Longitudinal intronic RNA-seq analysis of Parkinson's Disease patients reveals disease-specific genome-wide nascent transcription

Sulev Koks [1,2,*], Abigail L. Pfaff [1,2], Vivien J. Bubb [3] and John P. Quinn [3]

[1] Perron Institute for Neurological and Translational Science, Perth, WA 6009, Australia

[2] Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Perth, WA 6150, Australia

[3] Department of Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 3BX, UK

[*] Correspondence: sulev.koks@perron.uwa.edu.au; Tel.: +61-(0)-8-6457-0313

Abstract

Background: Transcriptomic studies usually focus on gene or exon-based annotations, and only limited experiments have reported the changes in the reads mapping to introns. The analysis of intronic reads allows the detection of nascent transcription that is not influenced by steady -state RNA levels and provides information on actively transcribed genes.

Methods: Here we describe substantial intronic transcriptional changes in Parkinson's Disease (PD) patients compared to healthy controls (CO) at two different timepoints, at the time of diagnosis (BL) and three years later (V08). We used blood RNA-seq data from the Parkinson's Progression Markers Initiative (PPMI) cohort subjects and identified significantly changed transcription of intronic reads only in PD patients during this follow up time.

Results: In CO subjects, only nine transcripts demonstrated differentially expressed introns between visits BL and V08. However, in PD patients 4,873 transcripts had differentially

expressed introns at visit V08 compared to BL, many of them with clear connection to neurodegenerative diseases, such as *LRRK2, C9orf72, LGALS3, KANSL1AS1, ALS2*. In addition, at the time of diagnosis (BL visit) we identified 836 transcripts (e.g. *SNCA, DNAJC19, PRRG4*) and at visit V08 2,184 transcripts (e.g. *PINK1, GBA, ALS2, PLEKHM1*) with differential intronic expression specific to PD patients. In contrast, reads mapping to exonic regions demonstrated little variation indicating high specificity of changes in intronic transcription.

Conclusion: Our study shows that Parkinson's disease is characterized by substantial changes in the nascent transcription and description of these changes could help to understand the molecular pathology underpinning this disease.

Keywords

Parkinson Disease, transcriptome; whole transcriptome analysis; introns; RNA-Seq; nascent transcript; nascent RNA; PPMI; blood transcriptome

Introduction

The analysis of the transcriptome is usually based on reads matching exons and gene-based annotations. A large proportion of the reads from RNA-sequencing map to the intronic sequences and this is valid for both ribosomal depleted and poly-A selected RNA protocols [1]. It has been described that 38% of total RNA seq reads and only 8% of polyA RNA reads map to the introns [1]. The significance of the intronic reads has remained controversial and this part of the transcriptome is mostly neglected in transcriptomic analyses [2]. However, recent studies have shown that the reads mapping to introns are genuine and reflect the immediate regulatory responses in transcription compared to post-transcriptional or steady-state changes [2, 3]. The intronic reads reflect the presence of the newly transcribed RNA and as such are useful to explore the complexity of the nascent RNA transcription and co-transcriptional splicing [4-6]. The analysis of intronic reads has been used to develop a detailed transcriptional model within a single sample showing the utility of intronic reads to estimate the genome-wide pre-mRNA synthesis rate [7]. It's also been reported that intronic coverage was related to nascent transcription and co-transcriptional splicing [1]. The levels of intronic reads precede the change in exonic reads by 15 minutes, making them very useful to detect immediate responsive changes in the transcriptome [5]. Therefore, analysis of the intronic reads allows the separation of nascent transcription from post-transcriptional changes during the formation of the genome-wide transcriptome.

Parkinson's Disease (PD) is one of the commonest neurodegenerative diseases with several clearly identified genetic mutations and variations associated with the disease [8-11]. The exact mechanisms of the disease are still not clear and given the complexity of the neuropathology of PD, the genomic network leading to the pathology are similarly likely to be both complex and multifactorial [12-14]. Several studies have attempted transcriptomic

analysis correlating with PD and identified transcriptional signatures specific to the disease or involved in the regulation of the splicing of genes expressed in the basal ganglia [15, 16]. The analysis of peripheral tissue transcriptomes (blood and skin) overlaid on CNS PD data has identified significant differences between PD and control transcriptomic profiles [12, 17, 18]. These studies underlined the importance of peripheral tissue analysis to not only determine biomarkers correlating with the disease but also gain insight into neurodegenerative diseases. In this longitudinal study we compared the blood intronic transcriptome of PD patients and healthy controls (Table 1) at two different timepoints of disease progression, at diagnosis (baseline, BL) and after three years follow up (V08). To address the immediate changes in the transcriptome reflected by the changes in nascent RNA we used only reads mapping to the intronic sequences. Intronic data were compared with the exonic reads to differentiate steady-state transcription from nascent transcription. We used blood whole transcriptome data from the Parkinson's Progression Markers Initiative (PPMI) cohort that contains data of PD and control (CO) subjects at these two different timepoints in the trajectory of PD.

Materials and Methods

Datasets

In this study we utilized the Parkinson's Progression Markers Initiative (PPMI) cohort data that were downloaded from www.ppmi-info.org/data (16 May 2021). The PPMI is a longitudinal cohort of Parkinson's patients with the aims to describe the progression and biomarkers of Parkinson's disease. The dataset contains whole transcriptome data from the blood together with genetic and clinical data. For the RNA-seq 1µg of RNA isolated from PaxGene tubes was used and sequencing was performed at Hudson's Alpha's Genomic Services Lab on an Illumina NovaSeq6000. All samples under went rRNA and globin reduction, followed by directional cDNA synthesis using the NEB kit. Following second-strand synthesis, library samples were prepared using the NEB/Kapa (NEBKAP) based library kit. Fastq files were merged and aligned to GRCh38p12 by STAR (v2.6.1d) on GENCODE v29.

Analytical Workflow

Bam files were imported to the R environment?ref/irl and intronic reads were called using packages "GenomicFeatures" and "GenomicAlignments". Differential expression was detected using "DESeq2" and functionally annotated using "ReactomePA", "DOSE" and "clusterProfiler" packages.

Statistical analysis

Formal statistical analysis of differential expression of the intronic transcripts was performed by using the "DESeq2" packages for R and only False Discovery Rate (FDR) corrected transcripts below 0.05 are reported. In addition, selection of FDR corrected transcripts was used for the pairwise comparison of the expression data and the plots were generated using "ggpubr" package.

Results

We analysed the intronic expression profiles in PD and CO subjects at the time of diagnosis (BL visit) and three years later (V08). There was little differential expression in exons in all our comparisons in both PD and CO groups, but there were many changes observed in intronic variation in the PD cohort which were highly significant and signify widespread active transcription in the blood from BL to V08. In addition, only few active transcription changes in the CO subjects were observed if we compared intronic transcripts at BL and V08 timepoints (Table 2). Remarkably, the differences in transcripts were identified almost exclusively only in introns and very few in exons (Table 2, Figure 1). More detailed description of our findings follows.

