



# Genomic convergence of locus-based GWAS meta-analysis identifies *AXIN1* as a novel Parkinson's gene

Mohammad Saeed<sup>1</sup>

Received: 17 May 2018 / Accepted: 7 June 2018 / Published online: 19 June 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Parkinson's disease (PD) is a common, disabling neurodegenerative disorder with significant genetic underpinnings. Multiple genome-wide association studies (GWAS) have been conducted with identification of several PD loci. However, these only explain about 25% of PD genetic risk indicating that additional loci of modest effect remain to be discovered. Association clustering methods such as gene-based tests are more powerful than single-variant analysis for identifying modest genetic effects. Combined with the locus-based algorithm, OASIS, the most significant association signals can be homed in. Here, two dbGAP GWAS datasets (7415 subjects (2750 PD and 4845 controls) genotyped for 0.78 million SNPs) were analyzed using combined clustering algorithms to identify 88 PD candidate genes in 24 loci. These were further investigated for gene expression in substantia nigra (SN) of PD and control subjects on GEO datasets. Expression differences were also assessed in normal brains SN versus white matter on BRAINEAC datasets. This genetic and functional analysis identified *AXIN1*, a key regulator of Wnt/ $\beta$ -catenin signaling, as a novel PD gene. This finding links PD with inflammation. Other significantly associated genes were *CSMD1*, *CLDN1*, *ZNF141*, *ZNF721*, *RHOT2*, *RICTOR*, *KANSL1*, and *ARHGAP27*. Novel PD genes were identified using genomic convergence of gene-wide and locus-based tests and expression studies on archived datasets.

**Keywords** Parkinson's · Locus-based test · Genome-wide association study · Gene-based tests · Gene expression

## Introduction

Parkinson's disease (PD) is characterized by resting tremor, rigidity, bradykinesia associated with loss of dopaminergic neurons in the substantia nigra (SN), and the neuronal presence of Lewy bodies. PD affects about 1% of people above the age of 60 years, making it the second most common neurodegenerative disease. Identification of genes for familial PD viz. *SNCA*, *PARKIN*, *PINK1*, *DJ-1*, *VPS35*, *LRRK2*, and *TMEM230* have provided insights into disease pathogenesis (Hernandez et al. 2016, Deng et al. 2016). Recently, several large genome-wide association studies (GWAS) have been

conducted to identify loci for sporadic PD. Yet majority of sporadic PD remains elusive with common variants explaining less a quarter of the PD heritability (Keller et al. 2012). This implies that genetic variants with modest effects on PD susceptibility remain to be discovered. GWAS is a powerful tool to map genes but is hampered by the large number of positive associations and multiple testing correction issues. Differentiating signal from noise by eliminating false-positive and negative associations is a huge challenge.

True genetic associations are accompanied by signals from surrounding markers due to linkage disequilibrium (LD). Recently, this observation was expanded upon by the development of a locus-based clustering algorithm for GWAS (Saeed, 2017). OASIS (Objective Assimilation of SNPs Interacting in Synchrony) is an algorithm that identifies and assimilates SNPs of even modest significance into clusters based on the LD principle (Saeed, 2017). Analogous to gene-based association tests, OASIS parses the GWAS data into physical proximity clusters. Yet OASIS is unique, in that it takes into account the entire association signal at each locus rather than assigning the most significant or a weighted *p* value to a gene (Saeed, 2017; Li et al. 2010). This enhances

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00251-018-1068-0>) contains supplementary material, which is available to authorized users.

✉ Mohammad Saeed  
msaeed.immunocure@gmail.com

<sup>1</sup> Consultant Rheumatology and Immunogenetics, ImmunoCure, Clinic and Lab, Suite 116, 1st Floor, The Plaza, 2-Talwar, Main Clifton Road, Karachi, Pakistan

the possibility of reducing statistical skewing by low-frequency alleles and consequent false-positive associations (type I error) as well as false negatives due to stringent corrections such as Bonferroni (type II error). OASIS has previously been used to identify genes for lupus and diabetic nephropathy (Saeed 2017; Saeed 2018).

Here, OASIS was applied to two independent PD database of Genotypes and Phenotypes (dbGAP) GWAS datasets (Simón-Sánchez et al. 2009; Pankratz et al. 2009) in a meta-analysis to identify significant PD loci. Candidate genes in OASIS loci were further evaluated using gene-based testing on both PD datasets. Significantly associated genes were then tested for gene expression in brain tissue using GEO and BRAINEAC datasets. These composite genomic and functional analyses led to the identification of nine novel PD candidate genes.

## Methods

### Datasets

GWAS datasets were obtained online from the publicly available dbGAP repository. Meta-analysis of two PD datasets, pha002865 (db1) (Simón-Sánchez et al. 2009) and pha002868 (db2) (Pankratz et al. 2009) was conducted using OASIS. The screening dataset, db1, consisted of 1713 PD cases and 3978 controls genotyped for 463,185 single nucleotide polymorphisms (SNPs) (Simón-Sánchez et al. 2009). The second dataset, db2, consisted of 857 familial PD cases and 867 controls (Pankratz et al. 2009) genotyped for 318,551 SNPs. Most significant SNPs in overlapping OASIS loci in the two datasets (db1, db2) were tested for single-variant replication in three separate datasets. These replication datasets included the dbGAP dataset pha002840 (db3) consisting of 443 discordant PD sib pairs and genotyped for 174,945 SNPs (Maraganore et al. 2005), the PDGene meta-data (db4) (Lill et al. 2012), and the publically available screening dataset of a recent large GWAS meta-analysis (db5) (Chang et al. 2017). Figure S2 shows the flow chart for the study methodology.

