182 carries the largest effect size and explains the largest proportion of variance in PD liability. Unlike

many other GWAS signals, there is no tightly linked group of correlated SNPs showing similarly strong association statistics, and conditional analyses strongly favor rs356182 over the proposed functional SNPs rs356165 and rs356168. We show from prediction tools and publicly available data that rs356182 is located in a likely brain-specific regulatory element overlapping conserved predicted transcription factor binding sites (see Fig 2B). Functional studies of this putative regulatory region are highly warranted.

Although precise mechanisms underlying GWAS signals for common complex disorders are often unknown, a high proportion of risk variants demonstrate an association with mRNA levels in disease-relevant tissue, indicating an effect on gene regulation.³⁴ Such an eQTL effect is particularly plausible in the case of *SNCA*, where the dose-dependent pathogenic effect of increased expression has been demonstrated in monogenic PD caused by genomic multiplications.² *SNCA* GWAS hits have been investigated as potential brain eQTLs in a number of previous studies without convincing results, including the large Genotype Tissue Expression³⁴ and BRAINEAC⁴⁰ databases, as well as a recent negative follow-up study of the proposed functional variant rs356168.⁴¹ This suggests that regulation is complex, and refined analyses may be needed to detect disease-relevant mechanisms. Previous studies have investigated *SNCA* splice variants affecting coding exons,⁴² but variation in 5'UTR expression has to our knowledge not been studied in relation to PD susceptibility and risk SNPs. We report a novel association between rs356182 and the NM_001146054 transcript variant, which is consistent across 3 independent datasets. The association is stronger for the abundance of the isoform relative to total *SNCA* than for its absolute level. Although the underlying mechanism linking genetic variation to transcription remains to be determined, our finding highlights a potential key role of specific transcription start sites and corresponding 5'UTR transcript variants. Our study investigated expression in human frontal cortex only, and it would be interesting to see whether the effect is also present in other brain regions. A mechanism specific to cortex would not fit well with an SNP that is strongly associated with PD, but not with DLB, which would lead us to hypothesize that this eQTL effect should be at least as strong in regions affected at an earlier stage of PD.

A suggestive positive association was seen between the high-risk allele of rs356182 and CSF levels of total α -synuclein when controlling for disease status in the PPMI cohort. This finding must be interpreted with caution, however, and has not been replicated. The result may appear paradoxical, as CSF α -synuclein has consistently been found to be lower in PD patients compared to controls,⁴³ as is also the case in the PPMI data.⁴⁴ Of note, the association with rs356182 was driven almost entirely by a QTL effect in the healthy control group, which might suggest 2 independent forces influencing CSF levels of α -synuclein: genetic variability in the *SNCA* region and the disease process in itself, the latter leading to lower α -synuclein levels. Further research into the relationship between genetic variation and α -synuclein in CSF is warranted.

In conclusion, we have conducted the most comprehensive genetic study to date of the *SNCA* locus in PD, taking several different approaches to elucidate how common variability in this important region contributes to disease risk. We identified a novel independent genetic association signal for PD risk at genome-wide significance level, demonstrated the combined effect of multiple risk variants on *SNCA*-related PD susceptibility, and reported

novel QTL associations with the top variant rs356182. Our study highlights how in-depth exploration of major disease loci will likely be instrumental in harvesting the translational potential of GWASs for future clinical trials and tailored treatment in complex neurological disorders.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

L.P. is supported by the Norwegian Health Association. V.B.-S. and M.T. are funded by grants from the South-Eastern Norway Regional Health Authority. M.T. and M.L. are supported by the Research Council of Norway. This research was supported by the Intramural Research Program of the NIH National Institute on Aging. W.D.J.v.d.B. is funded by grants from ZonMW Memorabel, Stichting Parkinson Fonds, and Alzheimer Nederland-LECMA, with additional funding from Hoffmann-La Roche (contract research and research collaboration) and from Lysosomal Therapeutics (contract research and research collaboration).