CO subjects

We analysed the longitudinal changes in the expression of the intronic reads of control subjects at BL and V08 timepoints and identified only limited differences. Nine transcripts showed differential intronic expression below FDR 0.05 (Supplementary Table 1). We also analysed the differential expression of the exonic transcripts, and nine transcripts were differentially expressed with exonic reads (Supplementary Table 2). This shows that on both intronic and exonic levels longitudinal differential expression in control subjects is quite limited and we can conclude that the transcriptome for control subjects is longitudinally very stable.

PD subjects

Next, we analysed only PD patients and compared their longitudinal intronic and exonic transcriptional profile. We identified 4,873 differentially expressed intronic reads in PD patients that differed between the time of diagnosis and after three years follow up (Supplementary Table 3). These 4,873 introns reflect the longitudinal change that is specific for PD patients during the three-year period. This is in stark contrast to the CO cohort that

exhibited a very stable transcriptome with very limited changes. At the same time, we detected only 8 exonic reads differentially expressed, showing the specificity of the intronic changes (Supplementary Table 4). Differentially expressed intronic reads were from several genes known to be involved in neurodegenerative diseases and PD, such as *LRRK2, VPS13C,*
130 *LGALS3, C9orf72* and *ALS2* (Figure 2). We identified statistically significant upregulation of the intron 1 and intron 2 expression of the *PINK1* gene, showing that the intronic changes are not limited to single introns (Figure 3). *FLACC1, KANLS1AS1* and *CASP8AP2* were other genes in which differential intronic expression occurred (Figure 3). Taken together, we can conclude here that PD patients express overwhelming longitudinal changes in nascent transcription.

135 Baseline differences

We analysed PD versus CO at different timepoints to have two separate cross-sectional snapshots of transcriptional changes. At the time of diagnosis, PD patients had 836 introns differentially expressed compared to the controls (Supplementary Table 5). Only one exon was differentially expressed showing again that the enhanced intronic expression is a very
140 specific feature of PD pathogenesis (Supplementary Table 6). Most remarkably, we identified *SNCA, GOLGA5* and *DNAJC19* genes with differentially expressed introns (Figure 4). *GOLGA5* is involved in the Golgi membrane and synaptic vesicle trafficking and docking [19]. *DNAJC19* or *TIMM14* is a chaperone that functions as an inner mitochondrial membrane co-chaperone and is responsible for mitochondrial protein import, a process that is involved in the
145 pathogenesis of PD among others [20-22]. The gene *WDFY3* showed eight introns to be differentially upregulated in PD patients (Figure 5). This gene is responsible for mitochondrial quality control and is associated with multiple severe neurological pathologies as it is involved in brain energetics and mitophagy [23]. We also identified genes with differential intronic upregulation related to the immune response or to the age at onset of PD, such as *PLXNC1*,

150 *TNFAIP6* and *PRRG4* (Figure 6). These changes were found at the time of diagnosis and indicate already existing active transcriptional changes from the moment of clinical presentation of PD.

Differences at the three-year follow up visit (V08)

While it is important to have a cross sectional snapshot at the time of diagnosis, longitudinal
155 study design allow exploration of time-dependent alterations in the transcriptome. Three years after diagnosis, PD patients had 2,184 intronic transcripts differentially expressed (Supplementary Table 7). At the same time only 17 differentially-expressed transcripts were from exon sequences indicating again very high specificity toward intronic transcription (Supplementary Table 8). From all the detected 2,184 differential transcripts, 329 were
160 identical to the transcripts at the BL timepoint demonstrating major changes in nascent transcriptional changes in the PD group. Also, one transcript, *RP11-403I13.4-002*, was identical in the exonic and intronic analysis of V08 timepoint in the PD group. Figure 7 gives some details about the genes that have been identified to be involved in neurodegeneration. We identified longitudinal changes in the expression of *GBA-204* and *GBA-206* isoforms,
165 *LRRK2* and *PINK1*; genes that are well known to be involved in the pathogenesis of PD (Figure 7).

Taken together, differential transcriptome analysis of introns and exons indicated specific and overwhelming changes in intronic transcription in PD patients compared to CO subjects, these differences were evident at the time of diagnosis and escalated during the
170 three-year progression of the disease.

Pathway analysis of longitudinal changes

We performed functional pathway analysis to identify the enriched pathways linked to the activated intronic transcription and to identify the common theme of these activated transcripts. As longitudinal PD changes were the most remarkable, we used only PD data for

175 pathway analysis and compared the intronic transcriptome profiles between the visits V08 and BL.

Although the analysis was performed in blood RNAseq, reactome enrichment analysis identified that nervous system related themes were enriched in the differential profile of the intronic transcription (Figure 8, Table 3). The pathways we identified were neuronal system,
180 protein-protein interactions at synapses, transmission across chemical synapses and muscle contraction related pathways to name some (Table 2). Remarkably, activation of the neuronal pathways was clear and is illustrated on Figure 8 as a bar plot and a dot plot showing the number of genes identified and gene ratio. To illustrate this enrichment further, we used heat and tree plots (Figure 9) showing again the large number of genes mapping to the pathways.
185 This indicated that the longitudinal changes in active transcription changes in the blood maybe reflected changes in the nervous system and most likely reflected longitudinal changes related to PD progression. This statement was further supported by the longitudinal comparison of PD and CO groups separately. As very limited transcriptional alterations in the longitudinal CO subgroup compared to PD subgroup were found, we can quite confidently
190 conclude that these pathways and biological functions are directly related to the progression of PD.

KEGG enrichment analysis was performed to validate our previous pathway analysis and to identify potential PD specific pathways (Figures 10 and 11). Indeed, when using KEGG annotation, the most significantly changed pathways were ubiquitin mediated proteolysis,
195 protein processing in endoplasmic reticulum, mitophagy and autophagy. The activation of protein processing in the endoplasmic reticulum (Figure 10) and the activation of mitophagy pathway (Figure 11) during the time course of PD supports the causal relationship between

the transcripts and disease progression. All these pathways are directly involved in the pathogenesis of PD.

While we analyzed the blood whole transcriptome data, using the intronic reads helped to identify genes with active transcriptional change reflecting the dynamic alterations in the transcriptional balance of the cells. By using blood data on clinical samples, we were able to identify PD specific pathogenetic networks that have been identified to be involved in PD using cellular models and biochemical experiments. These results show that blood transcriptome can reflect nascent transcription specific for the disease condition.

Taken together, analysis of the intronic reads in the whole genome transcriptome study allows to detect active nascent transcription and gives much more detailed information compared to the exon centric steady-state analysis of transcription. We were able to show large scale intronic activation that helped to detect actively transcribed transcripts that we were able to map into functionally and clinically relevant genetic networks. Intronic transcriptome analysis provides information about the condition-specific transcripts that can help to identify genomic patterns specific for the condition being studied.

Discussion

Whole transcriptome analysis is typically based on gene- or exon-based annotation, quite seldom transcript-based annotation is used [24, 25]. Gene-based annotation is an amalgamated approach where the reads mapping to different exons and transcripts will be merged under a single functional identifier, the gene [26]. This approach is the most widely used and therefore most of the whole transcriptome studies provide this aggregated information. Inevitably, this approach leads to the loss of power to detect precise and detailed changes in transcription leading to the loss of the sensitivity of the tool. The avoidance of transcript-based annotation is understandable as the bioinformatics tools to call transcripts accurately from short read data are limited. Only recently Kallisto and Salmon were developed to provide quasi-mapping approach and accurate transcript calling to overcome the issue with transcript detection [27, 28]. Nevertheless, using these tools nascent cannot be distinguished from steady state transcription.