### Locus-based analysis

The OASIS algorithm and software have been described in detail previously (Saeed 2017). Briefly, the OASIS algorithm sums the  $-\log [P]$  values of all significant SNPs ( $P \leq 0.05$ ) in a locus cluster of 200 kb to generate an OASIS score. This process is iterated across each GWAS dataset. The 200-kb cut off has previously been used to define a locus cluster and is the default value in the OASIS software, though it can be altered to yield a variety of analyses (Saeed 2017). OASIS calculates the 3-sigma ( $3\sigma$ ; three standard deviations or a value  $\geq 99.7\%$  of the data) cut offs for the  $-\log [P]$  values and the

OASIS scores, structuring the GWAS meta-analysis data in to two axes ( $-\log [P]$  and OASIS) and four groups viz quadrants A–D (Figure S1). Quadrant A loci crossed the  $3\sigma$  cut offs on both axes; quadrant B loci only on  $-\log [P]$  and quadrant C loci only on OASIS scores. Quadrant D loci failed to meet the  $3\sigma$  cut offs on either axes. OASIS compiles the overlapping (within 2 Mb distance)  $3\sigma$  significant regions in two GWAS datasets into a single table (Saeed 2017). The SNIPPER software was used to map genes on genome build hg19 to the OASIS loci. Positional information for each OASIS locus was used to define the chromosomal location for the gene search. The gene list for all OASIS loci was used as a “seed list” for subsequent gene-based analysis.

### Gene-based testing

Candidate genes in the OASIS seed list were evaluated for gene-based association using GATES as implemented in the KGG software (Li et al. 2010). SNPs were mapped onto genes according to positional information from the NCBI GRCh37 database, and SNPs within 10 kb upstream and 10 kb downstream of each gene were included as well. This analysis is distinctly different from the locus-based clustering approach incorporated in OASIS (Saeed 2017; Li et al. 2010). Mostly the gene-set enrichment analysis (GSEA) algorithms map SNPs to genes, assign a weighted or the most significant  $P$  value to each gene, and then iteratively process the genes for association with disease (Li et al. 2010).

### Gene expression analysis

Genomic convergence of results from multiple data types, such as genetic association and expression data, implies biological relationship of genes with disease. Gene expression in SN of PD patients ( $n = 16$ ) versus controls ( $n = 9$ ) was assessed using Gene Expression Omnibus (GEO) dataset GDS2821. The Affymetrix Human Genome U133 Plus 2.0 Array was used for gene expression analysis in this dataset. Genes that showed significant gene-wide association in GATES analysis were tested for expression differences in the GEO dataset. One-tailed Student's  $t$  test of unequal variance was used to calculate statistical significance of expression levels for each gene between PD cases and controls. The genes that showed statistical significance were evaluated for gene expression in SN versus white matter (WM) in normal brains using the BRAINEAC datasets. BRAINEAC integrates whole genome genotype and transcript expression data from 134 normal human brain samples of 10 brain regions. Fold change was calculated as  $2^{\text{SN}}/2^{\text{WM}}$  and the statistical significance of expression levels using the one-tailed Student's  $t$  test of unequal variance. Expression quantitative trait locus (eQTL) analysis was done using BRAINEAC for all

significant SNPs in db1 and db2, at locus 22 containing the gene, *AXINI*. Corresponding LD data was obtained from LDlink.

### Protein network analysis

All genes that were significantly overexpressed or repressed in PD SN in the GEO dataset GDS2821 were used as input for STRING. To test the interaction of these candidate genes with known PD pathways, familial PD genes were also entered (*PARKIN*, *PINK1*, *DJ-1*, *VPS35*, *LRRK2*, *TMEM230*, *SNCA*, *MAPT*, *ATP13A2*, *FBXO7*, and *PLA2GB*). Gene pairs that were either co-expressed or involved in experimentally validated protein–protein interactions with at least a low score (0.15) were considered.