For up-to-date information on the PPMI, visit www.ppmi-info.org. PPMI—a public–private partnership—is funded by the Michael J. Fox Foundation for Parkinson’s Research, and funding partners including AbbVie, Avid Radiopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Sanofi Genzyme, Servier, Takeda, Teva, and UCB.

We thank all study participants and investigators providing data for the study; and O. Andreassen and the Dem-Gen consortium for genotyping of the Oslo cohort.

References

1. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson’s disease. *Science* 1997;276:2045–2047. [PubMed: 9197268]
2. Singleton AB, Farrer M, Johnson J, et al. alpha-Synuclein locus triplication causes Parkinson’s disease. *Science* 2003;302:841. [PubMed: 14593171]
3. Simon-Sanchez J, Schulte C, Bras JM, et al. Genome-wide association study reveals genetic risk underlying Parkinson’s disease. *Nat Genet* 2009;41:1308–1312. [PubMed: 19915575]
4. Do CB, Tung JY, Dorfman E, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson’s disease. *PLoS Genet* 2011;7:e1002141. [PubMed: 21738487]
5. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson’s disease. *Nat Genet* 2014;46:989–993. [PubMed: 25064009]
6. Pankratz N, Beecham GW, DeStefano AL, et al. Meta-analysis of Parkinson’s disease: identification of a novel locus, RIT2. *Ann Neurol* 2012;71:370–384. [PubMed: 22451204]
7. Satake W, Nakabayashi Y, Mizuta I, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson’s disease. *Nat Genet* 2009;41:1303–1307. [PubMed: 19915576]
8. Spencer CC, Plagnol V, Strange A, et al. Dissection of the genetics of Parkinson’s disease identifies an additional association 5’ of *SNCA* and multiple associated haplotypes at 17q21. *Hum Mol Genet* 2011; 20:345–353. [PubMed: 21044948]
9. Hamza TH, Zabetian CP, Tenesa A, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson’s disease. *Nat Genet* 2010;42:781–785. [PubMed: 20711177]
10. International Parkinson’s Disease Genomics Consortium, Wellcome Trust Case Control Consortium. A two-stage meta-analysis identifies several new loci for Parkinson’s disease. *PLoS Genet* 2011;7: e1002142. [PubMed: 21738488]

11. Chang D, Nalls MA, Hallgrimsdottir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* 2017;49:1511–1516. [PubMed: 28892059]
12. Chu Y, Kordower JH. The prion hypothesis of Parkinson's disease. *Curr Neurol Neurosci Rep* 2015;15:28. [PubMed: 25868519]
13. Rodriguez JA, Ivanova MI, Sawaya MR, et al. Structure of the toxic core of alpha-synuclein from invisible crystals. *Nature* 2015;525: 486–490. [PubMed: 26352473]
14. Spillantini MG, Schmidt ML, Lee VM, et al. Alpha-synuclein in Lewy bodies. *Nature* 1997;388:839–840. [PubMed: 9278044]
15. Olanow CW, Kordower JH. Targeting alpha-synuclein as a therapy for Parkinson's disease: the battle begins. *Mov Disord* 2017;32: 203–207. [PubMed: 28218461]
16. Guerreiro R, Ross OA, Kun-Rodrigues C, et al. Investigating the genetic architecture of dementia with Lewy bodies: a two-stage genome-wide association study. *Lancet Neurol* 2018;17:64–74. [PubMed: 29263008]
17. Blauwendraat C, Francescato M, Gibbs JR, et al. Comprehensive promoter level expression quantitative trait loci analysis of the human frontal lobe. *Genome Med* 2016;8:65. [PubMed: 27287230]
18. Rhinn H, Qiang L, Yamashita T, et al. Alternative alpha-synuclein transcript usage as a convergent mechanism in Parkinson's disease pathology. *Nat Commun* 2012;3:1084. [PubMed: 23011138]
19. 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. [PubMed: 27096366]
20. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet* 2016;48:1284–1287. [PubMed: 27571263]
21. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;48: 1279–1283. [PubMed: 27548312]
22. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004;74:765–769. [PubMed: 14997420]
23. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 2005;95: 221–227. [PubMed: 16077740]
24. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7. [PubMed: 25722852]
25. Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 2002;70:124–141. [PubMed: 11719900]
26. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* 2006;78: 629–644. [PubMed: 16532393]
27. Stegle O, Parts L, Piipari M, et al. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* 2012;7:500–507. [PubMed: 22343431]
28. Nalls MA, Bras J, Hernandez DG, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. *Neurobiol Aging* 2015;36:1605.e7–1605.e12.
29. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med* 2009; 361:1651–1661. [PubMed: 19846850]
30. Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–315. [PubMed: 24487276]
31. Huang YF, Gulko B, Siepel A. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. *Nat Genet* 2017;49:618–624. [PubMed: 28288115]
32. Fullard JF, Giambartolomei C, Hauberg ME, et al. Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. *Hum Mol Genet* 2017;26:1942–1951. [PubMed: 28335009]
33. Vermunt MW, Reinink P, Korving J, et al. Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Rep* 2014;9:767–779. [PubMed: 25373911]