Recently intronic mapping was suggested to be an alternative to identify the transcripts that were recently transcribed [2]. This approach, based on the co-transcriptional splicing and detection of introns, indicates that the transcripts still have introns included. Detecting this event in the transcriptional process gives important additional power to measure the genes that are actively and newly transcribed compared to the steady state stable transcription in the background. The analysis of intronic transcripts has successfully been applied to human basal ganglia data with clear evidence for the reproducibility of the intronic eQTLs (i-eQTLs) and their utility to analyze the rate of transcription [15]. Interestingly, in this paper the authors also identified highly specific enrichment of disease-specific transcription suggesting the suitability of the intron based transcriptional analysis. The

authors also confirmed the mapping of intronic signals to novel transcripts and validated a/the annotation-independent approach in transcriptomic analysis.

We have identified the longitudinal changes in intronic transcription in PD patients. The PD changes could be specific for the disease condition as only minor differences were found in the analysis of the controls. Indeed, functional annotation of the intronic transcripts revealed activation of pathways closely related to the pathophysiology of PD although only blood RNAseq was analyzed. The functional significance of that data awaits further analysis using patient derived cell lines, but the identified differences are intriguing as in many cases they reflected changes in the pathways related to neurodegenerative processes.

Out of many different transcripts identified in the present study, the ones with clear evidence for the mitochondrial function and proteostasis support the relevance of our findings in the context of the pathophysiology and progression of PD. The gene DNAJC19 or TIMM14 encodes for an inner mitochondrial membrane translocase that imports proteins to mitochondria and is directly involved in the pathogenesis of neurodegenerative diseases [20-22]. Products of TIMM genes interact with translocases of the outer mitochondrial membrane (TOMM) to form a transport pathway for the nuclearly encoded precursor proteins [20]. In our earlier study we identified TOMM20 to be involved in the neurodegeneration caused by the downregulation of WFS1 gene leading to mitochondrial damage [29]. In addition, in the present study we identified eight introns of the WDFY3 gene to be differentially upregulated in PD patients. WDFY3 functions as a conductor for aggregate clearance by autophagy and colocalises with the aggregated proteins [30]. WDFY3 is associated with multiple severe neurological pathologies by regulating brain bioenergetics, autophagy and mitophagy [23, 31-33]. The findings from single transcripts and intronic changes were further confirmed by functional annotation of differential expression and identification of the activation of

autophagy and mitophagy in PD patients. Changes in intronic transcription were evident at the time of diagnosis indicating concomitant transcriptional changes with the clinical presentation of the PD.

Comparison of the intronic and exonic signals has recently been systemically analyzed and the reflection of the nascent transcription by the intronic reads confirmed [2]. Several experimental studies indicate that intronic transcription is a reliable proxy to measure nascent transcription [7]. Moreover, comparing exonic reads to intronic signals helps to differentiate transcriptional changes from post-transcriptional changes. Therefore, the differences we have identified are caused by changes in active transcription and are not due to changes related to the normal physiological processes.

Conclusions

In conclusion, we identified highly specific longitudinal nascent transcriptional profile in the blood of Parkinson patients that possibly reflects the changes caused by the molecular pathological processes of the disease and are relevant to improve our understanding about the progression of the disease.

Supplementary Materials

Table S1: SupplementaryTable1.xls, Table S2: SupplementaryTable2.xls,

Table S3: SupplementaryTable3.xls, Table S4: SupplementaryTable4.xls,

Table S5: SupplementaryTable5.xls, Table S6: SupplementaryTable6.xls,

Table S7: SupplementaryTable7.xls, Table S8: SupplementaryTable8.xls.

Author Contributions

Conceptualization, S.K.; methodology, A.L.P. and S.K.; formal analysis, S.K.; data interpretation, A.L.P., V.J.B., J.P.Q. and S.K.; writing—original draft preparation, S.K.; writing—review and editing, A.L.P., V.J.B., J.P.Q. and S.K.; funding acquisition, A.L.P. and S.K. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by MSWA, The Michael J. Fox Foundation, Shake It Up Australia and Perron Institute for Neurological and Translational Science.

Institutional Review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Human Ethics Research Office of The University of Western Australia (protocol code RA/4/20/5308 approved on 05.08.2019)

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study by PPMI.

Data Availability Statement

Raw data are available from the PPMI website (www.ppmi-info.org/data (accessed on 19 January 2021)).

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 (accessed on 19 January 2021)).

For up-to-date information on the study, visit www.ppmi-info.org. PPMI is sponsored and partially funded by The Michael J. Fox Foundation for Parkinson's Research. 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.

Conflicts of Interest

The authors declare no conflict of interest related to this study. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Table 1. Overview of the study samples and design. Subject numbers with the blood whole transcriptome data of the PPMI cohort are given.

Study group	Baseline (BL)	Three years (V08)
CO	189	157
PD	390	338

Table 2. Comparison of the differential expression of the intronic and exonic reads. For baseline (BL) and three-year (V08) time points PD was compared to CO. For PD and CO groups, the effect of time (three years) was measured (V08 versus BL), significance threshold was FDR < 0.05.

Group	Introns	Exons	Comparison
BL	836	1	PD versus CO
V08	2184	17	PD versus CO
PD	4873	8	V08 versus BL
CO	9	9	V08 versus BL

Table 3. Reactome pathway over-representation analysis. Longitudinal changes of the intronic transcription in PD patients were associated with the biological themes related to the nervous system.

Description	Gene Ratio	P-value	P-adjusted
Extracellular matrix organization	80/1505	2.6×10^{-9}	3.1×10^{-6}
Neuronal System	100/1505	4.3×10^{-9}	3.1×10^{-6}
Protein-protein interactions at synapses	30/1505	6.7×10^{-7}	0.0003
Muscle contraction	50/1505	1.0×10^{-5}	0.003
Striated Muscle Contraction	14/1505	0.0001	0.02
Transmission across Chemical Synapses	59/1505	0.0002	0.02
Neurexins and neuroligins	18/1505	0.0003	0.02

335

Legends to the figures

340 **Figure 1.** The figure illustrates the number of exonic and intronic reads in different subgroups and at different timepoints. The number of intronic reads is very high in PD patients compared to CO subjects. Exonic reads didn't change by the disease status or by the progression of the disease. It is also evident that the number of intronic reads increases from the baseline visit to three-year follow up.

345 **Figure 2.** Longitudinal differential expression of the intronic transcripts in PD patients, baseline (BL) compared to the V08 visit three years later. Panels A, B, C, D, E, F reflect changes in the introns for different transcripts. LRRK2, ALS2 and VPS13C are downregulated during the progression of PD. C9orf72, LGALS3 and LGALSL are upregulated within three years of PD.

Figure 3. Longitudinal differential expression of the intronic transcripts in PD patients, BL

350 compared to the V08 visit. The panels A, B, E, F, G, H reflect changes in the introns for different transcripts. Panels C and D illustrate changes for PINK1-201 transcripts, introns 1 and 2.