### Results

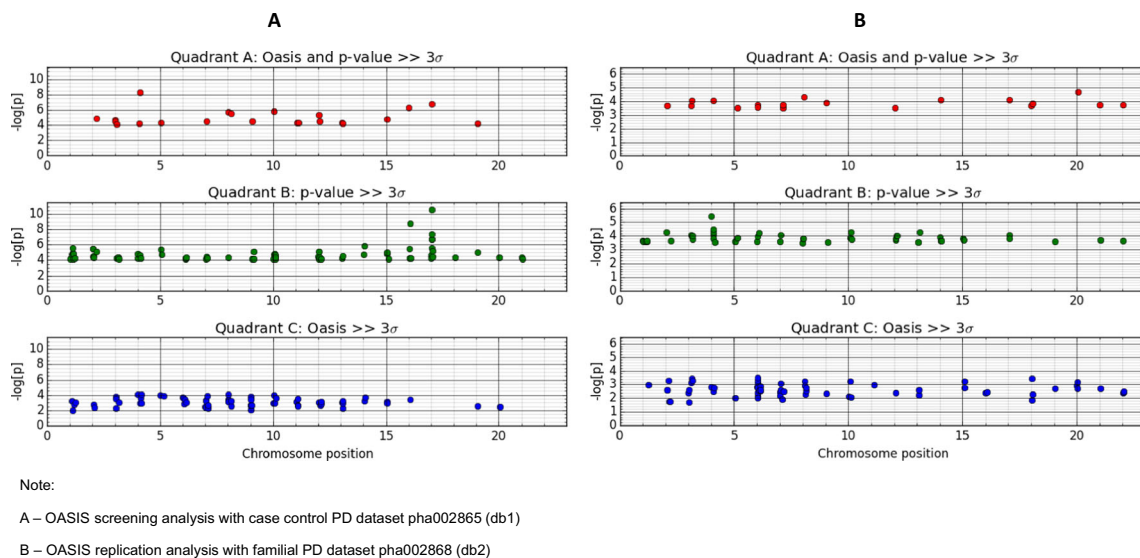
Initial OASIS analyses of db1 and db2 led to the identification of 289 loci that were significant (crossed the 3-sigma cut off) in these datasets (Fig. 1). Of these, 36 loci replicated in both datasets which were found to overlap into 24 unique loci (Table S1). The highest association signal was for rs2736990 ( $P = 5.7 \times 10^{-9}$ ) located at 4q22 containing *SCNA* and rs12185268 ( $P = 1.9 \times 10^{-7}$ ) at 17q21 containing *MAPT*. This was followed by rs11648673 ( $P = 4.8 \times 10^{-7}$ ) at a novel locus 16p13 (Table S1). All three loci were found in quadrant A suggestive of highly significant association with PD. In contrast, the known PD locus 4p16 containing *GAK/DGKQ/TMEM175* was found in quadrant B (rs11248060;  $P = 3.4 \times 10^{-6}$ ). Two further loci, 8p23 (locus 15) and 14q22

(locus 21), were also found in quadrant A in at least one dataset (Table S1).

Positional mapping using SNIPPER generated a list of 379 genes in these 24 overlapping OASIS loci (Table S2). These genes were used as a seed list for KGG gene-wide association analysis of db1 and db2, which identified 88 genes that were significantly associated with PD in at least one dataset (Table S1). Genes in three OASIS loci (6, 22, and 23) crossed the Benjamini and Hochberg (BH) correction in db1. *SCNA* ( $P = 1.0 \times 10^{-7}$ ;  $P_{BH} = 0.0012$ ) in locus 6 (4q22), *AXINI* ( $P = 1.0 \times 10^{-5}$ ;  $P_{BH} = 0.0176$ ) in locus 22 (16p13), and several genes in locus 23 (17q21) including *MAPT* ( $P = 2.1 \times 10^{-6}$ ;  $P_{BH} = 0.0048$ ) were significant (Table S3). No gene crossed the BH correction in db 2 (Table S4).

KGG analysis identified 57 genes in db1 and 52 genes in db2 which were nominally significant (Tables S3 and S4). Of these, 88 genes were unique and 21 overlapped between the two datasets. db1 *SCNA* in locus 6 and in db2 *DGKQ* ( $P = 6.78 \times 10^{-6}$ ) in locus 4 (4p16) had the highest gene-wide significance. *AXINI* was significant in db1 but not in db2. The reason is that in OASIS analysis locus 22 containing *AXINI* was found in quadrant A in db1, whereas it was found in quadrant C in db2. As indicated earlier, OASIS analysis is completely different from GATES analysis incorporated in KGG as well as other GSEA algorithms. OASIS has been shown to identify known disease susceptibility genes in quadrant C while they were missed by standard analysis (Saeed 2017).

Single-variant replication of SNPs with the highest  $-\log [P]$  values (Max\_SNP) in OASIS identified loci was carried out in db3, db4, and db5. In db4, 24 SNPs in nine OASIS loci replicated (Table S5), whereas in db3 and db5, SNPs in only



**Fig. 1** a (db1), b (db2) OASIS genome-wide association in the two PD datasets. The  $-\log [P]$  values for variants across the genome according to their categorization in quadrants (A, B, or C) in the two PD datasets (db1,

db2) are shown. This is based on the intersection of  $-\log [P]$  and OASIS score for each variant

three loci (4, 6, and 11) could be validated (Table S5). PDGene (db4) is a much larger database and all 60 OASIS identified Max\_SNPs were found, whereas only 15 SNPs in db3 and 9 SNPs in db5 were found. Significantly replicated SNPs mapped on to known PD loci including *SCNA* on 4q22, *DGKQ/TMEM175* on 4p16, and *MAPT* locus on 17q21. Max\_SNPs in locus 22 (16p13) containing *AXINI* were not found to be associated with PD in the replication datasets indicating that this and 14 other OASIS loci could potentially harbor novel PD genes.

Using GEO dataset GDS2821, the KGG identified genes ( $n = 88$ ) were further investigated for gene expression changes in SN of PD patients and controls. Of these, 21 genes were found to be nominally significant and 4 genes crossed the false discovery rate (FDR) (Table 1). Two-thirds of these genes were found in three known PD loci (4p16, 4q22, and 17q21) and one novel locus on 16p13. The most significant change in gene expression in PD SN was for *AXINI* located in 16p13. In normal brains,

*AXINI* was repressed in SN compared to WM ( $4.55 \times 10^{-19}$ ); however, in PD brains, it was upregulated in SN compared to SN in normal brains ( $2.97 \times 10^{-5}$ ;  $FDR = 2.35 \times 10^{-3}$ ) (Table 1). *CLDN1* was 2.3-fold upregulated in SN of normal brains compared to WM, but significantly repressed in PD SN (Table 1). *ZNF141* and *ZNF721* normally had lower expression in SN but were upregulated in PD (Table 1).