34. GTEx Consortium, Battle A, Brown CD, et al. Genetic effects on gene expression across human tissues. *Nature* 2017;550:204–213. [PubMed: 29022597]
35. Fantom Consortium, RIKEN PMI, CLST (DGT). A promoter-level mammalian expression atlas. *Nature* 2014;507:462–470. [PubMed: 24670764]
36. Escott-Price V, International Parkinson's Disease Genomics Consortium, Nalls MA, et al. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann Neurol* 2015;77:582–591. [PubMed: 25773351]
37. Fagan ES, Pihlstrom L. Genetic risk factors for cognitive decline in Parkinson's disease: a review of the literature. *Eur J Neurol* 2017;24: 561–e20. [PubMed: 28220571]
38. Pihlstrom L, Morset KR, Grimstad E, et al. A cumulative genetic risk score predicts progression in Parkinson's disease. *Mov Disord* 2016; 31:487–490. [PubMed: 26853697]
39. Mittal S, Bjornevik K, Im DS, et al. beta2-Adrenoreceptor is a regulator of the alpha-synuclein gene driving risk of Parkinson's disease. *Science* 2017;357:891–898. [PubMed: 28860381]
40. Ramasamy A, Trabzuni D, Guelfi S, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* 2014;17:1418–1428. [PubMed: 25174004]
41. Glenn OC, Tagliaferro L, Beach TG, et al. Interpreting gene expression effects of disease-associated variants: a lesson from *SNCA* rs356168. *Front Genet* 2017;8:133. [PubMed: 28979294]
42. McCarthy JJ, Linnertz C, Saucier L, et al. The effect of *SNCA* 3' region on the levels of *SNCA* –112 splicing variant. *Neurogenetics* 2011;12:59–64. [PubMed: 21046180]
43. Mollenhauer B, Locascio JJ, Schulz-Schaeffer W, et al. alpha-Synuclein and tau concentrations in cerebrospinal fluid of patients presenting with parkinsonism: a cohort study. *Lancet Neurol* 2011; 10:230–240. [PubMed: 21317042]
44. Kang JH, Irwin DJ, Chen-Plotkin AS, et al. Association of cerebrospinal fluid beta-amyloid 1–42, T-tau, P-tau181, and alpha-synuclein levels with clinical features of drug-naïve patients with early Parkinson disease. *JAMA Neurol* 2013;70:1277–1287. [PubMed: 23979011]
45. Pankratz N, Wilk JB, Latourelle JC, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum Genet* 2009;124:593–605. [PubMed: 18985386]
46. Pihlstrom L, Axelsson G, Bjornara KA, et al. Supportive evidence for 11 loci from genome-wide association studies in Parkinson's disease. *Neurobiol Aging* 2013;34:1708.e7–1708.e13.
47. Nalls MA, Keller MF, Hernandez DG, et al. Baseline genetic associations in the Parkinson's Progression Markers Initiative (PPMI). *Mov Disord* 2016;31:79–85. [PubMed: 26268663]
48. Hernandez DG, Nalls MA, Ylikotila P, et al. Genome wide assessment of young onset Parkinson's disease from Finland. *PLoS One* 2012;7:e41859. [PubMed: 22911860]
49. Rosenthal LS, Drake D, Alcalay RN, et al. The NINDS Parkinson's disease biomarkers program. *Mov Disord* 2016;31:915–923. [PubMed: 26442452]
50. Bandres-Ciga S, Price TR, Barrero FJ, et al. Genome-wide assessment of Parkinson's disease in a Southern Spanish population. *Neurobiol Aging* 2016;45:213.e3–213.e9.