Figure 4. Cross-sectional analysis of the intronic transcripts at the baseline visit (BL), at the time of enrolment of the subjects. PD patients had decreased expression of the SNCA-216, DNAJC19-211 and MRPS10-201 transcripts. The transcripts GOLGA5-201, CPQ-201 and UBR5-

355 208 were upregulated.

Figure 5. Cross-sectional analysis of the intronic transcripts at the baseline visit (BL), at the time of enrolment of the subjects. PD patients had increase expression of 8 introns of the WDFY3 gene.

Figure 6. Cross-sectional analysis of the intronic transcripts at the baseline visit (BL), at the

360 time of enrolment of the subjects. Different introns of the transcripts related to the immune response were up regulated in PD patients.

Figure 7. Cross-sectional analysis of the intronic transcripts at the V08 visit, three years after enrolment of the subjects. Two different transcripts of GBA, LRRK2-203, ARHGAP27-204 and PLEKHM1-212 were upregulated in PD patients. Intorns1 and 2 of the PINK1-201 transcript

365 were also upregulated. ALS2-251 transcript was downregulated in PD patients.

Figure 8. Reactome pathway analysis indicates longitudinal enrichment of the neuronal themes in the blood transcriptome of the PD patients. (**a**) Bar plot of all statistically significant

enriched pathways indicates the different number of genes for each pathway. (**b**) Dot plot of the enriched pathways shows the relation between the gene count, gene ratio and adjusted
370 p-values.

Figure 9. Reactome pathway analysis identified several neuronal pathways enriched longitudinally in PD blood transcriptome. (**a**) Heat plot showing individual genes and their mapping to pathways; (**b**) Tree plot illustrates clustering of the pathways, number of matched genes and adjusted p-values.

375 **Figure 10.** KEGG pathway of the protein processing in the endoplasmic reticulum and the mapped genes from the intronic transcriptome. Mapped genes are in red, green, or grey colors.

Figure 11. KEGG mitophagy pathway. Longitudinal blood intronic transcriptome profile of Parkinson disease mapped significantly to the mitophagy pathway. Mapped genes are in red,
380 green, or grey colours.

References

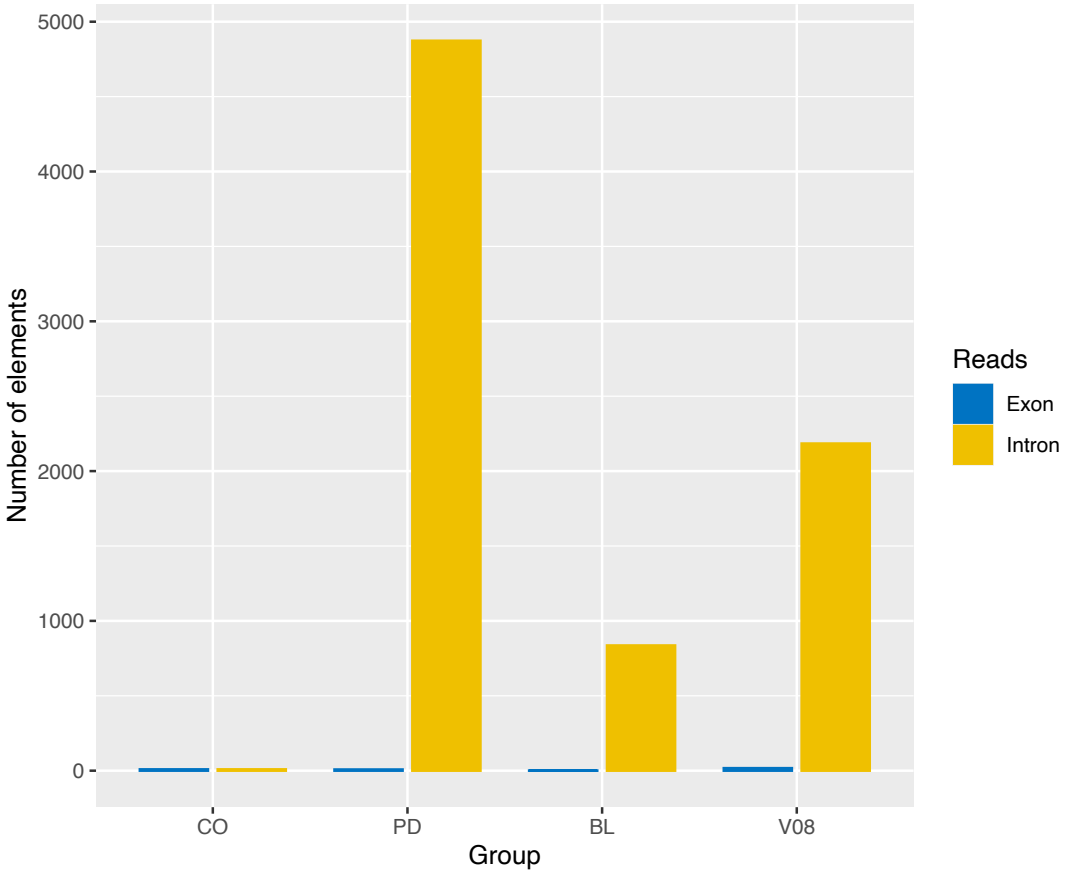
1. Ameur, A.; Zaghlool, A.; Halvardson, J.; Wetterbom, A.; Gyllensten, U.; Cavelier, L.; Feuk, L., Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* **2011,** 18, (12), 1435-40.
2. Gaidatzis, D.; Burger, L.; Florescu, M.; Stadler, M. B., Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat Biotechnol* **2015,** 33, (7), 722-9.
3. Wissink, E. M.; Vihervaara, A.; Tippens, N. D.; Lis, J. T., Nascent RNA analyses: tracking transcription and its regulation. *Nat Rev Genet* **2019,** 20, (12), 705-723.
4. Core, L. J.; Waterfall, J. J.; Lis, J. T., Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **2008,** 322, (5909), 1845-8.
5. Hendriks, G. J.; Gaidatzis, D.; Aeschimann, F.; Großhans, H., Extensive oscillatory gene expression during C. elegans larval development. *Mol Cell* **2014,** 53, (3), 380-92.
6. Khodor, Y. L.; Rodriguez, J.; Abruzzi, K. C.; Tang, C. H.; Marr, M. T., 2nd; Rosbash, M., Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev* **2011,** 25, (23), 2502-12.
7. Gray, J. M.; Harmin, D. A.; Boswell, S. A.; Cloonan, N.; Mullen, T. E.; Ling, J. J.; Miller, N.; Kuersten, S.; Ma, Y. C.; McCarroll, S. A.; Grimmond, S. M.; Springer, M., SnapShot-Seq: a method for extracting genome-wide, in vivo mRNA dynamics from a single total RNA sample. *PLoS One* **2014,** 9, (2), e89673.
8. Iwaki, H.; Blauwendraat, C.; Leonard, H. L.; Kim, J. J.; Liu, G.; Maple-Grødem, J.; Corvol, J. C.; Pihlstrøm, L.; van Nimwegen, M.; Hutten, S. J.; Nguyen, K. H.; Rick, J.; Eberly, S.; Faghri, F.; Auinger, P.; Scott, K. M.; Wijeyekoon, R.; Van Deerlin, V. M.; Hernandez, D. G.; Gibbs, J. R.; Chitrala, K. N.; Day-Williams, A. G.; Brice, A.; Alves, G.; Noyce, A. J.; Tysnes, O. B.; Evans, J. R.; Breen, D. P.; Estrada, K.; Wegel, C. E.; Danjou, F.; Simon, D. K.; Andreassen, O.; Ravina, B.; Toft, M.; Heutink, P.; Bloem, B. R.; Weintraub, D.; Barker, R. A.; Williams-Gray, C. H.; van de Warrenburg, B. P.; Van Hilten, J. J.; Scherzer, C. R.; Singleton, A. B.; Nalls, M. A., Genomewide association study of Parkinson's disease clinical biomarkers in 12 longitudinal patients' cohorts. *Mov Disord* **2019,** 34, (12), 1839-1850.
9. Bandres-Ciga, S.; Ahmed, S.; Sabir, M. S.; Blauwendraat, C.; Adarmes-Gómez, A. D.; Bernal-Bernal, I.; Bonilla-Toribio, M.; Buiza-Rueda, D.; Carrillo, F.; Carrión-Claro, M.; Gómez-Garre, P.; Jesús, S.; Labrador-Espinosa, M. A.; Macias, D.; Méndez-Del-Barrio, C.; Periñán-Tocino, T.; Tejera-Parrado, C.; Vargas-González, L.; Diez-Fairen, M.; Alvarez, I.; Tartari, J. P.; Buongiorno, M.; Aguilar, M.; Gorostidi, A.; Bergareche, J. A.; Mondragon, E.; Vinagre-Aragon, A.; Croitoru, I.; Ruiz-Martínez, J.; Dols-Icardo, O.; Kulisevsky, J.; Marín-Lahoz, J.; Pagonabarraga, J.; Pascual-Sedano, B.; Ezquerra, M.; Cámara, A.; Compta, Y.; Fernández, M.; Fernández-Santiago, R.; Muñoz, E.; Tolosa, E.; Valldeoriola, F.; Gonzalez-Aramburu, I.; Sanchez Rodriguez, A.; Sierra, M.; Menéndez-González, M.; Blazquez, M.; Garcia, C.; Suarez-San Martin, E.; García-Ruiz, P.; Martínez-Castrillo, J. C.; Vela-Desojo, L.; Ruz, C.; Barrero, F. J.; Escamilla-Sevilla, F.; Mínguez-Castellanos, A.; Cerdan, D.; Tabernero, C.; Gomez Heredia, M. J.; Perez Errazquin, F.; Romero-Acebal, M.; Feliz, C.; Lopez-Sendon, J. L.; Mata, M.; Martínez Torres, I.; Kim, J. J.; Dalgard, C. L.; Brooks, J.; Saez-Atienzar, S.; Gibbs, J. R.; Jorda, R.;

