

# Reference SVA insertion polymorphisms are associated with dopaminergic degeneration in Parkinson's Disease and differential gene expression in the PPMI cohort

**Abigail L Pfaff**

Murdoch University <https://orcid.org/0000-0002-2231-9800>

**Vivien J. Bubb**

University of Liverpool <https://orcid.org/0000-0003-2763-7004>

**John P. Quinn**

University of Liverpool

**Sulev Koks** (✉ [sulev.koks@murdoch.edu.au](mailto:sulev.koks@murdoch.edu.au))

Murdoch University <https://orcid.org/0000-0001-6087-6643>

---

## Research article

**Keywords:** Parkinson's disease, SINE-VNTR-Alus, retrotransposons, whole genome sequencing, gene expression

**Posted Date:** July 21st, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-41600/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background:** The development of Parkinson's disease (PD) involves a complex interaction of genetic and environmental factors. The majority of studies investigating the genetic component of complex diseases, including PD, have focused on single nucleotide polymorphisms as this enables genome wide analysis of a large number of samples. Genome wide association studies have been crucial in identifying PD risk variants, however a large proportion of the heritability of PD remains to be identified. To investigate the component of PD that may involve complex genetic variants we characterised SINE-VNTR-Alus (SVAs), a retrotransposon known to affect gene expression, in the Parkinson's Progression Markers Initiative (PPMI) cohort.

**Results:** Utilising whole genome sequencing from the PPMI cohort that consisted of 179 healthy controls, 371 individuals with PD and 58 individuals classified as SWEDD (scans without evidence of dopaminergic deficit) we genotyped SVAs in the reference genome for their presence or absence identifying 81 such SVAs. Seven of these SVAs were associated with progression of the disease, including four whose specific genotypes were linked to an increase in the gradient of dopaminergic loss when comparing the caudate to putamen from DaTscan imaging analysis. These seven SVAs also demonstrated regulatory properties as they were associated with differential gene expression in whole blood RNA sequencing data.

**Conclusion:** This study highlights the importance of addressing variation of SVAs and potentially other types of retrotransposons in PD genetics, furthermore these SVA elements should be considered as regulatory domains that could play a role in disease progression.

## Background

Parkinson's disease (PD) is the second most common neurodegenerative disease affecting 1% of the population over 60 years of age (1). PD pathology is characterised by the loss of dopaminergic neurons from the substantia nigra and the formation of neuronal inclusions called Lewy bodies. The primary symptoms are related to motor dysfunction (bradykinesia, resting tremor, rigidity and postural instability), however individuals with PD can also develop a range of non-motor symptoms that include sleep disturbances, constipation and impaired olfaction (2). The precise mechanisms underlying the development of PD are unknown. Genetic studies have played a key role in highlighting pathways involved in disease pathogenesis that include autophagy, mitophagy and the innate immune system (3, 4). There are highly penetrant mutations in several genes that cause rare monogenic forms of PD in either an autosomal dominant (e.g. *SNCA* and *LRRK2*) or an autosomal recessive (e.g. *PRKN* and *PARK7*) manner (5). However, monogenic forms of the disease contribute a small percentage to the total incidence of the disease, with the majority of cases classed as sporadic and likely caused by a complex interaction of genetic and environmental risk factors. Large-scale genome wide association studies (GWAS) have been crucial in identifying genetic risk factors of idiopathic disease, however a large proportion of the disease heritability has yet to be identified (6, 7). In addition, many of the single nucleotide polymorphisms (SNPs) that are associated with disease development are in non-coding regions of the genome and their functional impact is often unknown. We aim to investigate the genetic component of PD further by studying other types of variation and their functional consequences and this study focuses on variation generated by a specific family of retrotransposons.

SINE-VNTR-*Alus* (SVAs) are hominid specific retrotransposons and comprise approximately 2700 elements in the reference human genome (8). SVAs are composite elements (domains outlined in Fig. 1a) consisting of seven subfamilies (A-F and F1) and are the evolutionarily youngest member of the non-long terminal repeat class of retrotransposons (9, 10). Members of the SVA family are still able to retrotranspose in the human genome (11), and this ongoing mobilisation has led to insertions that are polymorphic in the human population for their presence or absence. Twelve such polymorphic SVA insertions have been identified as the cause of diseases including X-linked dystonia parkinsonism (XDP), Neurofibromatosis type 1 and haemophilia B through mechanisms such as loss of function mutation, modulation of splicing and deletions at the site of insertion (12). XDP is associated with a founder haplotype that contains a SVA in intron 32 of the *TAF1* gene, reduced expression of *TAF1* and alternative splicing and intron retention directed by the SVA insertion (13). Polymorphic retrotransposons, including members of the SVA family, have been shown to regulate gene expression in a population specific manner and could have an impact on phenotypic differences (14). In the literature retrotransposon insertion polymorphisms (RIPs) have also been identified to be in linkage disequilibrium (LD) with SNPs associated with a range of diseases through GWAS, further analysis of selected candidates identified six RIPs that could lead to disease through changes in gene regulation (15).

Our study focused on the potential role of reference SVA RIPs established in the population in the risk and progression of PD. To address this we utilised whole genome sequence, transcriptomic and clinical data from the Parkinson's Progression Markers Initiative (PPMI) which was established to identify PD progression markers to help understand disease aetiology and ultimately aid in the development of novel therapeutics (16). The PPMI is a longitudinal study collecting clinical and imaging data and biospecimens from PD subjects, healthy controls and those subjects classified as SWEDD (scans without evidence of dopaminergic deficit) (17). Using the extensive dataset from the PPMI cohort, we identified 81 SVAs that were polymorphic for their presence or absence in the reference genome. Seven of the identified SVA RIPs

were associated with PD progression and subsequent analysis of these seven RIPs correlated specific genotypes with changes in gene expression in the blood.

## Methods

### Identifying SVAs polymorphic for their presence/absence in the reference genome from whole genome sequencing data from the PPMI cohort

Whole genome sequencing data in bam file format aligned to hg38 were obtained from the PPMI. The structural variant caller Delly2 (<https://github.com/dellytools/delly>) (18), with default settings, was used to genotype structural variants in 608 individuals (179 healthy controls, 371 PD and 58 SWEDD) from the PPMI cohort. Those structural variants located within the coordinates of the human specific SVAs, available on the UCSC genome browser (<https://genome.ucsc.edu/>) (19), +/-100 bp were extracted. Structural variants classified as deletions in these regions were inspected manually to determine if the break points were consistent with the reported coordinates of the reference human specific SVAs. Association analysis of the 81 SVA RIPs identified with PD was performed using logistic regression in PLINK (v1.07) (20), using data from the control and PD samples only, and p values adjusted for multiple testing (Bonferroni correction).