FIGURE 1:

The *SNCA* locus with association results and linkage disequilibrium. The top and bottom left panels show regional Manhattan plots for 3 rounds of stepwise conditional analysis in the discovery data and degree of linkage disequilibrium (LD) with the lead single nucleotide polymorphism (SNP). Note the lack of SNPs in high LD with rs356182 and the different scale of the y-axis across plots. At the bottom right, the pattern of LD between these SNPs as well as the 2 proposed functional variants rs356165 and rs356168 is illustrated as an LD plot. The numbers represent r^2 values, and darker shade represents higher D' . The figure

was made using LocusZoom (locuszoom.org) and Haploview (www.broadinstitute.org/haploview). [Color figure can be viewed at www.annalsofneurology.org]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 2: Transcript variants at the 5' end of *SNCA* and genomic context of rs356182. (A) Figure zooming in on the 5' end of the *SNCA* gene, showing variable 5' untranslated region (UTR) corresponding to the 3 individual RefSeq transcript variants investigated in expression quantitative trait locus analysis. Translation starts at the ATG codon 26 base pairs into exon 2. (B) Illustration showing the chromosomal location of rs356182 downstream of *SNCA*. Note the 3' end of the *SNCA* gene at top right. The variant is located within a putative brain-specific regulatory element identified by H3K27Ac chromatin immunoprecipitation sequencing in human brain, showing no similar signal in 7 cell lines (Layered H3K27Ac) from ENCODE. Furthermore, rs356182 falls within an open chromatin

region differentially accessible across neuronal and nonneuronal cells, and overlaps a conserved predicted transcription factor binding site (TFBS) from the Transfac database. [Color figure can be viewed at www.annalsofneurology.org]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 3:

Increasing susceptibility across *SNCA* risk score quintiles. The figure shows odds ratios of risk quintiles incorporating the independent effect of 3 significant association signals at the *SNCA* locus in a combined score, assessed by logistic regression. Odds ratios are shown relative to the first quintile. Error bars represent 95% confidence intervals. Note that score performance was assessed in the replication dataset, which is independent from the discovery data that were used to generate the allele weights for the scoring algorithm.

FIGURE 4:

Expression of NM_001146054 by number of rs356182 risk alleles. Note the increased expression of NM_001146054 for each rs356182 risk allele across all 3 datasets. For comparison of effect sizes across gene expression analysis methods, mRNA expression levels have been converted to z scores. CAGE = cap analysis gene expression; PCR = polymerase chain reaction. [Color figure can be viewed at www.annalsofneurology.org]

TABLE 1.