Protein interactions of genes in Table 1 were tested with familial PD genes (Fig. 2). There was experimentally determined interaction between *LRKK2* and *AXINI*, *ZNF141* and *ZNF721*. *RHOT2*, also located in 16p13 and repressed in PD SN (Table 1), significantly interacted with *PARK2* and *PINK1*. *RICTOR* interacted with *PINK1* and was co-expressed (0.221) with *LRKK2*. *RICTOR*, located on 5p13, is normally repressed in SN but overexpressed in PD (Table 1).

To identify the variant(s) modulating *AXINI* expression, eQTL analysis was performed on all significant db1 and db2 SNPs (Table S6). The most significant eQTL SNPs reported

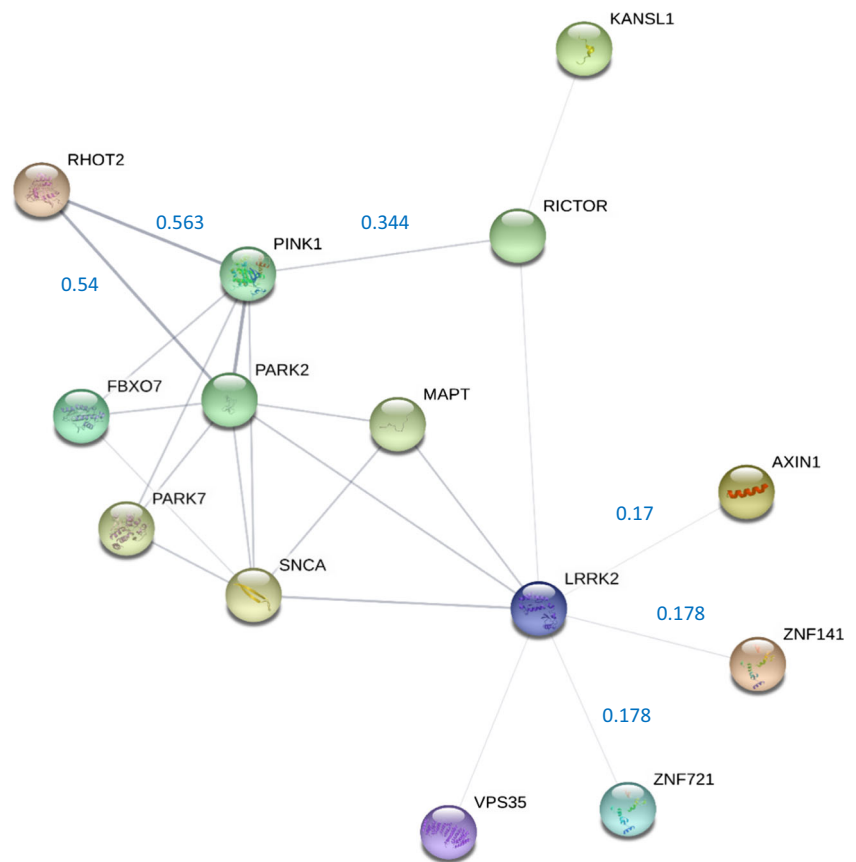
**Table 1** Genomic convergence of gene-wide association and gene expression analysis of genes in OASIS loci

Chr	Dataset	OASIS locus	Gene	$P_{KGG}$	Probe <sub>GEO</sub>	PD/N SN	$T(P)_{GEO}$	FDR (P) <sub>GEO</sub>	Probe <sub>BRC</sub>	$T(P)_{BRC}$	Fold <sub>BRC</sub>
3q28	db2	3	CLDN1	1.21E-02	222549_at	0.55	7.04E-04	1.93E-03	t2710599	9.91E-32	2.27
3q28	db2	3	P3H2	1.20E-02	218717_s_at	1.68	1.90E-02	N.S	NA	NA	NA
4p16.3	db1	4	ZNF141	1.28E-02	235580_at	1.75	9.56E-05	2.40E-03	t2713950	2.32E-04	0.86
4p16.3	db1	4	ZNF721	3.48E-02	228029_at	1.51	1.10E-03	1.68E-03	t2756404	2.79E-03	0.90
4p16.3	db1	4	ABCA11P	1.30E-02	220159_at	0.67	1.20E-02	N.S	t2756309	1.65E-06	1.17
4p16.3	db2	4	LOC100130872	8.10E-03	1569496_s_at	1.41	5.00E-02	N.S	t2756965	1.35E-02	0.96
4p16.1	db2	5	JAKMIP1	4.39E-02	238600_at	0.62	3.81E-02	N.S	t2759158	5.47E-41	1.99
4q22.1	db1	6	SNCA	1.02E-07	211546_x_at	0.64	2.98E-02	N.S	t2777714	2.82E-07	1.28
4q22.1	db2	6	MMRN1	4.90E-04	205612_at	1.89	3.90E-02	N.S	t2735759	5.15E-02	1.05
4q28.1	db1	8	FAT4	2.91E-02	219427_at	1.26	4.68E-02	N.S	t2742581	1.07E-07	0.82
5p13.1	db1	10	RICTOR	3.47E-02	226310_at	1.36	5.10E-03	N.S	t2854241	2.44E-14	0.69
8p21.2	db1	16	CDCA2	3.05E-02	236957_at	1.60	4.10E-02	N.S	t3090697	4.01E-01	1.00
9p21.2	db1	18	LINC00032	3.96E-02	1559292_s_at	1.80	6.10E-03	N.S	NA	NA	NA
14q22.3	db1	21	WDHD1	1.36E-03	204727_at	1.71	5.00E-02	N.S	t3565571	2.00E-07	0.91
16p13.3	db1	22	AXIN1	1.00E-05	212849_at	5.78	2.97E-05	2.35E-03	t3675047	4.55E-19	0.86
16p13.3	db2	22	MCRIP2	1.60E-02	225861_at	1.16	2.80E-02	N.S	NA	NA	NA
16p13.3	db2	22	RHOT2	2.71E-02	65770_at	0.81	4.20E-02	N.S	t3643229	1.77E-04	0.96
16p13.3	db2	22	RHBDL1	4.94E-02	207383_s_at	0.58	4.30E-02	N.S	t3643281	1.79E-25	1.22
16p13.3	db2	22	FBXL16	1.56E-02	227641_at	0.69	4.70E-02	N.S	t3675308	2.07E-24	1.54
17q21.31	db1	23	KANSL1	1.44E-06	225117_at	1.21	3.50E-02	N.S	NA	NA	NA
17q21.31	db1	23	ARHGAP27	1.71E-05	225618_at	1.34	4.92E-02	N.S	t3759778	1.19E-01	1.02