VCF files for the PPMI samples were downloaded from the PPMI website ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)). SNPs with a minor allele frequency greater than 0.001 located in 250 kb either side of the 81 SVA RIPs were extracted and PLINK (v1.07) was used to identify those SNPs in strong LD with the RIPs ( $r^2 > 0.8$ ). For each SVA RIP the SNP in strongest LD was selected and searched for in the summary statistics from Nalls et al (6) to determine if this proxy SNP for the SVA RIP genotype was associated with PD risk.

### Association analysis of SVA RIPs with clinical features and progression markers of PD in the PPMI cohort

For longitudinal analysis, a linear mixed-effects model combined with FDR correction measured the changes in phenotype scores during follow-up visits related to the 81 SVA RIPs in the PPMI cohort. A pairwise analysis was then performed on the SVAs that were significant after FDR correction for an effect size comparison with Tukey adjusted p value reported. Longitudinal data was obtained from PPMI ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)) and altogether, 111 different clinical variables were analysed in the PD subjects from four visits at baseline (0), 12, 24 and 36 months for the majority of the data points (data relating to the DaTscan SPECT imaging was not available at 36 months).

### Analysis of SVA RIP genotype and gene expression in blood RNA-sequencing data from the PPMI cohort

In order to evaluate the effect of SVA RIP genotypes on the gene expression profile, differential whole transcriptome analysis based on whole blood RNA sequencing data (obtained from [www.ppmi-info.org/data](http://www.ppmi-info.org/data)) was performed on all subjects (healthy controls, PD and SWEDD). The DESeq2 package in R was used to detect statistically significant differences in the gene expression profiles between the different genotypes of the seven SVAs associated with clinical features of PD at baseline (0 months) and 36 months.

## Results

### Characterisation of reference SVA RIPs in the PPMI cohort

The PPMI cohort whose whole genome sequence data were analysed in this study consisted of 179 healthy controls, 371 PD and 58 subjects classified as SWEDD. The subjects from each group showed similar demographics in terms of age and sex (Table 1) and all subjects included reported race as white. In the 608 subjects analysed, 83 SVAs in the human reference genome were identified as absent in at least one genome compared to the reference (hg38) (Additional Table 1). Two of these were detected as absent in all genomes analysed and therefore were not classified as polymorphic in this cohort. Of the 81 SVAs RIPs, 28 (34.6%) had not been reported previously in the dbRIP track available on UCSC genome browser. The majority, 85.2%, of the SVA RIPs belonged to subtypes E, F and F1 (Fig. 1b), which was consistent with these human specific subfamilies being the most recently active (9). The SVA RIPs showed a wide range of insertion allele frequencies from 0.0008–0.994 (Fig. 1c). Using logistic regression and Bonferroni correction for multiple testing there were no SVA RIPs significantly associated with PD when comparing control and PD subjects in this cohort (Additional Table 1). Tagging SNPs ( $r^2 > 0.8$ ) were identified for 76 (93.8%) of the SVA RIPs and were used to act as a proxy for the SVA genotype, enabling the evaluation of these elements in the largest PD and healthy control dataset available from the recent GWA meta-analysis performed by Nalls et al (6). In the summary statistics from this meta-analysis, 75 of the SVA RIP

tagging SNPs were present (Additional Table 2). One of these SNPs (rs55653937) was reported as associated with PD at genome wide significance ( $p = 1.64 \times 10^{-19}$ ) and was in strong LD ( $r^2 = 0.994$ ) with SVA\_67 located 12 kb upstream of the *KANSL1* gene.

Table 1  
Demographics of the PPMI subjects.

Variable	PPMI subjects included in genome wide genotyping of reference SVAs		
	PD (n = 371)	Healthy controls (n = 179)	SWEDD (n = 58)
Sex			
Male	241 (65%)	117 (65%)	38 (66%)
Female	130 (35%)	62 (35%)	20 (34%)
Age			
Mean (min, max)	62.0 (33.5–84.9)	61.3 (30.6–82.7)	60.7 (38.3–78.8)
Family history			
1st degree family member with PD	51 (14%)	0 (0%)	13 (24%)
Other family member with PD	43 (12%)	10 (6%)	5 (9%)
No family member with PD	276 (74%)	169 (94%)	39 (67%)

## Reference SVA RIPs are associated with progression markers of PD in the PPMI cohort

The PPMI cohort has extensive clinical and phenotypic data available at four time points, baseline (0 months) and subsequent measurements at 12, 24 and 36 months. This data includes motor, cognitive and olfactory testing, dopamine transporter (DAT) imaging assessments and analysis of markers in biospecimens. Analysis of the 81 SVA RIPs identified seven associated with features of PD (Table 2) at 12 months or later. The nearest gene to each of these SVAs are reported in Table 3 and for chromosomal loci see Additional Table 1. Four of the SVA RIPs (36, 47, 67 and 79) in Table 3 were associated with changes in the count density ratios (CDR) of the caudate/putamen, which had been calculated from DaTscan single photon emission computed tomography (SPECT) imaging. Figure 2a highlights this for SVA\_67, located upstream of the *KANSL1* gene, showing PD subjects with AA genotype of SVA\_67 had a significantly higher mean caudate/putamen CDR (3.81, 95% CI 3.12–4.51) than those with both PA (2.63, 95% CI 2.43–2.83,  $p = 0.004$ ) and PP (2.64, 95% CI 2.49–2.79,  $p = 0.004$ ) genotypes at 12 months. The AA PD subjects also had a significantly higher mean (5.10, 95% CI 4.43–5.76) at 24 months compared to PA (2.67, 95% CI 2.46–2.88,  $p < 0.001$ ) and PP (2.66, 95% CI 2.51–2.81,  $p < 0.001$ ). The remaining three SVA RIPs (7, 11 and 24) in Table 3 were associated with the following measurements reflecting the progression of PD: Hoehn and Yahr staging and the Movement Disorder Society – Unified Parkinson's Disease Rating Scale (MDS-UPDRS). For example, the presence of SVA\_11 was significantly associated with a lower mean MDS-UPDRS part II score at 36 months (Fig. 2b) when comparing PP (7.6, 95% CI 6.3–8.9) and PA (8.9, 95% CI 8.2–9.6) genotypes to AA (10.9, 95% CI 9.6–12.2,  $p = 0.001$  and  $p = 0.02$ ). SVA\_24 was significantly associated with changes in the categorical Hoehn and Yahr stage in PD subjects (Fig. 2c). Those with a PA genotype had a higher mean Hoehn and Yahr stage at 36 months (3.07, 95% CI 2.96–3.18) compared to AA (2.87, 95% CI 2.78–2.96,  $p = 0.02$ ) and PP genotypes (2.78, 95% CI 2.60–2.97,  $p = 0.03$ ).