Botia, J. A.; Bonet-Ponce, L.; Morrison, K. E.; Clarke, C.; Tan, M.; Morris, H.; Edsall, C.; Hernandez, D.; Simon-Sanchez, J.; Nalls, M. A.; Scholz, S. W.; Jimenez-Escrig, A.; Duarte, J.; Vives, F.; Duran, R.; Hoenicka, J.; Alvarez, V.; Infante, J.; Marti, M. J.; Clarimón, J.; López de Munain, A.; Pastor, P.; Mir, P.; Singleton, A., The Genetic Architecture of Parkinson Disease in Spain: Characterizing Population-Specific Risk, Differential Haplotype Structures, and Providing Etiologic Insight. *Mov Disord* **2019,** 34, (12), 1851-1863.

10. Nalls, M. A.; Blauwendraat, C.; Vallerga, C. L.; Heilbron, K.; Bandres-Ciga, S.; Chang, D.; Tan, M.; Kia, D. A.; Noyce, A. J.; Xue, A.; Bras, J.; Young, E.; von Coelln, R.; Simón-Sánchez, J.; Schulte, C.; Sharma, M.; Krohn, L.; Pihlstrøm, L.; Siitonen, A.; Iwaki, H.; Leonard, H.; Faghri, F.; Gibbs, J. R.; Hernandez, D. G.; Scholz, S. W.; Botia, J. A.; Martinez, M.; Corvol, J. C.; Lesage, S.; Jankovic, J.; Shulman, L. M.; Sutherland, M.; Tienari, P.; Majamaa, K.; Toft, M.; Andreassen, O. A.; Bangale, T.; Brice, A.; Yang, J.; Gan-Or, Z.; Gasser, T.; Heutink, P.; Shulman, J. M.; Wood, N. W.; Hinds, D. A.; Hardy, J. A.; Morris, H. R.; Gratten, J.; Visscher, P. M.; Graham, R. R.; Singleton, A. B., 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), 1091-1102.
11. Billingsley, K. J.; Barbosa, I. A.; Bandres-Ciga, S.; Quinn, J. P.; Bubb, V. J.; Deshpande, C.; Botia, J. A.; Reynolds, R. H.; Zhang, D.; Simpson, M. A.; Blauwendraat, C.; Gan-Or, Z.; Gibbs, J. R.; Nalls, M. A.; Singleton, A.; International Parkinson's Disease Genomics, C.; Ryten, M.; Koks, S., Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset. *NPJ Parkinsons Dis* **2019,** 5, 8.
12. Kurvits, L.; Lattekivi, F.; Reimann, E.; Kadastik-Eerme, L.; Kasterpalu, K. M.; Koks, S.; Taba, P.; Planken, A., Transcriptomic profiles in Parkinson's disease. *Exp Biol Med (Maywood)* **2021,** 246, (5), 584-595.
13. Müller-Nedebock, A. C.; van der Westhuizen, F. H.; Kõks, S.; Bardien, S., Nuclear Genes Associated with Mitochondrial DNA Processes as Contributors to Parkinson's Disease Risk. *Mov Disord* **2021,** 36, (4), 815-831.
14. Pfaff, A. L.; Bubb, V. J.; Quinn, J. P.; Koks, S., An Increased Burden of Highly Active Retrotransposition Competent L1s Is Associated with Parkinson's Disease Risk and Progression in the PPMI Cohort. *Int J Mol Sci* **2020,** 21, (18).
15. Guelfi, S.; D'Sa, K.; Botía, J. A.; Vandrovcova, J.; Reynolds, R. H.; Zhang, D.; Trabzuni, D.; Collado-Torres, L.; Thomason, A.; Quijada Leyton, P.; Gagliano Taliun, S. A.; Nalls, M. A.; Small, K. S.; Smith, C.; Ramasamy, A.; Hardy, J.; Weale, M. E.; Ryten, M., Regulatory sites for splicing in human basal ganglia are enriched for disease-relevant information. *Nat Commun* **2020,** 11, (1), 1041.
16. Reynolds, R. H.; Botía, J.; Nalls, M. A.; Hardy, J.; Gagliano Taliun, S. A.; Ryten, M., Moving beyond neurons: the role of cell type-specific gene regulation in Parkinson's disease heritability. *NPJ Parkinsons Dis* **2019,** 5, 6.
17. Kurvits, L.; Reimann, E.; Kadastik-Eerme, L.; Truu, L.; Kingo, K.; Erm, T.; Kõks, S.; Taba, P.; Planken, A., Serum Amyloid Alpha Is Downregulated in Peripheral Tissues of Parkinson's Disease Patients. *Front Neurosci* **2019,** 13, 13.
18. Planken, A.; Kurvits, L.; Reimann, E.; Kadastik-Eerme, L.; Kingo, K.; Kõks, S.; Taba, P., Looking beyond the brain to improve the pathogenic understanding of Parkinson's disease: implications of whole transcriptome profiling of Patients' skin. *BMC Neurol* **2017,** 17, (1), 6.

