



OPEN

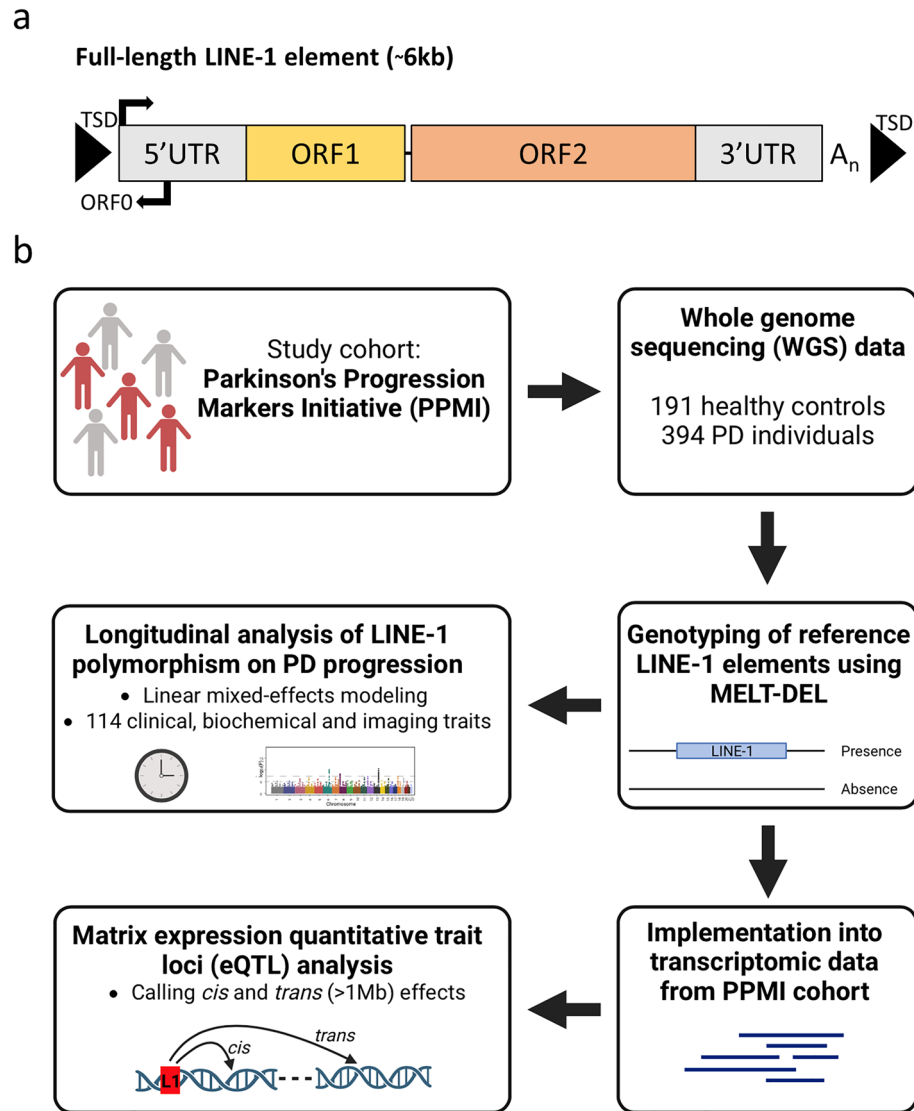
## Reference LINE-1 insertion polymorphisms correlate with Parkinson's disease progression and differential transcript expression in the PPMI cohort

Alexander Fröhlich<sup>1,2</sup>, Abigail L. Pfaff<sup>2,3</sup>, Vivien J. Bubb<sup>1</sup>, John P. Quinn<sup>1</sup>✉ & Sulev Koks<sup>2,3</sup>✉

Long interspersed nuclear element-1 (LINE-1/L1) retrotransposons make up 17% of the human genome. They represent one class of transposable elements with the capacity to both mobilize autonomously and in *trans* via the mobilization of other elements, primarily *Alu* and SVA elements. Reference LINE-1 elements are, by definition, found in the reference genome, however, due to the polymorphic nature of these elements, variation for presence or absence is present within the population. We used a combination of clinical and transcriptomic data from the Parkinson's Progression Markers Initiative (PPMI) and applied matrix expression quantitative trait loci analysis and linear mixed-effects models involving 114 clinical, biochemical and imaging data from the PPMI cohort to elucidate the role of reference LINE-1 insertion polymorphism on both gene expression genome-wide and progression of Parkinson's disease (PD). We demonstrate that most LINE-1 insertion polymorphisms are capable of regulating gene expression, preferentially in *trans*, including previously identified PD risk loci. In addition, we show that 70 LINE-1 elements were associated with longitudinal changes of at least one PD progression marker, including ipsilateral count density ratio and UPDRS scores which are indicators of degeneration and severity. In conclusion, this study highlights the effect of the polymorphic nature of LINE-1 retrotransposons on gene regulation and progression of PD which underlines the importance of analyzing transposable elements within complex diseases.