Table 2

SVAs RIPs that are associated with clinical features and markers of progression of PD in the PPMI cohort: The mean and 95% confidence intervals (CI) are reported for each genotype at each visit with data available and adjusted p values are shown for each pair of genotypes.

SVA ID	Associated feature	Time of visit (months)	AA	PA	PP	AA vs PA	AA vs PP	PA vs PP
			mean (95% CI)	mean (95% CI)	mean (95% CI)	p-value	p-value	p-value
SVA_7	MDS-UPDRS Part III Score (OFF)	0	24.5 (9.5–39.5)	20.9 (18.5–23.3)	21.0 (19.8–22.3)	0.88	0.89	0.99
		12	28.5 (13.5–43.5)	23.9 (21.4–26.4)	25.6 (24.2–26.9)	0.82	0.92	0.49
		24	27.0 (12.1–42.0)	26.0 (23.4–28.7)	27.5 (26.1–29.0)	0.99	0.99	0.57
		36	10.0 (-4.96–25.0)	28.4 (25.7–31.1)	29.5 (28.1–31.0)	<b>0.05</b>	<b>0.03</b>	0.75
SVA_11	MDS-UPDRS Part II Score	0	6.3 (5.1–7.52)	5.6 (4.9–6.3)	6.2 (5.0–7.5)	0.57	>0.99	0.62
		12	7.4 (6.1–8.6)	7.2 (6.5–7.8)	7.9 (6.7–9.2)	0.95	0.82	0.55
		24	8.3 (7.1–9.6)	7.9 (7.2–8.5)	7.5 (6.2–8.8)	0.79	0.61	0.85
		36	10.9 (9.6–12.2)	8.9 (8.2–9.6)	7.6 (6.3–8.9)	<b>0.02</b>	<b>0.001</b>	0.16
SVA_24	Categorical Hoehn & Yahr	0	2.62 (2.55–2.69)	2.53 (2.44–2.62)	2.47 (2.31–2.64)	0.30	0.26	0.82
		12	2.76 (2.68–2.84)	2.77 (2.67–2.87)	2.62 (2.44–2.81)	0.99	0.38	0.37
		24	2.79 (2.71–2.88)	2.86 (2.75–2.96)	2.61 (2.43–2.79)	0.62	0.18	0.06
		36	2.87 (2.78–2.96)	3.07 (2.96–3.18)	2.78 (2.60–2.97)	<b>0.02</b>	0.70	<b>0.03</b>
SVA_36	Ipsilateral count density ratio (caudate/putamen)	0	2.39 (1.98–2.79)	2.40 (2.22–2.58)	2.43 (2.27–2.59)	>0.99	0.98	0.96
		12	3.36 (2.93–3.79)	2.55 (2.36–2.74)	2.64 (2.48–2.81)	<b>0.003</b>	<b>0.007</b>	0.77
		24	3.72 (3.28–4.17)	2.63 (2.44–2.82)	2.68 (2.51–2.86)	<b>&lt; 0.0001</b>	<b>0.0001</b>	0.91
SVA_47	Ipsilateral count density ratio (caudate/putamen)	0	2.40 (2.29–2.52)	2.69 (2.15–3.23)	2.53 (1.27–3.78)	0.57	0.98	0.97
		12	2.64 (2.51–2.76)	3.39 (2.85–3.93)	2.89 (1.40–4.38)	<b>0.02</b>	0.94	0.81
		24	2.66 (2.54–2.79)	4.44 (3.87–5.01)	3.18 (1.18–5.19)	<b>&lt; 0.0001</b>	0.87	0.46

SVA ID	Associated feature	Time of visit (months)	AA	PA	PP	AA vs PA	AA vs PP	PA vs PP
			mean (95% CI)	mean (95% CI)	mean (95% CI)	p-value	p-value	p-value
SVA_67	Ipsilateral count density ratio (caudate/putamen)	0	2.64 (2.04–3.24)	2.43 (2.24–2.62)	2.38 (2.24–2.53)	0.79	0.69	0.91
		12	3.81 (3.12–4.51)	2.63 (2.43–2.83)	2.64 (2.49–2.79)	<b>0.004</b>	<b>0.004</b>	> 0.99
		24	5.10 (4.43–5.76)	2.67 (2.46–2.88)	2.66 (2.51–2.81)	<b>&lt; 0.0001</b>	<b>&lt; 0.0001</b>	> 0.99
SVA_79	Contralateral count density ratio (caudate/putamen)	0	-	3.26 (2.75–3.78)	2.87 (2.76–2.98)	-	-	0.14
		12	-	2.40 (1.87–2.93)	2.93 (2.82–3.04)	-	-	0.06
		24	-	2.18 (1.67–2.69)	2.93 (2.81–3.04)	-	-	<b>0.005</b>

## The genotype of specific SVA RIPs is associated with differential gene expression

Transcriptomic data from whole blood was utilised to determine if the seven SVA RIPs associated with PD progression (Table 3) had any influence on gene expression. RNA sequencing data from whole blood was available at baseline (0 months) and four additional time points (6, 12, 24 and 36 months). In the analysis, all subjects in the cohort were used to compare gene expression profiles between subjects with the three different genotypes (PP, PA and AA) based on the presence or absence of the specific SVA RIP with statistical analysis performed at 0 and 36 months. For all of the seven SVA RIPs analysed, at least one genotype at one time point was associated with differential gene expression (Table 3). The number of genes whose differential expression was significantly associated with specific SVA genotypes (FDR < 0.1) at 0 and 36 months is summarised in Table 3 and the gene with the lowest adjusted p value is shown in brackets. Genotypes of SVA\_7 and SVA\_36 were associated with changes in gene expression of hundreds or thousands of genes at many different loci. Whereas the effects of other SVA RIPs (24 and 67) were restricted to a smaller number of genes (Table 3), many of which were located at the SVA locus, and several of those genes were affected at both time points analysed. However, one SVA RIP demonstrated a narrower effect on gene expression with only the expression of one gene (*RAB3B*) associated with SVA\_79 genotype. The effect of the SVA RIPs on gene expression varied between the different elements and below we highlight specific examples of genes or clusters whose expression was associated with multiple SVA genotypes and/or at both time points at which statistical analysis was performed.