The genes that showed significantly altered expression in SN of PD cases compared to controls in the GEO dataset GDS2821 ( $T(P)_{GEO}$ ) are shown. These genes were also evaluated in BRAINEAC datasets for expression difference between SN and WM in normal brains. The gene-wide significance based on GATES analysis using KGG is also shown. Fold<sub>BRC</sub> was calculated as described in the “Methods” section

Chr chromosome,  $P_{KGG}$  gene-wide significance based on KGG analysis, Probe<sub>GEO</sub> the probe used to measure the gene expression in the GEO dataset GDS2821, PD/N SN ratio of expression in GEO GDS2821 for the gene in PD SN and control (N),  $T(P)_{GEO}$  P value based on one-tailed Student's *t* test of unequal variance between PD SN and controls for the respective gene in the GEO dataset GDS2821, FDR<sub>GEO</sub> false discovery rate correction for  $T(P)_{GEO}$ , Probe<sub>BRC</sub>  $T(P)_{BRC}$  represent similar abbreviations but for the BRAINEAC dataset

**Fig. 2** STRING protein–protein interactions of PD candidate genes. The interactions of proteins, whose genes had significantly altered expression in PD SN, with familial PD proteins are shown. The gene names are shown and thickness of the connecting lines represents the strength of interaction. The numbers represent STRING calculated strength of experimentally determined interaction between a protein pair



by BRAINEAC were also included. Variants in *AXINI* predominantly modulated its expression in the basal ganglia (Table S6). The Max\_SNP at locus 22, rs11648673 ( $P = 4.77 \times 10^{-7}$ ) was not an eQTL for *AXINI*. It is in strong LD with rs11644916 ( $r^2 = 0.892$ ) which also showed minimally significant eQTL association ( $P = 0.043$ ). SNPs 17–22 (Table S6) are in strong LD with each other ( $r^2 > 0.98$ ) and though they were reported as the top eQTL in BRAINEAC datasets for *AXINI*, the lack of association of three of these SNPs in db1 and db2 make them unlikely candidates in PD (Table S6). The intronic SNP rs13337493 (eQTL  $P = 1.4 \times 10^{-4}$ ) significantly modulated *AXINI* expression in SN; however, it was not genotyped in either db1 or db2. This SNP is in LD with rs758033 ( $r^2 = 0.962$ ), rs2361988 ( $r^2 = 0.962$ ) which showed significant associations in db1 (Table S6).

## Discussion

In this study, four major loci for PD were identified, of which three were previously known (4q22, 17q21, and 4p16). The novel PD locus 16p13 harbors the gene *AXINI* which is a critical component of Wnt/ $\beta$ -catenin signaling and functions as its negative regulator, inducing apoptosis (Clevers and Nusse 2012). OASIS showed significant association of the 16p13 locus with both sporadic and familial PD (db1 and

db2). Moreover, the strength of association was nearly as much as the *MAPT* locus on 17q21 and higher than the *TMEM175* locus, 4p16. *AXINI* was the most significant gene at the 16p13 locus on GATES analysis. Gene expression in SN of PD patients was found to be most significantly altered for *AXINI* from amongst 88 PD associated genes across the genome. *AXINI* was also shown to have a modest protein interaction with *LRRK2*. These findings demonstrate a significant genetic association of *AXINI* with PD and implicate it in its pathobiology.