Long interspersed nuclear element 1 (LINE-1/L1) retrotransposons represent transposable elements which are capable of autonomous propagation throughout the genome via a "copy-&-paste" mechanism, called target-primed reverse transcription (TPRT)<sup>1,2</sup>. A full-length LINE-1 element is about 6kb in length and consists of a 5' untranslated region (UTR), three open reading frames (ORF0, ORF1, and ORF2), a 3'UTR, and a poly A-tail (Fig. 1a)<sup>3</sup>. Through the process of TPRT, new LINE-1 insertions are flanked 5' and 3' by duplicated sequences called target site duplications (TSD) which are regarded as hallmarks of LINE-1 mediated retrotransposition<sup>3</sup>. While LINE-1 elements contribute to 17% of the human genome (more than 1 million copies present), only a small subset of 146 elements are full-length; in other words, most LINE-1 elements are 5' truncated and therefore unable to mobilize<sup>4,5</sup>. LINE-1 insertion events can contribute to genome variability within the population and can also lead to disease<sup>6-8</sup>. To date, more than 124 LINE-1-mediated retrotransposition events can be attributed to a variety of diseases, caused by dysfunctional splicing, exon skipping and double strand DNA breaks among others<sup>6-8</sup>. Previous studies have established a connection between the increase in LINE-1 activity and a heightened risk of developing Parkinson's disease (PD). Nielsen et al. conducted a study in 2012 that associated smoking with the loss of LINE-1 DNA methylation, which can result in the de-repression of LINE-1 and potentially increase the risk of PD<sup>9</sup>. Another study indicated that mitochondrial stress may also trigger LINE-1 activation

<sup>1</sup>Department of Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK. <sup>2</sup>Perron Institute for Neurological and Translational Science, Perth, WA, Australia. <sup>3</sup>Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Perth, WA, Australia. ✉email: jquinn@liverpool.ac.uk; Sulev.Koks@murdoch.edu.au



**Figure 1.** Structure of LINE-1 element and overview of study. **(a)** A full-length LINE-1 element is about 6kb in length and consists of a 5' and 3' untranslated region (UTR), three open reading frames (ORF0, ORF1, ORF2) and a poly(A) tail. The element is flanked by target site duplications (TSDs). The 5'UTR involves a sense and antisense promoter. **(b)** In this study, a combination of whole genome sequencing and transcriptomic datasets from the Parkinson's Progression Markers Initiative (PPMI) was used. Aim of the study was to associate reference LINE-1 polymorphism with Parkinson's disease progression and to assess its role as expression quantitative trait loci (eQTL).

by causing a loss of LINE-1 methylation<sup>10</sup>. Here, the induction of reactive oxygen species production resulted in reduced levels of LINE-1 methylation, leading to the activation of LINE-1 expression via de-repression<sup>10</sup>.

In 2019, Nalls et al. conducted the largest genome-wide association study (GWAS) for PD to date<sup>11</sup>. The authors identified 90 independent risk signals; in addition, LD score regression (LDSC) and polygenic risk scores (PRS) estimated that 26–36% of PD's genetic predisposition could be attributed to the identified variants, highlighting the significant genetic contribution to PD development<sup>11</sup>. However, a substantial proportion of the heritability remains unexplained. Therefore, other genomic elements, such as repetitive DNA in form of LINE-1 retrotransposons, which are frequently excluded in bioinformatic studies, may account for the missing heritability not detected by GWAS. A more detailed and targeted analysis is required to understand the role of transposable elements within PD and other neurodegenerative diseases. Billingsley et al. performed a genome-wide structural variant analysis by testing over 3,154 structural variant loci in PD which increased the understanding of genetic variations at PD risk loci<sup>12</sup>.

Prior research has demonstrated the potential for analyzing transposable elements (TEs) in the context of complex diseases, with a focus on studying their presence or absence in the genome, termed retrotransposon insertion polymorphism (RIP). For example, an SVA insertion in the *TAF1* gene's intron was found to cause X-linked Dystonia Parkinsonism (XDP)<sup>13–15</sup>. Our recent studies also suggest that polymorphic non-reference

transposable elements and reference SVAs are involved in PD progression<sup>16,17</sup>. In the present study, we focused on reference LINE-1 insertion polymorphisms, which means that these elements are generally found in the human reference genome, however, these elements can be polymorphic for their presence or absence in the population. For analysis, we used the Parkinson's Progression Markers Initiative (PPMI) cohort which allows the impact of LINE-1 elements in PD traits due to its availability of clinical, imaging, transcriptomic and whole-genome sequencing (WGS) data (Fig. 1a). We combined available datasets and reference LINE-1 genotypes in order to elucidate the role of LINE-1 insertion polymorphism on both global gene expression and progression of PD (Fig. 1b). We demonstrate that LINE-1 insertion polymorphisms are capable of significantly regulating gene expression, preferentially in *trans*, including previously identified PD risk loci. These involve transcripts encoded by inositol polyphosphate-5-phosphatase F (*INPP5F*), potassium voltage-gated channel modifier subfamily S member 3 (*KCNS3*) and golgi brefeldin A resistant guanine nucleotide exchange factor 1 (*GBF1*). In addition, we show that 70 LINE-1 elements were associated with longitudinal changes of at least one PD progression marker, including ipsilateral count density ratio and UPDRS scores. Taken together, this study highlights the effect of the polymorphic nature of reference LINE-1 retrotransposons on gene regulation and progression of PD which underlines the importance of analyzing transposable elements within complex diseases and their missing heritability.

## Results

**Reference LINE-1 insertion polymorphisms correlate with PD progression.** In order to analyse the impact of reference LINE-1 retrotransposons on PD progression, we correlated genotypes from 116 LINE-1 retrotransposons with longitudinal data from the PPMI cohort. This involves 114 clinical, biochemical and imaging traits from up to five visits (Baseline/BL, V04, V06, V08, V10, V12). Here we demonstrate that in total 72 clinical progression markers (clinical scores or traits) were affected by LINE-1 insertion polymorphism; in other words, presence or absence of reference LINE-1 retrotransposons significantly correlated with this number of clinical progression markers (FDR  $P < 0.05$ ).