The expression of *FLACC1* was significantly different for individuals with PP genotype compared to AA for SVA\_11 (located in an intron of *CASP8*) at 0 (FDR p = 0.04) and at 36 months (FDR p =  $2.22 \times 10^{-5}$ ) (Fig. 3a). *FLACC1* is adjacent to *CASP8* in the genome and SVA\_11 is 3.5 kb downstream of its 3'UTR. The expression of *HLA-A* was repeatedly the most significant gene to be associated with SVA\_24 (located 8 kb upstream of *HLA-A*) genotypes at 0 and 36 months (Table 3). At both 0 and 36 months the subjects with PP genotypes had significantly higher *HLA-A* expression than those with AA genotypes (FDR p =  $6.22 \times 10^{-51}$  and p =  $2.49 \times 10^{-38}$ ) (Fig. 3b). Those with PA genotypes also showed significantly higher *HLA-A* expression than AA genotypes at 0 and 36 months (FDR p =  $1.90 \times 10^{-78}$  and p =  $7.51 \times 10^{-60}$ ), however there was no significant difference when comparing PP and PA genotypes (Fig. 3b).

SVA\_67 at the chromosomal locus 17q21-31 was associated with differential expression of multiple genes, including six in a 1.15 Mb region centred around the SVA RIP (Fig. 4a). At baseline when comparing PP and AA genotypes three (*PLEKHM1*, *ARL17A* and *CRHR1*) out of the four significantly associated genes were located in this region as were five (*PLEKHM1*, *ARL17A*, *CRHR1*, *MAPT* and *LRCC37A*) out of seven genes whose expression was significantly different when comparing PP to PA genotypes (Fig. 4b-f). Extending the analysis to expression data at 36 months there were 22 genes whose expression was significantly different when comparing PP and AA genotypes of SVA\_67, 4 of which were located in the 1.15 Mb region (*PLEKHM1*, *ARL17A*, *CRHR1* and *MAPT*) (Fig. 4b-e). These same four genes as well as two others (*LRCC37A* and *KANSL1*) in this genomic region were also differentially expressed when comparing PP and PA genotypes at 36 months (Fig. 4b-g). Four of the genes in this region whose expression was associated with SVA\_67 (*PLEKHM1*, *ARL17A*, *CRHR1* and *MAPT*) all showed higher expression in

individuals with PP genotypes and the individuals with the lowest expression had an AA genotype. This was in contrast with the levels of expression of *LRCC37A* and *KANSL1* where the opposite pattern was observed.

Table 3

**Differentially expressed genes significantly associated with the genotype of SVA RIPs at baseline (0 months) and 36 months follow up visit-** The number of genes significantly differentially expressed (FDR corrected  $p < 0.1$ ) are reported for each SVA RIP genotype followed by the most significant gene in brackets.

SVA ID	Nearest gene to SVA	No. of significantly differentially expressed genes at baseline			No. of significantly differentially expressed genes at 36 month visit			Tagging SNP for SVA reported as an eQTL in GTEx portal	
		PP v AA	PP v PA	PA v AA	PP v AA	PP v PA	PA v AA	Whole blood	Other tissues
SVA_7	<i>OR1411</i> (4 kb downstream)	0	686 ( <i>EVPL</i> )	2 ( <i>HBA2</i> )	1 ( <i>AC003102.1</i> )	28 ( <i>RAB3B</i> )	3 ( <i>AC003102.1</i> )	N	Y
SVA_11	<i>CASP8</i> (intron)	1 ( <i>FLACC1</i> )	0	64 ( <i>IGSF3</i> )	36 ( <i>IGLV7-46</i> )	15 ( <i>IGLV7-43</i> )	5 ( <i>IFI27</i> )	no tag $r^2 > 0.8$	no tag $r^2 > 0.8$
SVA_24	<i>HLA-A</i> (8 kb upstream)	4 ( <i>HLA-A</i> )	13 ( <i>IGKV1-16</i> )	3 ( <i>HLA-A</i> )	10 ( <i>HLA-A</i> )	7 ( <i>SERINC4</i> )	13 ( <i>HLA-A</i> )	SNP not present	SNP not present
SVA_36	<i>NUPR2</i> (30 kb upstream)	1467 ( <i>TSPAN11</i> )	5 ( <i>PCSK1N</i> )	207 ( <i>FAM155B</i> )	1474 ( <i>RPL31</i> )	36 ( <i>ADGRG7</i> )	823 ( <i>GIMAP2</i> )	Y	Y
SVA_47	<i>RAD23B</i> (10 kb upstream)	0	0	12 ( <i>IL32</i> )	75 ( <i>AC01177.3</i> )	42 ( <i>SNORC</i> )	11 ( <i>PLAU</i> )	N	N
SVA_67	<i>KANSL1</i> (12 kb upstream)	4 ( <i>PLEKHM1</i> )	7 ( <i>ARL17A</i> )	0	22 ( <i>AC120194.1</i> )	25 ( <i>PLEKHM1</i> )	13 ( <i>AC120194.1</i> )	Y	Y
SVA_79	<i>PFDN4</i> (30 kb downstream)	na	1 ( <i>RAB3B</i> )	na	na	0	na	N	N

## Discussion

This study incorporated genetic, clinical and gene expression data from the PPMI cohort to evaluate the role of SVAs in PD, identifying elements associated with disease progression that can modify the transcriptome. In this analysis, we characterised novel variation in the human genome identifying 81 reference SVA RIPs. In the PPMI cohort we did not identify any SVAs that were associated with disease risk. However, in the literature RIPs have been reported as candidate causative variants at known disease associated loci by identifying those in strong LD with trait associated SNPs (15, 21). Utilising this approach we identified one SVA RIP in our cohort in strong LD with a SNP associated with PD risk at genome wide significance. This SVA (SVA\_67) is located in a 1.8 Mb region of high LD on chromosome 17 where there are two predominant haplotypes (H1 and H2) and the presence of SVA\_67 is part of H1 with the SVA is absent in H2. H1 is associated with increased risk of developing PD in multiple studies (22–24). This region encompasses several genes and includes *MAPT* in which mutations can cause the neurodegenerative diseases frontotemporal dementia with parkinsonism and progressive supranuclear palsy (25, 26). However, it remains unclear which variant or gene within the H1 haplotype is responsible for the increased risk to PD and our study has highlighted additional types of variation to be considered at this locus.