Wnt signaling is mediated by  $\beta$ -catenin, a nuclear factor that promotes cell cycling (Clevers and Nusse 2012). Concentration of  $\beta$ -catenin is controlled by a degradation complex whose rate-limiting component is Axin (coded by *AXINI*) (Clevers and Nusse 2012). Wnt/ $\beta$ -catenin signaling has been shown to play a crucial role in the neurogenesis of the of midbrain dopaminergic neurons (L'Episcopo et al. 2014). Moreover,  $\beta$ -catenin accumulates specifically in mature thalamic neurons and is an intrinsic feature of these cells (Misztal et al. 2011). Normally, Axin levels and that of other components of the  $\beta$ -catenin degradation complex are low in the thalamic neurons which allows the nuclear localization of  $\beta$ -catenin, independent of Wnt signaling (Misztal et al. 2011). Axin is inhibited by tankyrases whose inhibitors, by stabilizing Axin, promote apoptosis (Huang et al. 2009).

Similarly, overexpression of Axin induces  $\beta$ -catenin degradation (Nakamura et al. 1998). It can therefore be postulated that upregulation of *AXINI* expression in SN of PD patients, as shown here, may be involved in neurodegeneration typical of PD.

*SCNA* was the obvious gene at 4q22 (Table S1); however, the association signal was more widespread at the 4p16 and 17q21 loci. Consequently, the specific gene at the 4p16 locus is as yet unclear. At the 17q21 locus, *MAPT* is a known PD gene; however, several nearby genes have consistently demonstrated association and it is possible that this locus may harbor additional PD susceptibility genes (Lill et al. 2012; Nagle et al. 2016). Here, two genes at each of these loci showed altered gene expression. At 4p16, *ZNF141* and *ZNF721*, though modestly associated with PD in GATES analysis, showed significant upregulation of expression in SN of PD patients (Table 1). It is possible that rare variants in one or both of these genes are linked to common variants at this locus generating the association signal in GWAS. Deep sequencing of *ZNF141* and *ZNF721* together with other candidate genes at 4p16 (Table S1), will provide the definite answer to this intriguing association signal. Interestingly, zinc finger protein genes, which function as transcriptional regulators, have been previously associated with PD in GWAS (Chang et al. 2017). At 17q21, *KANSL1* and *ARHGAP27* showed significant gene-wide association (Tables S3 and S4) as well as were upregulated in PD SN (Table 1). *ARHGAP27* is involved in the clathrin-mediated endocytosis of receptor tyrosine kinases (Sakakibara et al. 2004). *KANSL1* is involved in cell cycling by microtubule stabilization (Meunier et al. 2015) and its haploinsufficiency causes intellectual disability (Zollino et al. 2012). Of the six consistently associated genes at this locus (Chang et al. 2017), *KANSL1* and *ARHGAP27* appear to be strong candidates for further investigation including sequencing and biological studies.

Other notable loci that this study identified include 8p23 (locus 15) and 14q22 (locus 21), both of which were found in quadrant A in at least one dataset (Table S1). Only two genes at 8p23 were significant on GATES analysis (Table S1). *ARHGEF10* is involved in cell cycling and a mutation has been shown to result in slowed nerve conduction velocity (Verhoeven et al. 2003; Aoki et al. 2009). The other candidate gene at 8p23 was *CSMD1* (gene-wide  $P = 0.003$ ). *CSMD1* is involved in complement activation and inflammation in the developing CNS (Kraus et al. 2006). Using whole exome sequencing, *CSMD1* was recently shown to harbor mutations for familial PD (Ruiz-Martínez et al. 2017). This major finding helps validate the methodology described in the present study of composite analysis with OASIS and GATES to map disease susceptibility genes in GWAS.

On 14q22, an interesting candidate gene is *WDHD1* (Table 1), a nucleoplasmic DNA-binding protein that influences cell cycling as it is a key regulator of centromere function (Hsieh et

al. 2011). *CLDN1* at 3q28 (Table 1) is involved in formation of tight junctions during neuroinflammation (Hornig et al. 2017). *CLDN1* repression in PD (Table 1) is suggestive of an increased propensity for inflammation which may help the spread of PD pathology (Sampson et al. 2016). *RHOT2* is a critical component of the mitochondrial trafficking complex together with Pink1 (Fig. 2) (Wang et al. 2011). Overexpression of *RHOT2* increases Pink1 mitochondrial localization preventing mitochondrial dysfunction (Weihsen et al. 2009). Here, *RHOT2* expression was decreased in PD SN (Table 1). Though the gene-wide and expression associations are modest, *RHOT2* should be considered an important PD candidate gene. Pink1 is not only cytoprotective in the mitochondria but also in the cytosol where it induces Akt phosphorylation that is mediated by Rictor (encoded by *RICTOR*) (Murata et al. 2011). Therefore, the locus 5p13 containing the gene *RICTOR* (Table S1) is another interesting candidate locus for PD identified in this study. Further studies are needed to experimentally determine the causal genes in the PD loci described here (Table S1). This study, however, provides a paradigm for the design of such functional studies. *MAPT* expression remained unaltered between PD SN and controls (data not shown), indicating that expression analysis may not suffice as the only type of functional study for genes identified by composite analysis described here.