19. Malsam, J.; Satoh, A.; Pelletier, L.; Warren, G., Golgin tethers define subpopulations of COPI vesicles. *Science* **2005,** 307, (5712), 1095-8.
20. Franco-Iborra, S.; Cuadros, T.; Parent, A.; Romero-Gimenez, J.; Vila, M.; Perier, C., Defective mitochondrial protein import contributes to complex I-induced mitochondrial dysfunction and neurodegeneration in Parkinson's disease. *Cell Death Dis* **2018,** 9, (11), 1122.
21. Davey, K. M.; Parboosingh, J. S.; McLeod, D. R.; Chan, A.; Casey, R.; Ferreira, P.; Snyder, F. F.; Bridge, P. J.; Bernier, F. P., Mutation of DNAJC19, a human homologue of yeast inner mitochondrial membrane co-chaperones, causes DCMA syndrome, a novel autosomal recessive Barth syndrome-like condition. *J Med Genet* **2006,** 43, (5), 385-93.
22. Heinemeyer, T.; Stemmet, M.; Bardien, S.; Neethling, A., Underappreciated Roles of the Translocase of the Outer and Inner Mitochondrial Membrane Protein Complexes in Human Disease. *DNA Cell Biol* **2019,** 38, (1), 23-40.
23. Napoli, E.; Song, G.; Panoutsopoulos, A.; Riyadh, M. A.; Kaushik, G.; Halmai, J.; Levenson, R.; Zarbalis, K. S.; Giulivi, C., Beyond autophagy: a novel role for autism-linked Wdfy3 in brain mitophagy. *Sci Rep* **2018,** 8, (1), 11348.
24. Yao, S.; Zhang, X.; Zou, S. C.; Zhu, Y.; Li, B.; Kuang, W. P.; Guo, Y.; Li, X. S.; Li, L.; Wang, X. Y., A transcriptome-wide association study identifies susceptibility genes for Parkinson's disease. *NPJ Parkinsons Dis* **2021,** 7, (1), 79.
25. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szcześniak, M. W.; Gaffney, D. J.; Elo, L. L.; Zhang, X.; Mortazavi, A., A survey of best practices for RNA-seq data analysis. *Genome Biol* **2016,** 17, 13.
26. Falchetti, M.; Prediger, R. D.; Zanotto-Filho, A., Classification algorithms applied to blood-based transcriptome meta-analysis to predict idiopathic Parkinson's disease. *Comput Biol Med* **2020,** 124, 103925.
27. Bray, N. L.; Pimentel, H.; Melsted, P.; Pachter, L., Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **2016,** 34, (5), 525-7.
28. Patro, R.; Duggal, G.; Love, M. I.; Irizarry, R. A.; Kingsford, C., Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **2017,** 14, (4), 417-419.
29. Kõks, S.; Overall, R. W.; Ivask, M.; Soomets, U.; Guha, M.; Vasar, E.; Fernandes, C.; Schalkwyk, L. C., Silencing of the WFS1 gene in HEK cells induces pathways related to neurodegeneration and mitochondrial damage. *Physiol Genomics* **2013,** 45, (5), 182-90.
30. Filimonenko, M.; Isakson, P.; Finley, K. D.; Anderson, M.; Jeong, H.; Melia, T. J.; Bartlett, B. J.; Myers, K. M.; Birkeland, H. C.; Lamark, T.; Krainc, D.; Brech, A.; Stenmark, H.; Simonsen, A.; Yamamoto, A., The selective macroautophagic degradation of aggregated proteins requires the PI3P-binding protein Alfy. *Mol Cell* **2010,** 38, (2), 265-79.
31. Fox, L. M.; Kim, K.; Johnson, C. W.; Chen, S.; Croce, K. R.; Victor, M. B.; Eenjes, E.; Bosco, J. R.; Randolph, L. K.; Dragatsis, I.; Dragich, J. M.; Yoo, A. S.; Yamamoto, A., Huntington's Disease Pathogenesis Is Modified In Vivo by Alfy/Wdfy3 and Selective Macroautophagy. *Neuron* **2020,** 105, (5), 813-821.e6.
32. Yamamoto, A.; Simonsen, A., Alfy-dependent elimination of aggregated proteins by macroautophagy: can there be too much of a good thing? *Autophagy* **2011,** 7, (3), 346-50.

33. Reinhart, E. F.; Litt, N. A.; Katzenell, S.; Pellegrini, M.; Yamamoto, A.; Ragusa, M. J., A
525 highly conserved glutamic acid in ALFY inhibits membrane binding to aid in aggregate clearance. *Traffic* **2021,** 22, (1-2), 23-37.



530

Figure 1. The figure illustrates the number of exonic and intronic reads in different subgroups and at different timepoints