The role of SVAs in the progression of PD has not been thoroughly addressed in the literature and here we identified seven SVAs associated with the progression of PD (Table 2). Clinical scales such as the Hoehn and Yahr and the MDS-UPDRS are used to measure disease severity and three of the SVA RIPs were associated with significant differences in these scores between the different genotypes of the insertions. The remaining four SVA RIPs were associated with changes observed in DaTscan SPECT analysis, specifically the ratio of caudate to putamen CDR (Table 2). DaTscan SPECT utilises a radioligand that binds to the DAT that is present on the presynaptic membrane of nigrostriatal neurons. The loss of the DAT from the striatum is characteristic of PD, correlated with the loss of dopaminergic neurons from the substantia nigra and used to distinguish PD from other neurodegenerative parkinsonisms such as essential tremor (27). The reduction of striatal DAT correlates with symptom severity and changes over the course of the disease, therefore this can be used to monitor dopaminergic degeneration (28, 29). The caudate/putamen ratio is a measure of the anterior to posterior gradient of dopaminergic loss and increases in this gradient were associated

with specific SVA RIP genotypes. The SVAs associated with these changes may not be directly causing these alterations to the gradient of dopaminergic loss but may be involved in modifications to the disease course that are measured by these features of DaTscan SPECT.

We have previously shown that SVAs can affect gene expression in *in vitro* and *in vivo* models using reporter gene assays (8, 30). More recently, studies have integrated genetic data regarding SVA RIPs and RNA sequencing to identify those elements that regulate transcription (14, 15). The SVA RIPs associated with progression of PD had a significant impact on the transcriptome and this could be one of the mechanisms by which the SVAs are exerting an effect of PD progression. Genes at both the SVA RIP locus and distant regions of the genome, including different chromosomes, were associated with specific SVA RIP genotypes. For example SVA\_24 was most significantly associated with robust changes of expression of its nearest gene (*HLA-A*) and SVA\_67 was associated with differential expression of a cluster of neighbouring genes. In contrast, SVA\_36 was associated with changes in expression of thousands of genes at many different loci including those on different chromosomes to the SVA RIP. These differences observed could be due to the SVA insertion itself, the genomic architecture at each particular locus or a combination of both. SVAs could influence gene expression through multiple mechanism such as the recruitment of transcription factors, the interaction of SVAs with distant promoters through 3D chromatin structure and intronic insertions affecting mRNA splicing. Regulatory domains can function in a tissue specific manner and therefore we utilised the GTEx database (<https://gtexportal.org/home/>) to evaluate the potential for the SVAs to act in other tissues using their proxy SNPs. Tagging SNPs for five out the seven SVAs were located in the GTEx database and three of these were identified as eQTLs in multiple tissues (Table 3), including different regions of the brain such as the caudate, putamen, substantia nigra and cerebellum. This suggests the SVAs (7, 36 and 67) could also be influencing gene expression in other tissues in addition to whole blood demonstrated in this study.

## Conclusions

Using a systematic approach to analyse SVAs in PD integrating multiple different data types we identified novel genetic variation in the PPMI cohort and specific SVA RIPs that were associated with dopaminergic degeneration. We were able to correlate these elements with a regulatory function and could therefore be influencing the disease course through modulation of gene expression.

## Abbreviations

DAT – dopamine transporter

LD – linkage disequilibrium

MDS-UPDRS - Movement Disorder Society – Unified Parkinson’s Disease Rating Scale

PD – Parkinson’s disease

PPMI – Parkinson’s progression marker initiative

RIP – retrotransposon insertion polymorphism

SWEDD - scans without evidence of dopaminergic deficit

SVA – SINE-VNTR-Alu

## Declarations

## Ethics approval and consent to participate

Written informed consent was obtained by PPMI investigators for subjects to be included in the PPMI study.

## Consent for publication

Not applicable

## Availability of data and materials

The data used in this study was obtained from the Parkinson’s Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)). Access to sequencing and phenotype data is controlled by PPMI.

## Competing interests

The authors declare that they have no competing interests.

## Funding

ALP and SK are funded by MSWA, The Michael J. Fox Foundation, Shake It Up Australia and The Perron Institute.

## Author Contributions

ALP and SK devised the study, performed the data analysis and generated figures for the manuscript. Data interpretation was completed by ALP, VJB, JPQ and SK. ALP and SK wrote the original draft of the manuscript with review and editing by VJB and JPQ. All authors read and approved the final manuscript.

## Acknowledgments

This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)). For up-to-date information on the study, visit [www.ppmi-info.org](http://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, Allergan, Amathus therapeutics, Avid Radiopharmaceuticals, Biogen Idec, Biolegend, Briston-Myers Squibb, Celgene, Denali, GE Healthcare, Genentech, GlaxoSmithKline, janssen neuroscience, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi Genzyme, Servier, Takeda, Teva, UCB, Verily and Voyager Therapeutics.