Datasets Included in the Study

Dataset	PD Patients			Controls		Genotyping Array	
	n	Female, %	Age at Onset, yr, Mean (SD)	Diagnostic Criteria	n	Female, %	Age at study, yr, Mean (SD)
Discovery							
IPDGC NeuroX ²⁸	5,542	35.5	61 (13)	UKPDBB or neurologist diagnosis	5,866	44.2	64 (15)
Replication							
phs000089.v3.p2, NINDS ³	916	40.3	58 (13)	UKPDBB	801	58.1	59 (16)
phs000126.v1.p1, CIDR ⁴⁵	864	40.0	60 (16)	Modified UKPDBB ^a	856	60.2	55 (13)
phs000196.v2.p1, NGRC ⁹	1,972	32.4	58 (12)	UKPDBB	1,984	61.3	70 (14)
Oslo ⁴⁶	474	36.3	56 (11)	Modified UKPDBB ^a	464	42.5	62 (11)
PPMI ⁴⁷	314	32.2	60 (8)	PPMI criteria	166	35.5	61 (11)
Finland ⁴⁸	386	45.9	55 (6)	Neurologist diagnosis	493	78.9	92 (4)
Germany ³	741	39.7	56 (12)	UKPDBB	944	57.5	47 (12)
Harvard Biomarker Study	541	34.6	66 (10)	Neurologist diagnosis	473	61.7	79 (9)
Parkinson's Disease Biomarker Program ⁴⁹	543	38.5	64 (9)	Neurologist diagnosis	284	51.1	62 (11)
Spain ⁵⁰	210	35.2	67(11)	UKPDBB	171	55.0	58 (14)
Sum replication	6,961				6,636		
Total	12,503				12,502		

Age data were missing for 2% of patients and 0.3% of controls in the replication sample.

^aModified UKPDBB criteria allow for affected first-degree relatives.

CIDR = Center for Inherited Disease Research; IPDGC = International Parkinson Disease Genomics Consortium; NGRC = NeuroGenetics Research Consortium; NINDS = National Institute of Neurological Disorders and Stroke; PD = Parkinson disease; PPMI = Parkinson's Progression Marker Initiative; SD = standard deviation; UKPDBB = United Kingdom Parkinson's Disease Brain Bank criteria.

TABLE 2.

Association Analyses of *SNCA* Risk Variants

SNP	rs356182	rs763443	rs2870004
Alleles, effect/reference	A/G	C/T	A/T
b37 chr4 base position	90626111	90819961	90471245
Discovery, n cases/controls = 5,542/5,886			
Effect allele frequency	0.37	0.52	0.21
Stepwise conditional analysis			
Conditioned on	Not conditioned	rs356182	rs356182 and rs763443
OR (95% CI)	0.78 (0.74–0.82)	1.20 (1.12–1.28)	0.88 (0.82–0.93)
<i>P</i>	2.7e-20	4.1e-10	0.00010
Combined model			
OR (95% CI)	0.70 (0.65–0.74)	1.21 (1.14–1.28)	0.88 (0.82–0.94)
<i>P</i>	9.7e-31	2.9e-10	0.00010
Independent replication, n cases/controls = 6,961/6,636			
OR (95% CI)	0.67 (0.63–0.71)	1.15 (1.09–1.22)	0.88 (0.83–0.94)
<i>P</i>	2.8e-42	4.5e-07	4.2e-05
Combined analysis discovery + replication			
OR (95% CI)	0.68 (0.66–0.71)	1.17 (1.13–1.22)	0.88 (0.84–0.92)
<i>P</i>	1.1e-69	4.9e-15	3.0e-08

The table lists variants passing our SNPSpD-adjusted significance threshold of $p < 0.00023$ using stepwise forward conditional analysis in the discovery phase. Stepwise conditional results are followed by statistics for the joint model in discovery data, independent replication, and finally combined analysis in a logistic regression model including all 3 SNPs, with sex, study, and top 5 principal components as covariates. Genomic coordinates correspond to GRCh37, and directions of effect are shown for the nonreference allele.

CI = confidence interval; OR = odds ratio; SNP = single nucleotide polymorphism.

TABLE 3.

LD and Functional Prediction Scores of Statistical Top Hit and Proposed Causal SNPs

SNP	r^2 with rs356182 1000 Genomes/Present Study	LINSIGHT Score	Phred-Scaled CADD
rs356182	—	10.0	10.68
rs356165	0.76/0.81	0.055	4.58
rs356168	0.49/0.55	0.096	5.87

Note the moderate LD values indicating that rs356182 would not be a likely top hit merely by virtue of tagging any of the other proposed functional SNPs. Prediction tools favor a hypothesis of rs356182 as functional in itself.