535

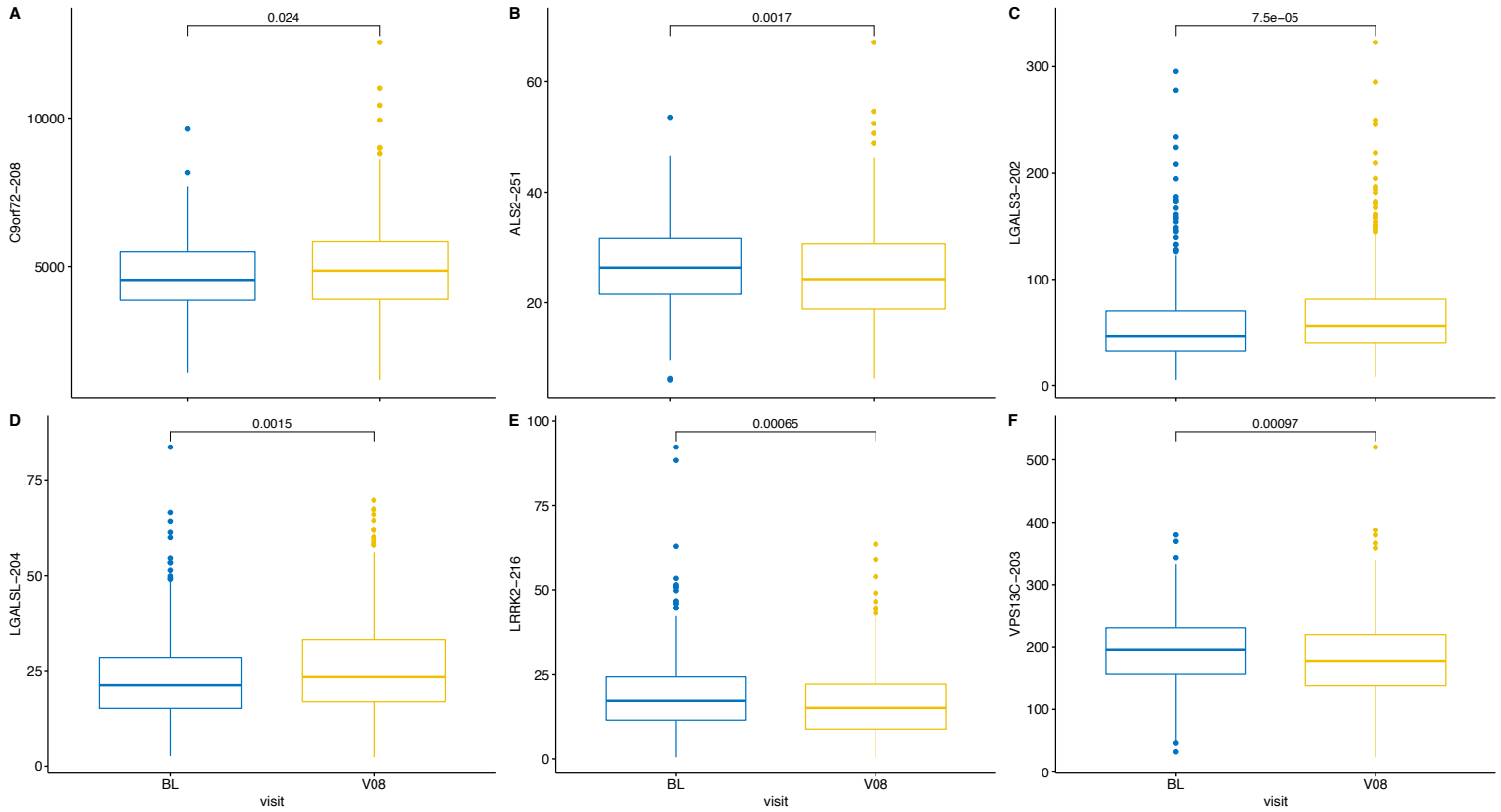
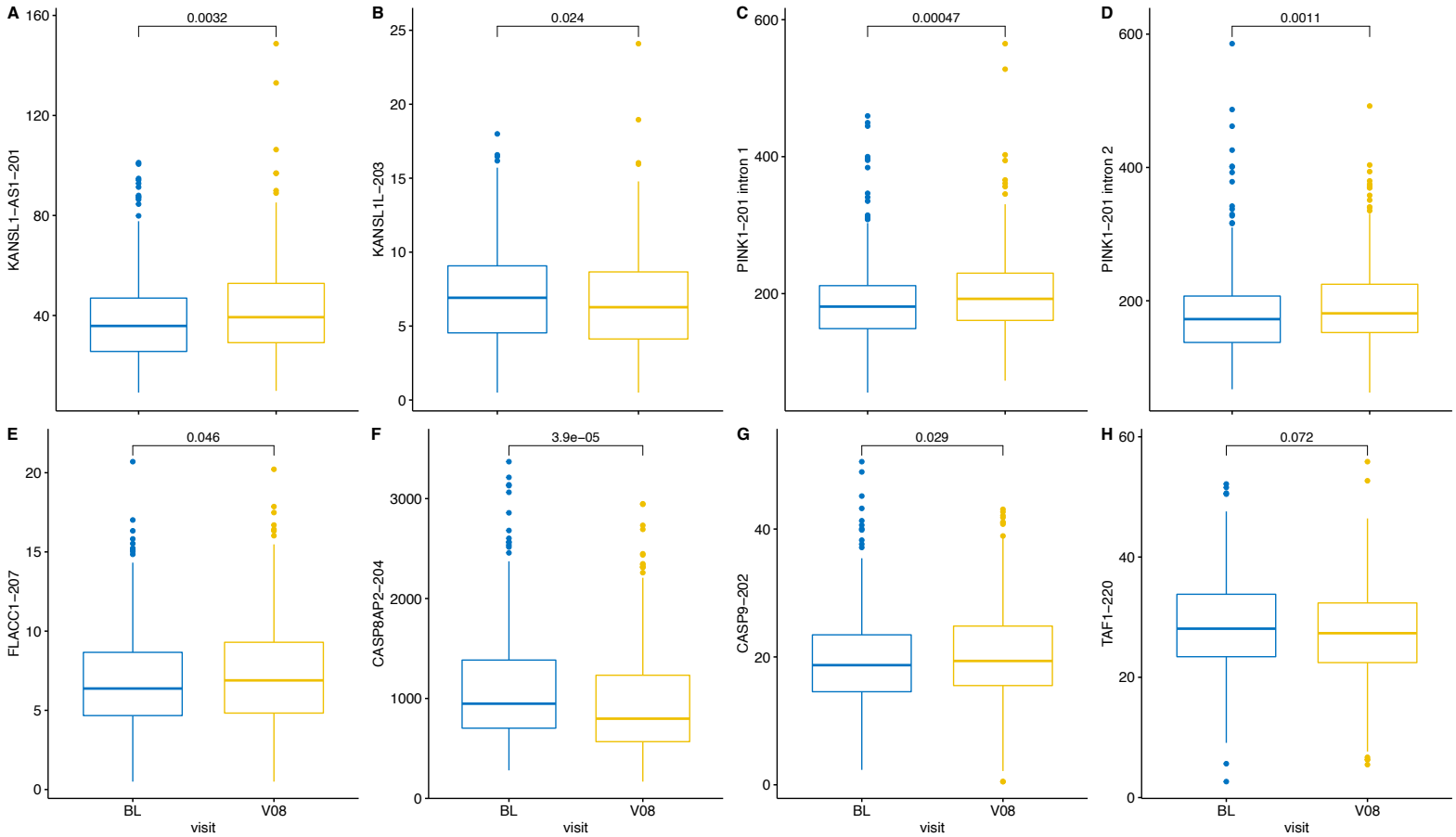


Figure 2. Longitudinal differential expression of the intronic transcripts in PD patients, baseline (BL) compared to the V08 visit three years later.
540 Panels A, B, C, D, E, F reflect changes in the introns for different transcripts. LRRK2, ALS2 and VPS13C are downregulated during the progression of PD. C9orf72, LGALS3 and LGALSL are upregulated within three years of PD.



545

Figure 3. Longitudinal differential expression of the intronic transcripts in PD patients, BL compared to the V08 visit. The panels A, B, E, F, G, H reflect changes in the introns for different transcripts. Panels C and D illustrate changes for PINK1-201 transcripts, introns 1 and 2.