## References

1. Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. *J Neural Transm (Vienna)*. 2017;124(8):901–5.
2. Shulman JM, De Jager PL, Feany MB. Parkinson's disease: genetics and pathogenesis. *Annu Rev Pathol*. 2011;6:193–222.
3. Singleton A, Hardy J. Progress in the genetic analysis of Parkinson's disease. *Hum Mol Genet*. 2019;28(R2):R215-R8.
4. Billingsley KJ, Barbosa IA, Bandres-Ciga S, Quinn JP, Bubb VJ, Deshpande C, et al. Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset. *NPJ Parkinsons Dis*. 2019;5:8.
5. Hernandez DG, Reed X, Singleton AB. Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. *J Neurochem*. 2016;139(Suppl 1):59–74.
6. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18(12):1091–102.
7. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol*. 2020;19(2):170–8.
8. Savage AL, Bubb VJ, Breen G, Quinn JP. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol Biol*. 2013;13:101.
9. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, et al. SVA elements: a hominid-specific retroposon family. *J Mol Biol*. 2005;354(4):994–1007.
10. Damert A, Raiz J, Horn AV, Lower J, Wang H, Xing J, et al. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res*. 2009;19(11):1992–2008.
11. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH. Jr. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet*. 2011;20(17):3386–400.
12. Hancks DC, Kazazian HH. Jr. Roles for retrotransposon insertions in human disease. *Mob DNA*. 2016;7:9.
13. Aneichyk T, Hendriks WT, Yadav R, Shin D, Gao D, Vaine CA, et al. Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. *Cell*. 2018;172(5):897–909 e21.
14. Wang L, Rishishwar L, Marino-Ramirez L, Jordan IK. Human population-specific gene expression and transcriptional network modification with polymorphic transposable elements. *Nucleic Acids Res*. 2017;45(5):2318–28.
15. Wang L, Norris ET, Jordan IK. Human Retrotransposon Insertion Polymorphisms Are Associated with Health and Disease via Gene Regulatory Phenotypes. *Front Microbiol*. 2017;8:1418.

16. Parkinson Progression Marker I. The Parkinson Progression Marker Initiative (PPMI). *Prog Neurobiol.* 2011;95(4):629–35.
17. Marek K, Chowdhury S, Siderowf A, Lasch S, Coffey CS, Caspell-Garcia C, et al. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Ann Clin Transl Neurol.* 2018;5(12):1460–77.
18. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics.* 2012;28(18):i333-i9.
19. Tang W, Mun S, Joshi A, Han K, Liang P. Mobile elements contribute to the uniqueness of human genome with 15,000 human-specific insertions and 14 Mbp sequence increase. *DNA Res.* 2018;25(5):521–33.
20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75.
21. Payer LM, Steranka JP, Yang WR, Kryatova M, Medabalimi S, Ardeljan D, et al. Structural variants caused by Alu insertions are associated with risks for many human diseases. *Proc Natl Acad Sci U S A.* 2017;114(20):E3984-E92.
22. Li J, Ruskey JA, Arnulf I, Dauvilliers Y, Hu MTM, Hogg B, et al. Full sequencing and haplotype analysis of MAPT in Parkinson's disease and rapid eye movement sleep behavior disorder. *Mov Disord.* 2018;33(6):1016–20.
23. Tobin JE, Latourelle JC, Lew MF, Klein C, Suchowersky O, Shill HA, et al. Haplotypes and gene expression implicate the MAPT region for Parkinson disease: the GenePD Study. *Neurology.* 2008;71(1):28–34.
24. Wider C, Vilarino-Guell C, Jasinska-Myga B, Heckman MG, Soto-Ortolaza AI, Cobb SA, et al. Association of the MAPT locus with Parkinson's disease. *Eur J Neurol.* 2010;17(3):483–6.
25. Im SY, Kim YE, Kim YJ. Genetics of Progressive Supranuclear Palsy. *J Mov Disord.* 2015;8(3):122–9.
26. Strang KH, Golde TE, Giasson BI. MAPT mutations, tauopathy, and mechanisms of neurodegeneration. *Lab Invest.* 2019;99(7):912–28.
27. Bajaj N, Hauser RA, Grachev ID. Clinical utility of dopamine transporter single photon emission CT (DaT-SPECT) with (123I) ioflupane in diagnosis of parkinsonian syndromes. *J Neurol Neurosurg Psychiatry.* 2013;84(11):1288–95.
28. Ikeda K, Ebina J, Kawabe K, Iwasaki Y. Dopamine Transporter Imaging in Parkinson Disease: Progressive Changes and Therapeutic Modification after Anti-parkinsonian Medications. *Intern Med.* 2019;58(12):1665–72.
29. Simuni T, Siderowf A, Lasch S, Coffey CS, Caspell-Garcia C, Jennings D, et al. Longitudinal Change of Clinical and Biological Measures in Early Parkinson's Disease: Parkinson's Progression Markers Initiative Cohort. *Mov Disord.* 2018;33(5):771–82.
30. Savage AL, Wilm TP, Khursheed K, Shatunov A, Morrison KE, Shaw PJ, et al. An evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS. *PLoS One.* 2014;9(6):e90833.