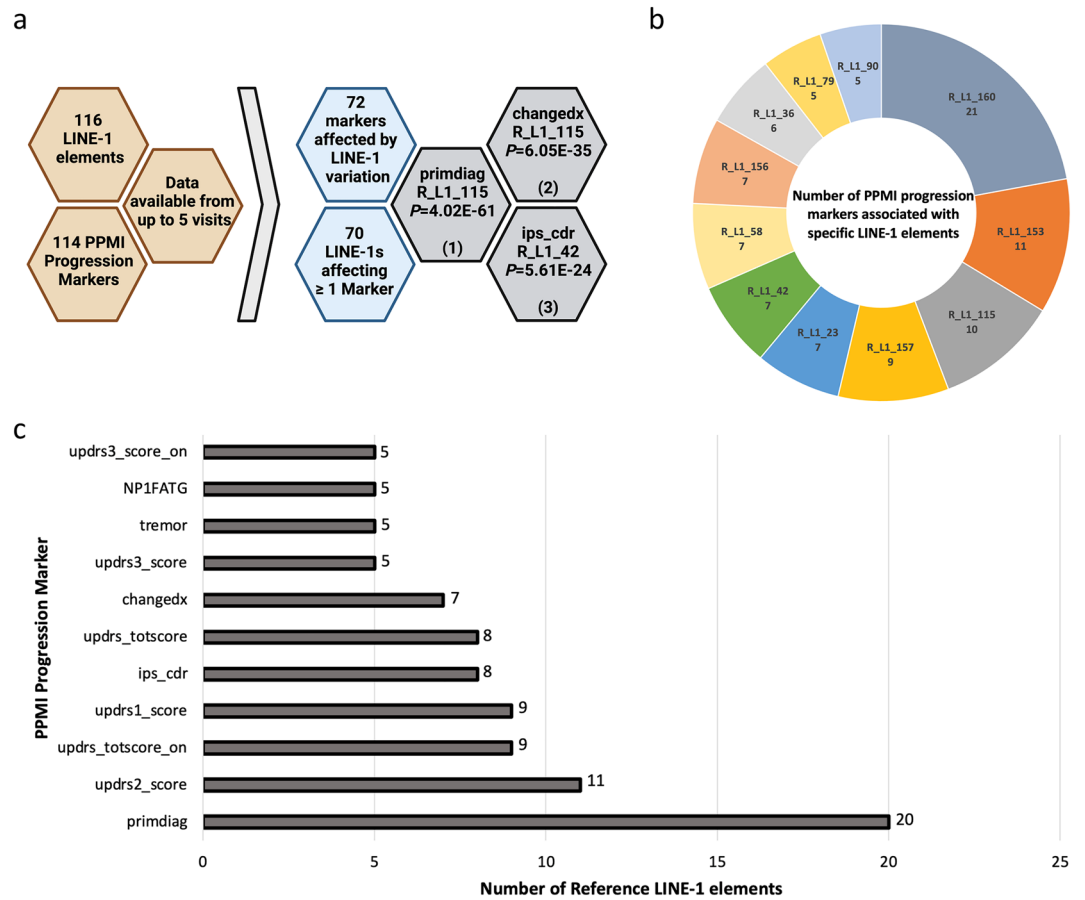
Interestingly, 70 LINE-1 elements correlated with changes of at least one progression marker indicating the plethora of PD progression changes associated with LINE-1 elements (Fig. 2a). The top three significant changes were found in primary diagnosis (R\_L1\_115,  $P = 4.02E-61$ ), changes in primary diagnosis (R\_L1\_115,  $P = 6.05E-35$ ) and ipsilateral count density ratio (R\_L1\_42,  $P = 5.61E-24$ ) (Fig. 2a). Eleven reference LINE-1 retrotransposons showed an effect on at least five progression markers, with R\_L1\_160 displaying the highest association with 21 markers followed by R\_L1\_153 with 11 markers (Fig. 2b). A full-list of LINE-1 elements significantly associating with PPMI progression markers is shown in Supplementary Data 1. In addition, eleven progression traits were affected by at least five LINE-1 elements (Fig. 2c). The highest number of LINE-1 retrotransposons were associated with primary diagnosis (20 elements). Other commonly affected markers represented MDS-UPDRS Part III Score (ON) (5 hits), MDS-UPDRS Part I Fatigue (5 hits), Tremor Score (OFF) (5 hits), MDS-UPDRS Part III Score (OFF) (5 hits), change in primary diagnosis (7 hits), MDS-UPDRS Total Score (OFF) (8 hits), ipsilateral count density ratio (8 hits), MDS-UPDRS Part I Score (9 hits), MDS-UPDRS Total Score (ON) (9 hits) and MDS-UPDRS Part II Score (11 hits) (Fig. 2c).

To expand the impact of LINE-1 elements on these progression traits, we specifically looked at the changes of ipsilateral count density ratio (IPS-CDR) and MDS-UPDRS Part II Score (UPDRS2). Figure 3 shows Manhattan plots indicating the location of LINE-1 elements and the corresponding  $P$  values associated with longitudinal changes of IPS-CDR (Fig. 3a) and UPDRS2 (Fig. 3b). For IPS-CDR, four elements (R\_L1\_6, R\_L1\_92, R\_L1\_123 and R\_L1\_42) showed significant association with FDR  $P < 0.01$  (Fig. 3a). Similarly, UPDRS2 score was significantly (FDR  $P < 0.01$ ) affected by the elements R\_L1\_60, R\_L1\_157 and R\_L1\_115 (Fig. 3b).