In summary, composite analysis with OASIS, a locus-based clustering algorithm and gene-based association testing with GATES, on existing GWAS datasets combined with expression studies has led to the identification of several novel PD candidate genes and loci. This gene-mapping strategy has immense potential to mine genes of low to moderate effect sizes in GWAS for complex diseases such as PD (Saeed 2017; Saeed 2018). The genes thus identified can be subsequently validated by deep sequencing and functional studies including expression in affected tissues. Convergence of genomic and functional data then provides robust evidence for disease susceptibility. The most noteworthy novel PD associated gene identified here is *AXINI* located at 16p13. Based on the eQTL data the SNPs rs13337493, rs758033, and rs2361988 seem to modulate *AXINI* expression in the basal ganglia. This finding needs further investigation using comprehensive *AXINI* genotyping and eQTL studies. It is also possible that a haplotype modulates *AXINI* expression in SN rather than a single variant, as in the case of *APOL1* (Larsen et al., 2013). Other important PD candidate genes that this study also identified include *CLDN1* at 3q28, *ZNF141* and *ZNF721* at 4p16, *RICTOR* at 5p13, *CSMD1* at 8q23 recently shown to be a familial PD gene, *RHOT2* also located at 16p13, and *KANSL1* and *ARHGAP27* located at 17q21. Biological studies including proteomic, immunohistopathologic, in vitro, and in vivo studies would help confirm the pathogenic relevance of these genes and may provide therapeutic targets for PD modulation.

**Acknowledgements** All GWAS data was obtained from dbGAP.

**Web links** *OASIS* [<https://github.com/dr-saeed/OASIS/blob/master/OASIS.py>], *SNIPPER* [<https://csg.sph.umich.edu/boehnke/snipper/>], *KGG* [<http://grass.cgs.hku.hk/limx/kgg/>], *BRAINEAC* [<http://www.braineac.org>] and *STRING* [<https://string-db.org/cgi/input.pl>], *LDlink* [<https://analysistools.nci.nih.gov/LDlink/?tab=ldmatrix>]

## Compliance with ethical standards

**Conflict of Interest** The author declares that he has no conflicts of interest.

## References

- Aoki T, Ueda S, Kataoka T, Satoh T (2009) Regulation of mitotic spindle formation by the RhoA guanine nucleotide exchange factor ARHGGEF10. *BMC Cell Biol* 10:56
- Chang D, Nalls MA, Hallgrímsson IB, Hunkapiller J, van der Brug M, Cai F, International Parkinson's Disease Genomics Consortium; 23andMe Research Team et al (2017) A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* 49(10):1511–1516
- Clevers H, Nusse R (2012) (2012). Wnt/ $\beta$ -catenin signaling and disease. *Cell* 149(6):1192–1205
- Deng HX, Shi Y, Yang Y, Ahmeti KB, Miller N, Huang C, Cheng L, Zhai H, Deng S, Nuytemans K, Corbett NJ, Kim MJ, Deng H, Tang B, Yang Z, Xu Y, Chan P, Huang B, Gao XP, Song Z, Liu Z, Fecto F, Siddique N, Foroud T, Jankovic J, Ghetti B, Nicholson DA, Krainc D, Melen O, Vance JM, Pericak-Vance MA, Ma YC, Rajput AH, Siddique T (2016). Identification of TMEM230 mutations in familial Parkinson's disease. *Nat Genet* 48(7):733–739.PMID: 27270108
- Hernandez DG, Reed X, Singleton AB. (2016). Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. *J Neurochem* 139 Suppl 1:59–74. PMID: 27090875
- Hornig S, Therattil A, Moyon S, Gordon A, Kim K, Argaw AT, Hara Y, Mariani JN, Sawai S, Flodby P, Crandall ED, Borok Z, Sofroniew MV, Chapouly C, John GR (2017) Astrocytic tight junctions control inflammatory CNS lesion pathogenesis. *J Clin Invest* 127(8):3136–3151
- Hsieh CL, Lin CL, Liu H, Chang YJ, Shih CJ, Zhong CZ, Lee SC, Tan BC (2011) WDHD1 modulates the post-transcriptional step of the centromeric silencing pathway. *Nucleic Acids Res* 39(10):4048–4062
- Huang SM, Mishina YM, Liu S, Cheung A, Stegmeier F, Michaud GA, Charlat O, Wiellette E, Zhang Y, Wiessner S et al (2009) Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* 461(7264):614–620
- Keller MF, Saad M, Bras J, Bettella F, Nicolaou N, Simón-Sánchez J, Mittag F, Büchel F, Sharma M, Gibbs JR, Schulte C, Moskva V, Durr A, Holmans P, Kilarski LL, Guerreiro R, Hernandez DG, Brice A, Yliskotila P, Stefánsson H, Majamaa K, Morris HR, Williams N, Gasser T, Heutink P, Wood NW, Hardy J, Martinez M, Singleton AB, Nalls MA, International Parkinson's Disease Genomics Consortium (IPDGC), Wellcome Trust Case Control Consortium 2 (WTCCC2) (2012) Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum Mol Genet* 21(22):4996–5009
- Kraus DM, Elliott GS, Chute H, Horan T, Pfenninger KH, Sanford SD, Foster S, Scully S, Welcher AA, Holers VM (2006) CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *J Immunol* 176(7):4419–4430
- L'Episcopo F, Tirolo C, Testa N, Caniglia S, Morale MC, Serapide MF, Pluchino S, Marchetti B (2014) Wnt/ $\beta$ -catenin signaling is required to rescue midbrain dopaminergic progenitors and promote neurorepair in ageing mouse model of Parkinson's disease. *Stem Cells* 32(8):2147–2163
- Larsen CP, Beggs ML, Saeed M, Walker PD (2013) Apolipoprotein L1 risk variants associate with systemic lupus erythematosus-associated collapsing glomerulopathy. *J Am Soc Nephrol* 24(5):722–725
- Li MX, Sham PC, Cherny SS, Song YQ (2010) A knowledge-based weighting framework to boost the power of genome-wide association studies. *PLoS One* 5:e14480
- Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BM, Schjeide LM, Meissner E, Zauft U, Allen NC, et al. (2012). Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene database. *PLoS Genet* 8(3):e1002548. PMID: 22438815
- Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PV, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet* 77(5):685–693
- Meunier S, Shvedunova M, Van Nguyen N, Avila L, Vemos I, Akhtar A (2015) An epigenetic regulator emerges as microtubule minus-end binding and stabilizing factor in mitosis. *Nat Commun* 6:7889
- Misztal K, Wisniewska MB, Ambrozkiwicz M, Nagalski A, Kuznicki J (2011) WNT protein-independent constitutive nuclear localization of beta-catenin protein and its low degradation rate in thalamic neurons. *J Biol Chem* 286(36):31781–31788
- Murata H, Sakaguchi M, Jin Y, Sakaguchi Y, Futami J, Yamada H, Kataoka K, Huh NH (2011) A new cytosolic pathway from a Parkinson disease-associated kinase, BRPK/PINK1: activation of AKT via mTORC2. *J Biol Chem* 286(9):7182–7189
- Nagle MW, Latourelle JC, Labadorf A, Dumitriu A, Hadzi TC, Beach TG, Myers RH (2016) The 4p16.3 Parkinson disease risk locus is associated with GAK expression and genes involved with the synaptic vesicle membrane. *PLoS One* 11(8):e0160925
- Nakamura T, Hamada F, Ishidate T, Anai K, Kawahara K, Toyoshima K, Akiyama T (1998) Axin, an inhibitor of the Wnt signalling pathway, interacts with beta-catenin, GSK-3beta and APC and reduces the beta-catenin level. *Genes Cells* 3(6):395–403
- Pankratz N, Wilk JB, Latourelle JC, AL DS, Halter C, Pugh EW, Doheny KF, Gusella JF, Nichols WC, Foroud T, Myers RH, PSG-PROGENI and GenePD Investigators, Coordinators and Molecular Genetic Laboratories (2009) Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum Genet* 124(6):593–605
- Ruiz-Martínez J, Azcona LJ, Bergareche A, Martí-Massó JF, Paisán-Ruiz C (2017) Whole-exome sequencing associates novel CSMD1 gene mutations with familial Parkinson disease. *Neurol Genet* 3(5):e177
- Saeed M (2017) Novel linkage disequilibrium clustering algorithm identifies new lupus genes on meta-analysis of GWAS datasets. *Immunogenetics* 69:295–302
- Saeed M (2018) Locus and gene-based GWAS meta-analysis identifies new diabetic nephropathy genes. *Immunogenetics* 70(6):347–353
- Sakakibara T, Nemoto Y, Nukiwa T, Takeshima H (2004) Identification and characterization of a novel rho GTPase activating protein implicated in receptor-mediated endocytosis. *FEBS Lett* 566(1–3):294–300
- Sampson TR, Debelius JW, Thron T, Janssen S, Shastri GG, Ilhan ZE, Challis C, Schretter CE, Rocha S, Gradinaru V, Chesselet MF, Keshavarzian A, Shannon KM, Krajmalnik-Brown R, Wittung-Stafshede P, Knight R, Mazmanian SK (2016) Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease. *Cell* 167(6):1469–1480.e12
- Simón-Sánchez J, Schulte C, Bras JM, Shama M, Gibbs JR, Berg D, Paisan-Ruiz C, Lichtner P, Scholz SW, Hernandez DG, Krüger R, Federoff M, Klein C, Goate A, Perlmutter J, Bonin M, Nalls MA,