CADD = Combined Annotation and Depletion algorithm; LD = linkage disequilibrium; SNP = single nucleotide polymorphism.

TABLE 4.

Expression Quantitative Trait Locus Analyses

Expression Outcome	SNP	Dataset	β	SE	P
Total <i>SNCA</i>	rs2870004	CAGEseq	0.095	0.11	0.40
		RNAseq	-0.018	0.16	0.91
		qPCR	0.052	0.20	0.80
	rs356182	CAGEseq	0.276	0.10	0.0088
		RNAseq	-0.191	0.14	0.17
		qPCR	0.179	0.20	0.38
NM_000345	rs763443	CAGEseq	-0.021	0.12	0.86
		RNAseq	0.016	0.09	0.86
		qPCR	-0.019	0.18	0.91
	rs2870004	CAGEseq	0.033	0.11	0.77
		RNAseq	-0.070	0.16	0.66
		qPCR	-0.036	0.20	0.86
NM_001146054	rs356182	CAGEseq	0.086	0.10	0.40
		RNAseq	-0.103	0.14	0.46
		qPCR	-0.176	0.20	0.38
	rs763443	CAGEseq	-0.125	0.09	0.17
		RNAseq	-0.024	0.12	0.84
		qPCR	0.018	0.18	0.92
NM_001146055	rs2870004	CAGEseq	0.283	0.12	0.018
		RNAseq	-0.033	0.16	0.83
		qPCR	0.179	0.20	0.36
	rs356182 ^a	CAGEseq	0.445	0.11	7.0e-05
		RNAseq	0.295	0.14	0.034
		qPCR	0.559	0.19	0.0055
NM_001146055	rs763443	CAGEseq	0.040	0.09	0.67
		RNAseq	0.211	0.12	0.08
		qPCR	-0.084	0.17	0.63
NM_001146055	rs2870004	CAGEseq	-0.103	0.11	0.33
		CAGEseq	-0.103	0.11	0.33

Expression Outcome	SNP	Dataset	β	SE	P
NM_001146054 fraction of total <i>SNCA</i>	rs356182	RNAseq	0.051	0.16	0.75
		qPCR	-0.379	0.19	0.046
	rs763443	CAGEseq	0.108	0.10	0.27
		RNAseq	-0.313	0.14	0.024
	rs763443	qPCR	0.128	0.19	0.49
		CAGEseq	0.087	0.09	0.31
	rs2870004	RNAseq	-0.041	0.12	0.73
		qPCR	0.259	0.16	0.12
	rs356182 ^a	CAGEseq	0.336	0.11	0.0020
		RNAseq	-0.029	0.15	0.85
CSF α -synuclein	rs356182 ^a	qPCR	0.156	0.19	0.42
		CAGEseq	0.439	0.10	1.7e-5
	rs763443	RNAseq	0.494	0.13	2.9e-4
		qPCR	0.584	0.19	0.0032
	rs2870004	CAGEseq	0.048	0.09	0.58
		RNAseq	0.263	0.12	0.024
	rs356182 ^a	qPCR	-0.084	0.17	0.62
		PPMI	0.071	0.03	0.038
	rs763443	PPMI	0.077	0.03	0.015
		PPMI	-0.004	0.03	0.89

The table shows results from linear regression of risk variant association with mRNA levels and CSF α -synuclein. For comparison of effect sizes across gene expression analysis methods, mRNA expression levels have been converted to z scores. Effect alleles correspond to Table 1.

^aSignificant associations, see main text.

CAGEseq = cap analysis gene expression sequencing; CSF = cerebrospinal fluid; PPMI = Parkinson's Progression Marker Initiative; qPCR = quantitative polymerase chain reaction; RNAseq = RNA sequencing; SE = standard error; SNP = single nucleotide polymorphism.