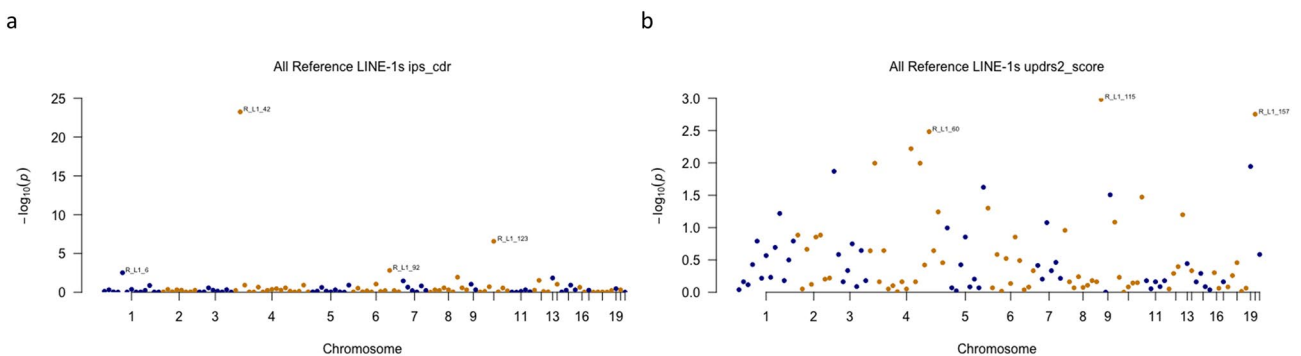
We next analyzed the specific effect of LINE-1 elements showing the most significant association with progression traits (R\_L1\_42 for IPS-CDR and R\_L1\_115 for UPDRS2) (Fig. 4). Figure 4A shows the longitudinal change in IPS-CDR based on presence or absence of R\_L1\_42. Interestingly, individuals with the absence of R\_L1\_42 (AA genotype) show significantly higher ipsilateral count density ratio at visits 4, 6 and 10 compared to the presence of one or two copies of this element (Fig. 4a). When looking at the change in MDS-UPDRS Part II score, only data for presence (PP genotype) or one allele missing (PA genotype) were available. However, individuals with one copy missing of R\_L1\_115 (PA genotype) show significantly higher UPDRS2 score at visit 10 ( $P < 0.01$ ) which indicates that these individuals progressed faster (scores more than 10 points) and the protective role of this specific LINE-1 element regarding PD progression (Fig. 4b).

**Reference LINE-1 retrotransposons act as eQTL genome-wide.** We combined genotyping and blood-derived transcriptomic data from the PPMI cohort to analyse the capacity of reference LINE-1 elements to act as eQTL. For the analysis, we analysed 140 reference LINE-1 elements, of which 53% are located intronic and 47% intergenic (Supplementary Data 4). All LINE-1 elements analysed in this study represent members of the youngest and human-specific L1HS subfamily. We performed matrix eQTL analysis on an isoform-based level which showed that in total 9014 transcripts were significantly (FDR  $P < 0.05$ ) affected by LINE-1 insertion polymorphisms with  $P$  values ranging up to  $2.62E-83$  (Fig. 5a). Interestingly, almost 99% of the transcripts were *trans*-regulated by LINE-1 elements (Fig. 5a) which means that the location of the corresponding affected transcript in relation to the LINE-1 element is  $> 1\text{M}$  bp. Only 94 of 9014 transcripts showed a closer location with respect to the LINE-1 element they are affected by. It is worth mentioning that only six LINE-1 elements had a *cis*-specific effect, specifically R\_L1\_60 (16 transcripts), R\_L1\_155 (9), R\_L1\_145 (4), R\_L1\_147 (3), R\_L1\_63 (3) and R\_L1\_126 (2).

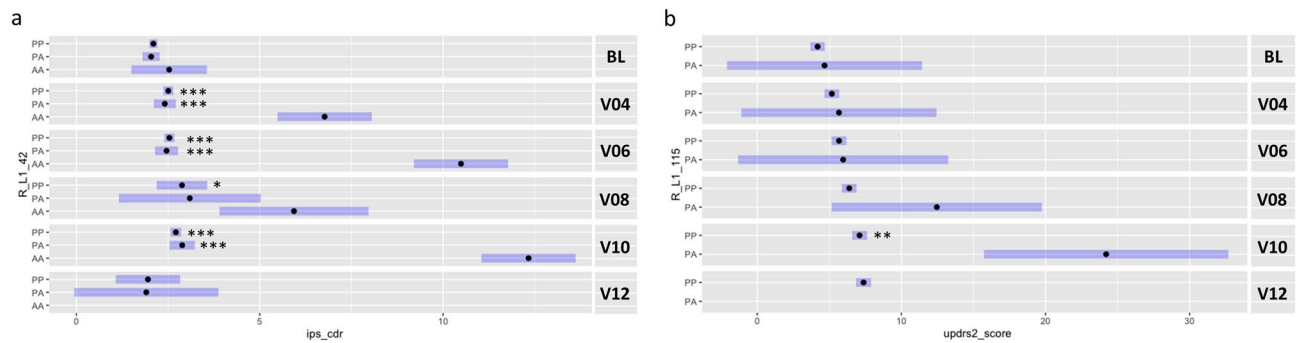
When looking on the element-based level, there are 18 LINE-1 elements present which affect 100 or more transcripts genome-wide. R\_L1\_17 represents the element that affects the highest number of transcripts (1222)



**Figure 2.** Reference LINE-1 elements are associated with PPMI progression markers. **(a)** Longitudinal and clinical data from the PPMI cohort was used for association analysis. 70/116 LINE-1 elements were associated with at least one progression marker. Top three hits were associated with primary diagnosis (primdiag), change in primary diagnosis (changedx) and ipsilateral count density ratio (ips\_cdr). FDR corrected  $P$  values are shown. **(b)** Pie chart representing the total number of PPMI progression markers associated with a single LINE-1 element. R\_L1\_160 correlated with the highest number of markers (21). **(c)** Bar chart displaying the number of LINE-1 elements associated with a single PPMI progression marker. Only progression markers with  $> 5$  hits are shown. Updrs3\_score\_on, MDS-UPDRS Part III Score (ON); NP1FATG, MDS\_UPDRS Part I Fatigue; tremor, Tremor Score (OFF); updrs3\_score, MDS-UPDRS Part III Score (OFF); changedx, Change in Primary Diagnosis from Baseline to current visit; updrs\_totscore, MDS-UPDRS Total Score (OFF); ips\_cdr, Ipsilateral count density ratio; updrs1\_score, MDS-UPDRS Part I Score; updrs\_totscore\_on, MDS\_UPDRS Total Score (ON); updrs2\_score (MDS-UPDRS Part II Score); primdiag, Primary Diagnosis.