- Illig T, Gieger C, Houlden H, Steffens M, Okun MS, Racette BA, Cookson MR, Foote KD, Fernandez HH, Traynor BJ, Schreiber S, Arepalli S, Zonozi R, Gwinn K, van der Brug M, Lopez G, Chanock SJ, Schatzkin A, Park Y, Hollenbeck A, Gao J, Huang X, Wood NW, Lorenz D, Deuschl G, Chen H, Riess O, Hardy JA, Singleton AB, Gasser T (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet* 41(12):1308–1312
- Verhoeven K, De Jonghe P, Van de Putte T, Nelis E, Zwijsen A, Verpoorten N, De Vriendt E, Jacobs A, Van Gerwen V et al (2003) Slowed conduction and thin myelination of peripheral nerves associated with mutant rho guanine-nucleotide exchange factor 10. *Am J Hum Genet* 73(4):926–932
- Wang X, Winter D, Ashrafi G, Schlehe J, Wong YL, Selkoe D, Rice S, Steen J, LaVoie MJ, Schwarz TL (2011) PINK1 and Parkin target Miro for phosphorylation and degradation to arrest mitochondrial motility. *Cell* 147(4):893–906
- Weihofen A, Thomas KJ, Ostaszewski BL, Cookson MR, Selkoe DJ (2009) Pink1 forms a multiprotein complex with Miro and Milton, linking Pink1 function to mitochondrial trafficking. *Biochemistry* 48(9):2045–2052
- Zollino M, Orteschi D, Murdolo M, Lattante S, Battaglia D, Stefanini C, Mercuri E, Chiurazzi P, Neri G, Marangi G (2012) Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat Genet* 44(6):636–638