## Figures

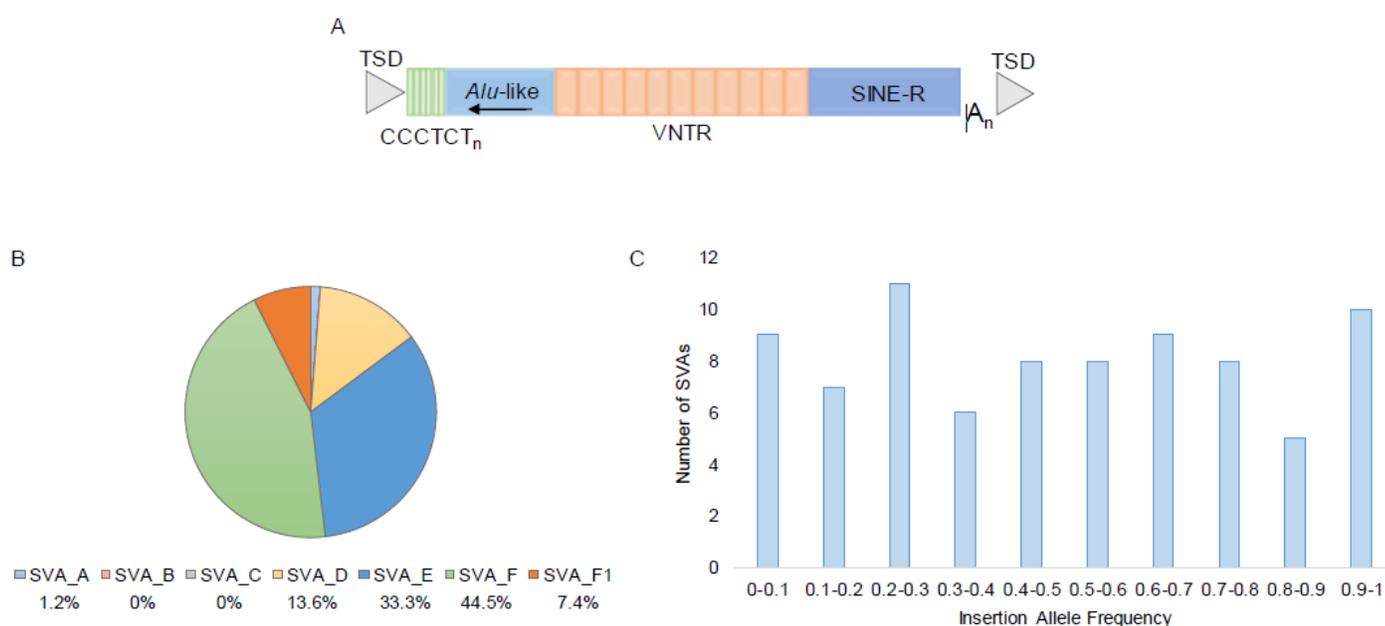
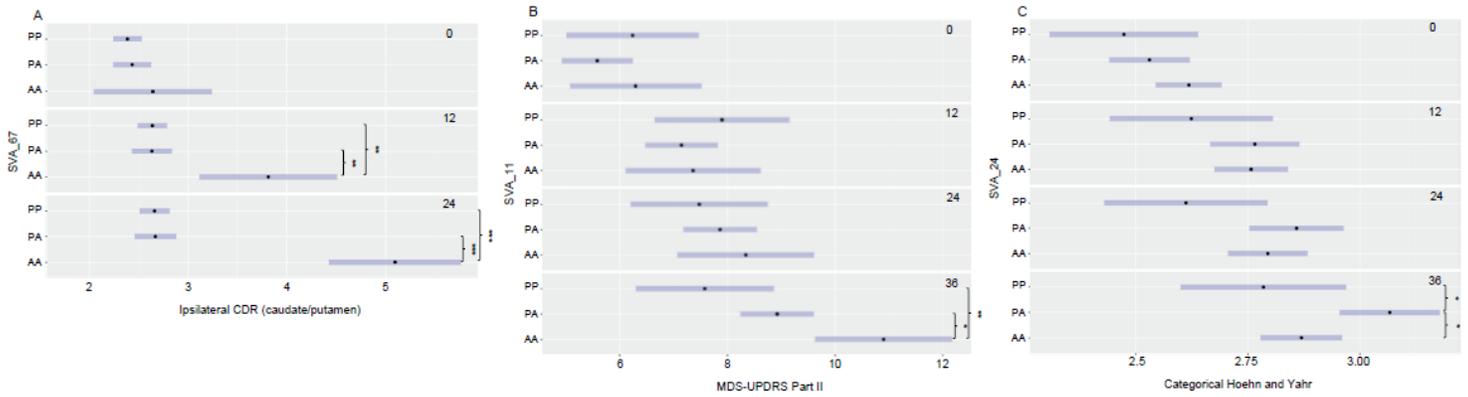


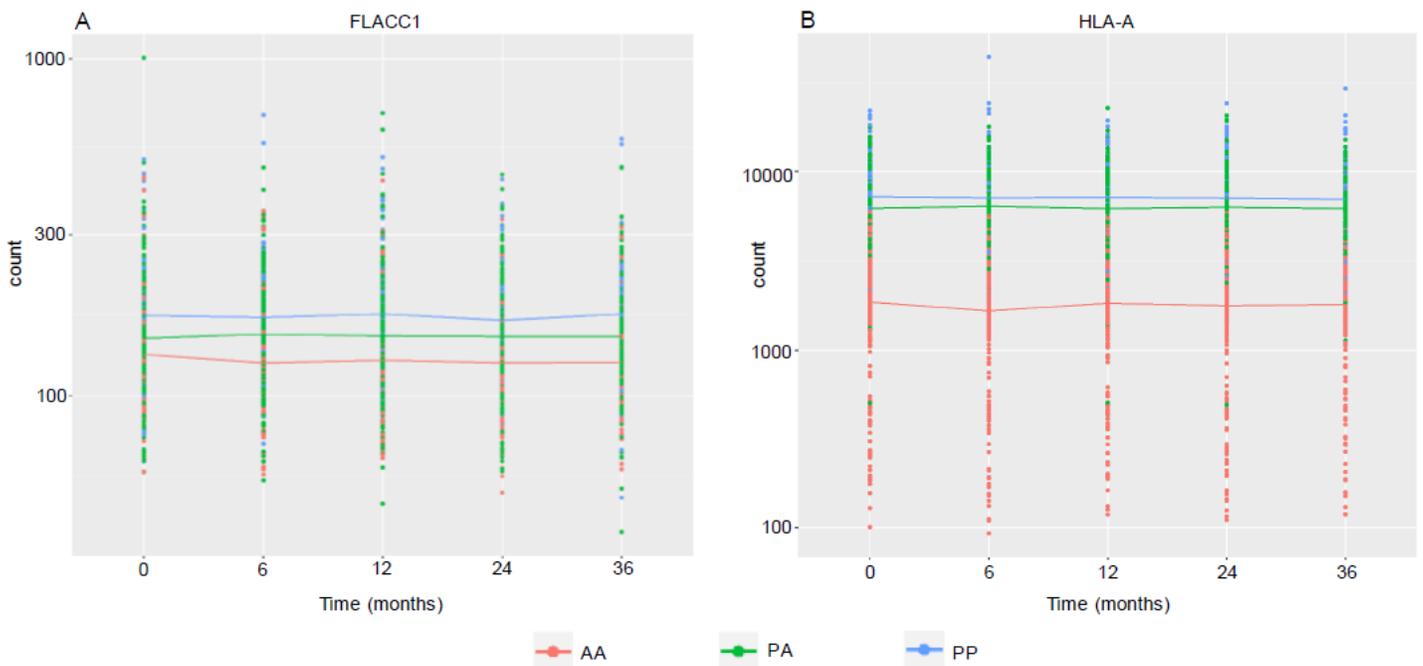
Figure 1

Structure, subtype classification and frequency of polymorphic reference SVAs in the PPMI cohort. A- The structure of a canonical SVA consists of a (CCCTCT)<sub>n</sub> hexamer repeat, an Alu-like region consisting of two antisense Alu fragments and an intervening unique sequence, a variable number tandem repeat (VNTR), a SINE-R domain and a poly A-tail. The length of an intact SVA can vary depending on the number of repeats present in the hexamer and VNTR domains. SVAs are flanked by target site duplications (TSDs). B- The proportion of the 81 reference SVA RIPs that belong to each of the seven subtypes of SVA. C- The number of reference SVA RIPs that belong in each category of frequency of allele insertion.