**Figure 3.** Manhattan plot for the association of all reference LINE-1 elements with PPMI progression markers. LINE-1 elements were correlated with ipsilateral count density ratio (ips\_cdr) **(a)** and MDS-UPDRS Part II Score (updrs2\_score) **(b)** using FDR  $P$  values for longitudinal changes. Chromosomal locations of these elements in the genome are indicated. Only elements with FDR  $P$  below 0.01 are labeled. FDR corrected  $P$  values are displayed on the Y-axis as  $-\log_{10}$  values.



**Figure 4.** Longitudinal changes in PPMI progression markers based on LINE-1 retrotransposon insertion polymorphism. Presence/absence of specific LINE-1 elements were associated with ipsilateral count density ratio (ips\_cdr, R\_L1\_42) (a) and MDS-UPDRS Part II Score (updrs2\_score, R\_L1\_115) (b). Blue bars represent 95% confidence intervals. Tukey adjusted  $P$  values are reported compared to AA (R\_L1\_42) and PA (R\_L1\_115) genotypes. Genotypes represent the status of LINE-1 elements in the human genome in form of homozygous present (PP), heterozygous (PA) and homozygous absent (AA), respectively. Longitudinal data at baseline/BL (0 months) and up to five visits (V04/12 months, V06/24 months, V08/36 months, V10/48 months, V12/60 months) were used for analysis. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

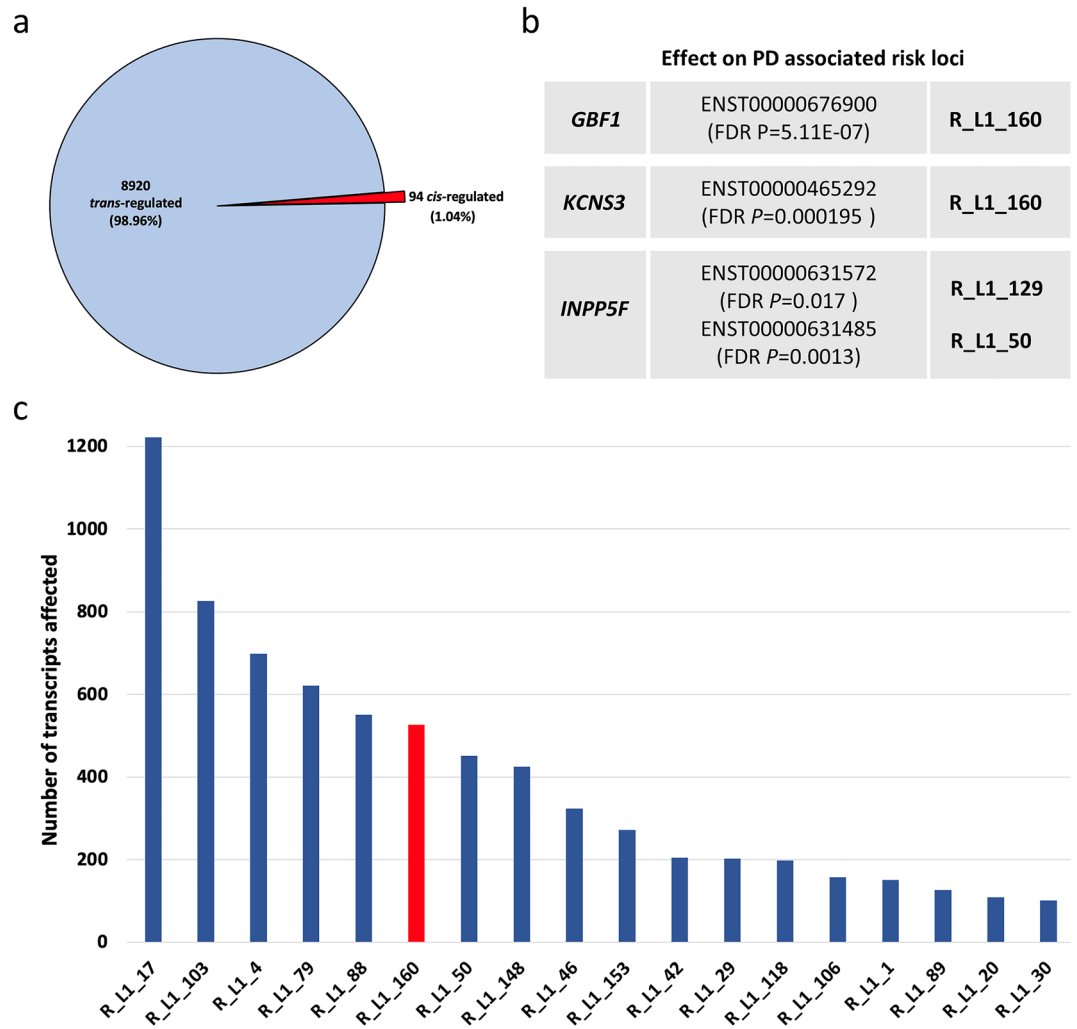
with only one transcript being *cis* regulated (Fig. 5c). Interestingly, the element R\_L1\_160, which previously showed to be associated with the highest number of PD progression markers (Fig. 2), affected 527 transcripts, with no transcripts regulated in *cis*. When looking specifically on the transcripts affected by R\_L1\_160, it can be highlighted that this element affects expression of previously identified novel PD risk loci, *GBF1* and *KCNS3* (Fig. 5b, Supplementary Data 2). Here, R\_L1\_160 insertion polymorphism significantly activated expression of *GBF1* transcript ENST00000676900 (FDR  $P = 5.11E-07$ ) and *KCNS3* transcript ENST00000465292 (FDR  $P = 0.000195$ ). Another PD risk locus represents *INPP5F*, whereby the elements R\_L1\_129 and R\_L1\_50 showed a positive effect on *INPP5F* ENST00000631572 (FDR  $P = 0.017$ ) and ENST00000631485 (FDR  $P = 0.0013$ ), respectively (Fig. 5b, Supplementary Data 2).