**Figure 2**

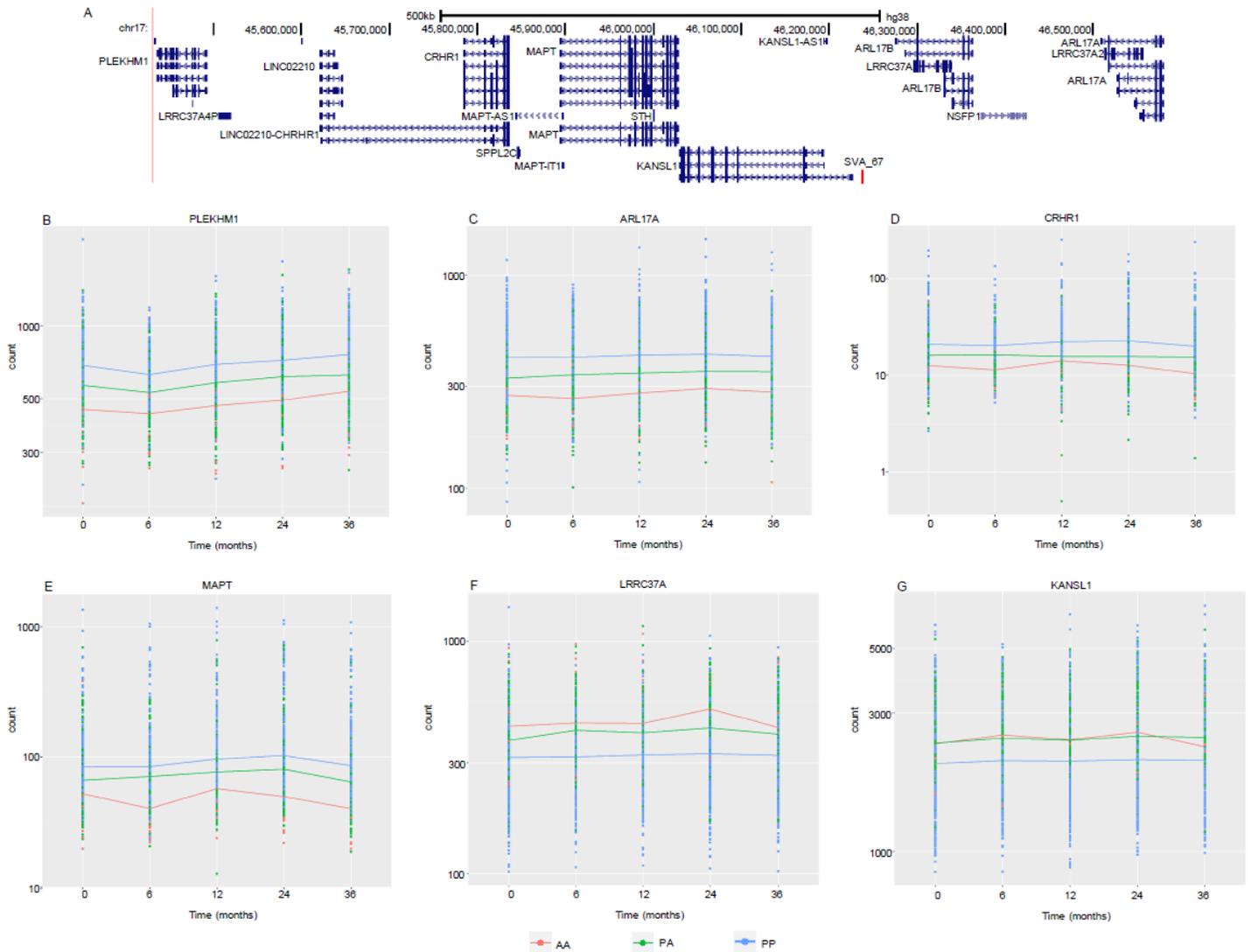
Polymorphic SVAs were associated with markers of PD progression in the PPMI cohort. A- PD individuals with the AA genotype of SVA<sub>67</sub> had a significantly higher ipsilateral caudate/putamen CDR at both 12 and 24 months compared to those with PA and PP genotypes. B- PD individuals with the AA genotype of SVA<sub>11</sub> had a significantly higher MDS-UPDRS part II score at 36 months compared to those with PA and PP genotypes. C- PD individuals with the PA genotype of SVA<sub>24</sub> had a significantly higher Hoehn and Yahr score at 36 months compared to those with AA and PP genotypes. 95% confidence intervals are represented by the blue bar. \*= $p < 0.05$ , \*\*= $p < 0.01$  and \*\*\*= $p < 0.001$



**Figure 3**

The genotype of reference SVAs were associated with differential gene expression. A- At 0 months a 1.3 fold change in expression of FLACC1 was significantly associated with SVA<sub>11</sub> genotype when comparing PP to AA (FDR  $p = 0.04$ ) and at 36 months a 1.4 fold change in expression when comparing PP to AA (FDR  $p = 2.22 \times 10^{-5}$ ). B- At 0 months a 3 fold change in expression of HLA-A was significantly associated with

SVA<sub>24</sub> genotype when comparing PA to AA (FDR  $p=1.90 \times 10^{-78}$ ) and a 3.7 fold change in expression when comparing PP to AA (FDR  $p=6.22 \times 10^{-51}$ ). At 36 months a 2.5 fold change in expression of HLA-A was significantly associated with SVA<sub>24</sub> genotype when comparing PA to AA (FDR  $p=7.51 \times 10^{-60}$ ) and a 3.6 fold change in expression when comparing PP to AA (FDR  $p=2.49 \times 10^{-38}$ ).  $n=608$



**Figure 4**

The expression of several genes at the 17q21.31 locus were associated with SVA<sub>67</sub>. A- A schematic modified from UCSC genome browser of the 1.15Mb region on chromosome 17 that contains six of the genes whose expression was associated with SVA<sub>67</sub>. B-G – The levels of expression of six genes (PLEKHM1, ARL17A, CRHR1, MAPT, LRRRC37A and KANSL1) from individuals with different genotypes of SVA<sub>67</sub> (PP, PA and AA). Differential expression of these genes was significantly associated with specific SVA<sub>67</sub> genotypes at 0 or 36 months or both.  $n=608$

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalTable1.xlsx](#)
- [AdditionalTable2.xlsx](#)