## Discussion

In this study we have analyzed the association of reference LINE-1 elements on PD progression and their ability to act as eQTL by regulating gene expression patterns based on polymorphism for presence or absence. Here, we could demonstrate that 70/116 RIP LINE-1 elements had significant influence on at least one PPMI progression marker which involve different clinical, imaging and biochemical traits. Intriguingly, the highest number of LINE-1 retrotransposon insertion polymorphisms (20 elements) were associated with primary diagnosis, in addition, more than 7 elements also correlated with change in primary diagnosis. This means that PD individuals with specific genetic variations related to LINE-1s may have a higher likelihood of initially receiving an incorrect diagnosis during early visits, which may later be corrected. This change in diagnosis could potentially highlight differences in the intermediate phenotypes between individuals with different LINE-1 genotypes. It should be emphasized that accurately diagnosing PD at the onset can be difficult due to the complex phenotype and overlapping extrapyramidal side effects<sup>18</sup>. The discovery that certain LINE-1s correlated with a significantly increased frequency of diagnostic challenges may reflect the genetic heterogeneity of PD and potentially propose the existence of a separate disease subtype.

In addition to that, reference LINE-1 elements also significantly correlated with UPDRS scores. These clinical scales are used to measure disease severity and include, for instance, non-motor (part I) or motor (part II) experiences of daily living. UPDRS Part II score involves measures such as an effect on speech, chewing, dressing, walking or doing activities<sup>19</sup>. Here we could show that presence or absence of 11 LINE-1 elements had a significant impact on the severity of this scale. To be more specifically, the element R\_L1\_115 had the highest impact; in other words, individuals with one copy missing of R\_L1\_115 (PA genotype) show significantly higher UPDRS2 score at visit 10 ( $P < 0.01$ ) which indicates that these individuals progressed faster (scores more than 10 points compared to PP genotype) and the protective role of this specific LINE-1 element on PD progression. This difference with more than 10 points on this clinical scale may make the difference if a PD patient is able to exert the activities mentioned before or not anymore. The study presented here focused on reference LINE-1 elements, however, in our previous study we showed that non-reference (non-ref) transposable elements were also able to affect PD progression<sup>16</sup>. These non-ref elements are not fixed in the reference human genome and can therefore be variable for presence or absence in the genome. In total, 2886 *Alu*, 360 LINE-1, and 128 SVA elements were analysed which showed the highest number of effects on two PD progression traits, namely primary diagnosis (268 hits) and UPDRS2 score (224 hits)<sup>16</sup>. More specifically, when looking at non-ref LINE-1 elements, the most common traits affected were UPDRS2 score, primary diagnosis, and ipsilateral count density ratio of caudate/putamen (ips-CDR). This is consistent with our current study which showed that these traits were also commonly affected by reference LINE-1 elements supporting the effect seen on this PD progression traits.

Another PD progression marker we want to highlight here is the ipsilateral count density ratio which is based on imaging data of the brain which provides a reliable source to characterize PD progression in the form of neurodegeneration. Count density ratios (CDR) of the caudate/putamen are calculated by DaTscan single photon



**Figure 5.** Reference LINE-1 elements act as eQTL in *cis* and *trans*. (a) Matrix eQTL analysis was performed in order to analyze the effect of LINE-1 elements to act as eQTL on a transcript-based level. 9014 transcripts were significantly affected (FDR  $P < 0.05$ ) by LINE-1 insertion polymorphism, whereby the majority (98.96%) being *trans*-regulated by LINE-1 elements. (b) Transcripts of previously identified novel PD risk loci, including *GBF1*, *KCNS3* and *INPP5F* were significantly affected by R\_L1\_160, R\_L1\_129 and R\_L1\_29 polymorphism, respectively. (c) Bar chart showing the number of transcripts affected by specific LINE-1 elements. Only LINE-1 elements with  $> 100$  affected transcripts are shown. R\_L1\_160 which was associated with the highest number of PPMI progression markers (Fig. 2) is highlighted in red.

emission computed tomography (DaTscan SPECT) imaging that uses a radioligand to bind to the dopamine transporter (DAT) located on the presynaptic membrane of nigrostriatal neurons. The loss of DAT from the striatum is a defining characteristic of PD and is correlated with the loss of dopaminergic neurons in the substantia nigra. This loss of DAT can be used to distinguish PD from other neurodegenerative parkinsonian disorders such as essential tremor<sup>20</sup>. The severity of symptoms in PD is related to the reduction of striatal DAT, which changes over the course of the disease, making it an important tool for detecting dopaminergic degeneration<sup>21,22</sup>. The caudate/putamen ratio is a measure of the gradient of dopaminergic loss from the anterior to posterior regions, which specifically increases when the R\_L1\_42 is absent in the genome. It's important to note that while the LINE-1s associated with these changes may not directly cause changes to the dopaminergic loss gradient, they may be a part of the process that modifies PD disease course. One mechanism involved could be transcriptional changes of genes located in PD risk loci. Transposable elements are known to have the ability to modulate gene expression levels<sup>23–25</sup>; we and others have previously shown that TE insertion polymorphisms can act as eQTL and have an activating or repressive effect on transcripts in proximity (*cis*) or across chromosomes (*trans*)<sup>26–28</sup>. Differential gene expression triggered by the presence/absence or primary sequence variation of specific LINE-1 or other TE elements may be one mechanism by which PD course and progression can be influenced. In this study we could show that in total 9014 transcripts were significantly modified by reference LINE-1 insertion polymorphisms. Interestingly, almost 99% of these transcripts were *trans*-regulated by LINE-1 presence or absence. This is consistent with previous literature targeting the impact of TEs on gene expression which involves a study by Wang et al. which identified a greater number of *trans*-regulating TE elements using data from the 1000 genome

project<sup>27</sup>. In our study, we used identical analysis models which increases the confidence of comparing these data with the aim to pinpoint the significant effect of transposable elements (here LINE-1 elements) on gene regulation which may also have an impact on pathways involved in neurodegeneration. It is worth mentioning that we identified transcripts encoded by previously identified PD risk loci *KCNS3*, *GBF1* and *INPP5F*<sup>11</sup> that were also affected by LINE-1 insertion polymorphism.

The results showed here benefited from the use of longitudinal data from the PPMI cohort which provides various clinical, biochemical and imaging data to monitor the course and progression of PD individuals. However, functional work in follow-up studies are necessary to validate the effects of LINE-1 insertion polymorphism. This can be achieved by analysing patient-derived cell lines harbouring the corresponding LINE-1 elements. The effects can be examined by CRISPR, followed by measurement of gene expression or cellular phenotype. Furthermore, transgenic mice can be utilised to model the impact of LINE-1 elements in form of overexpression experiments. Such an approach enables the observation of changes in brain function and behaviour, which can be tailored to the specific phenotype of Parkinson's disease. In summary, this study described the significant impact of reference LINE-1 insertion polymorphism on 1) progression of PD by using a variety of PD progression traits, whereby a significant association with severity and degeneration of PD was observed, and 2) transcriptional regulation genome wide; thousands of transcripts were significantly *cis*- and *trans*-regulated by LINE-1 presence or absence, including transcripts encoded by PD risk genes. This study therefore highlights the plethora of transcriptomic and phenotypic changes associated with reference LINE-1 polymorphism which should be considered when analysing the role of transposable elements within neurodegenerative diseases and their missing heritability.

## Methods

**Study cohort.** For this research, we obtained data from the Parkinson's Progression Markers Initiative (PPMI) cohort, which can be accessed at <http://www.ppmi-info.org/data>. The PPMI is a longitudinal cohort study designed to track the course of Parkinson's disease. The dataset contains 423 PD and 196 control individuals including whole transcriptome data from blood samples as well as genetic and clinical information.

**Genotyping of reference LINE-1 elements in the PPMI cohort.** Reference LINE-1 elements were identified using whole-genome sequencing (WGS) data obtained from the PPMI cohort. The mobile element locator tool-deletion (MELT-DEL) (<https://melt.igs.umaryland.edu/>) was used to genotype reference LINE-1 elements. WGS data in bam format aligned to Hg38 were used as input for MELT-DEL. In each individual, MELT-DEL genotyped reference LINE-1 elements based on their coordinates, which were provided in a bed file. The output files were then merged to produce a final VCF. Reference LINE-1 polymorphic insertions that were not in Hardy-Weinberg equilibrium ( $p < 1 \times 10E-06$  in healthy controls) were removed using plink v1.07<sup>29</sup>. The chromosomal location and coordinates of each LINE-1 element are listed in Supplementary Data 4.

**Longitudinal analysis of reference LINE-1 elements on PD progression.** The longitudinal analysis focused solely on the data of PD patients from the PPMI cohort, while control subjects' data was excluded as the aim of the study was to investigate the impact of reference LINE-1 retrotransposons on PD progression. The PPMI dataset comprises longitudinal data at baseline/BL (0 months) and up to five visits (V04/12 months, V06/24 months, V08/36 months, V10/48 months, V12/60 months) of 423 PD patients. In total, 114 clinical, imaging and biochemical traits were selected for analysis (Supplementary Data 3). The impact of the presence or absence polymorphism of LINE-1 on clinical traits was analyzed using linear mixed-effects modelling and the *LmerTest* R package. The following formula was employed for longitudinal modelling of the effect of reference LINE-1s on changes in traits between visits:

$$\text{anova(lmerTest::lmer(\text{TRAIT} \sim \text{REF-L1} * \text{months} + (1|\text{PATNO}), \text{na.action} = \text{na.omit}, \text{data} = \text{PD}))}$$

The *P* values obtained were subjected to false discovery rate (FDR) adjustment for multiple corrections, and only FDR values below 0.05 were deemed statistically significant. The corrected FDR values were then utilized to identify significant traits for the pairwise analysis of genotype effects using the *emmeans* package in R. Pairwise comparisons *P* values were adjusted with Tukey for family-wise error rate. The use of longitudinal PPMI clinical and genomic data was approved (approval number: 2021/123) by the Human ethics committee of the Murdoch University.

**Expression quantitative trait loci (eQTL) analysis of reference LINE-1 retrotransposons.** We combined determined LINE-1 genotypes with transcriptomic data obtained from the PPMI cohort to investigate the capacity of reference LINE-1 retrotransposons to act as eQTL. The Salmon tool was used to quantify transcriptomic data at the transcript-based level, and the *tximport* function from the *tximport* package was used to import the Salmon-generated quant files into R. Raw counts were extracted and then normalized using the median-of-ratios method included in the *DESeq2* package. We combined PD and healthy control individual data for the analysis. Matrix eQTL was used to determine the genetic loci regulating the expression of transcript variants. We used an additive linear model with age and sex as covariates and set the FDR threshold to 0.05. During the eQTL analysis, we called local (*cis*) and distant (*trans*) quantitative loci and set the distant locus threshold to 1M bp. The correction for multiple testing of eQTL was performed using FDR, and we only report results that remained significant after FDR correction.

## Data availability

The datasets used and/or analysed during the current study are publicly available from the Parkinson's Progression Marker Initiative ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)).

Received: 18 May 2023; Accepted: 21 August 2023

Published online: 24 August 2023

## References

1. Raiz, J. *et al.* The non-autonomous retrotransposon SVA is trans -mobilized by the human LINE-1 protein machinery. *Nucl. Acids Res.* **40**, 1666–1683 (2012).
2. Kim, S., Cho, C.-S., Han, K. & Lee, J. Structural variation of Alu element and human disease. *Genom. Inform.* **14**, 70 (2016).
3. Hancks, D. C. & Kazazian, H. H. Roles for retrotransposon insertions in human disease. *Mob DNA* **7**, 9 (2016).
4. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
5. Penzkofer, T. *et al.* L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucl. Acids Res.* **45**, D68–D73 (2017).
6. Gasior, S. L., Wakeman, T. P., Xu, B. & Deininger, P. L. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* **357**, 1383–1393 (2006).
7. Belancio, V. P., Hedges, D. J. & Deininger, P. Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. *Genome Res.* **18**, 343–358 (2008).
8. Wallace, N. A., Belancio, V. P. & Deininger, P. L. L1 mobile element expression causes multiple types of toxicity. *Gene* **419**, 75–81 (2008).
9. Nielsen, S. S. *et al.* LINE-1 DNA methylation, smoking and risk of Parkinson's disease. *J. Parkinsons Dis.* **2**, 303–308 (2012).
10. Baeken, M. W., Moosmann, B. & Hajjeva, P. Retrotransposon activation by distressed mitochondria in neurons. *Biochem. Biophys. Res. Commun.* **525**, 570–575 (2020).
11. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
12. Billingsley, K. J. *et al.* Genome-wide analysis of structural variants in Parkinson disease. *Ann. Neurol.* <https://doi.org/10.1002/ana.26608> (2023).
13. Anechych, T. *et al.* Dissecting the causal mechanism of X-linked dystonia-parkinsonism by integrating genome and transcriptome assembly. *Cell* **172**, 897–909.e21 (2018).
14. Makino, S. *et al.* Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am. J. Hum. Genet.* **80**, 393–406 (2007).
15. Bragg, D. C. *et al.* Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. *Proc. Natl. Acad. Sci.* **114**, E11020–E11028 (2017).
16. Koks, S., Pfaff, A. L., Singleton, L. M., Bubb, V. J. & Quinn, J. P. Non-reference genome transposable elements (TEs) have a significant impact on the progression of the Parkinson's disease. *Exp. Biol. Med.* **247**, 1680–1690 (2022).
17. Pfaff, A. L., Bubb, V. J., Quinn, J. P. & Koks, S. Reference SVA insertion polymorphisms are associated with Parkinson's disease progression and differential gene expression. *NPJ. Parkinsons Dis.* **7**, 44 (2021).
18. Zhang, P.-L., Chen, Y., Zhang, C.-H., Wang, Y.-X. & Fernandez-Funez, P. Genetics of Parkinson's disease and related disorders. *J. Med. Genet.* **55**, 73–80 (2018).
19. Goetz, C. G. *et al.* Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Mov. Disord.* **23**, 2129–2170 (2008).
20. Bajaj, N., Hauser, R. A. & Grachev, I. D. Clinical utility of dopamine transporter single photon emission CT (DaT-SPECT) with (123 I) ioflupane in diagnosis of parkinsonian syndromes. *J. Neurol. Neurosurg. Psychiatr.* **84**, 1288–1295 (2013).
21. 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.* **58**, 1665–1672 (2019).
22. Simuni, T. *et al.* Longitudinal change of clinical and biological measures in early Parkinson's disease: Parkinson's progression markers initiative cohort. *Mov. Disord.* **33**, 771–782 (2018).
23. Savage, A. L., Bubb, V. J., Breen, G. & Quinn, J. P. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol. Biol.* **13**, 101 (2013).
24. Savage, A. L. *et al.* An Evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS. *PLoS ONE* **9**, e90833 (2014).
25. Fröhlich, A., Pfaff, A. L., Bubb, V. J., Koks, S. & Quinn, J. P. Characterisation of the function of a SINE-VNTR-Alu retrotransposon to modulate isoform expression at the MAPT locus. *Front. Mol. Neurosci.* **15**, 815695 (2022).
26. Wang, L., Norris, E. T. & Jordan, I. K. Human retrotransposon insertion polymorphisms are associated with health and disease via gene regulatory phenotypes. *Front. Microbiol.* **8**, 1418 (2017).
27. Wang, L., Rishishwar, L., Mariño-Ramírez, L. & Jordan, I. K. Human population-specific gene expression and transcriptional network modification with polymorphic transposable elements. *Nucl. Acids Res.* <https://doi.org/10.1093/nar/gkw1286> (2016).
28. Koks, S., Pfaff, A. L., Bubb, V. J. & Quinn, J. P. Expression quantitative trait loci (eQTLs) associated with retrotransposons demonstrate their modulatory effect on the transcriptome. *Int. J. Mol. Sci.* **22**, 6319 (2021).
29. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

## Acknowledgements

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, Biogen, Bristol-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. This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

## Author contributions

A.F. and S.K. designed and coordinated experiments. A.F., A.L.P., V.J.B., J.P.Q. and S.K. interpreted and analysed the data, A.L.P. performed genotyping experiments. A.F. and S.K. performed bioinformatic analysis of eQTL and longitudinal association with progression markers. A.F. wrote the manuscript. Each author contributed to manuscript editing and review.



## Funding

Andrzej Wlodarski Memorial Research Fund, MSWA and Perron Institute for Neurological and Translational Science, MND, Quinn/Apr20/875-791, Quinn/Apr20/875-791, Quinn/Apr20/875-791, Quinn/Apr20/875-791, Darby Rimmer Foundation.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-41052-1>.

**Correspondence** and requests for materials should be addressed to J.P.Q. or S.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023