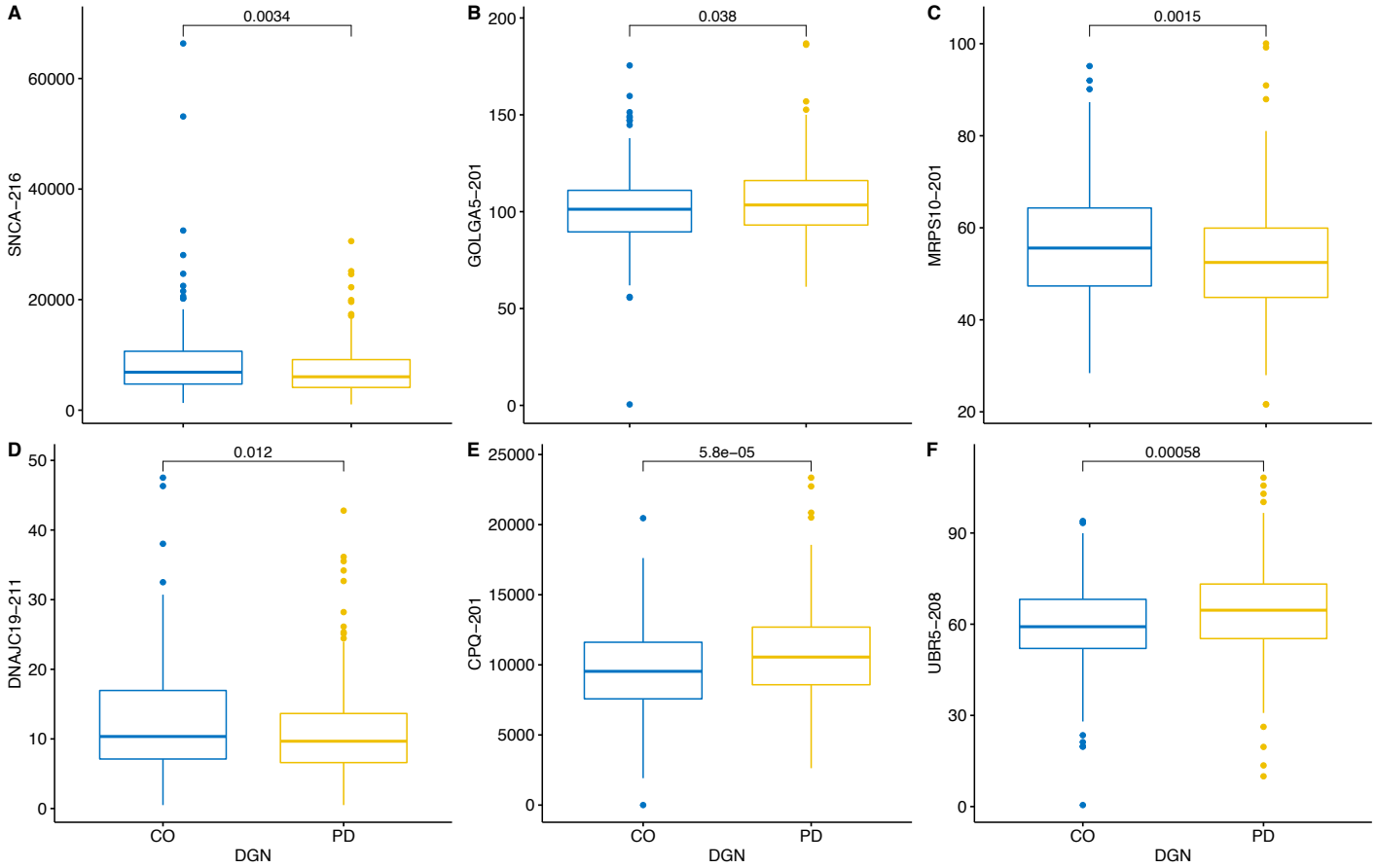


Figure 4. Cross-sectional analysis of the intronic transcripts at the baseline visit (BL), at the time of enrolment of the subjects. PD patients had decreased expression of the SNCA-216, DNAJC19-211 and MRPS10-201 transcripts. The transcripts GOLGA5-201, CPQ-201 and UBR5-208 were upregulated.

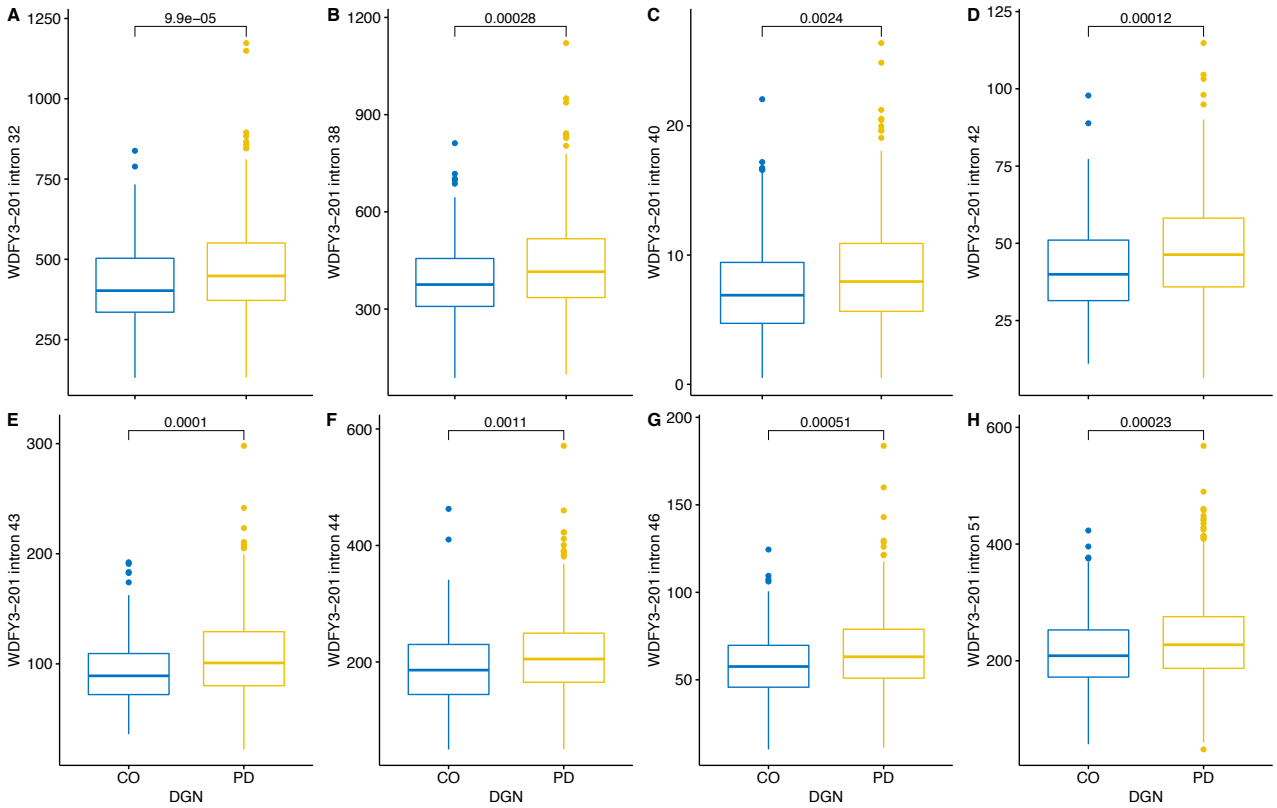


Figure 5. Cross-sectional analysis of the intronic transcripts at the baseline visit (BL), at the time of enrolment of the subjects. PD patients had increase expression of 8 introns of the WDFY3 gene.

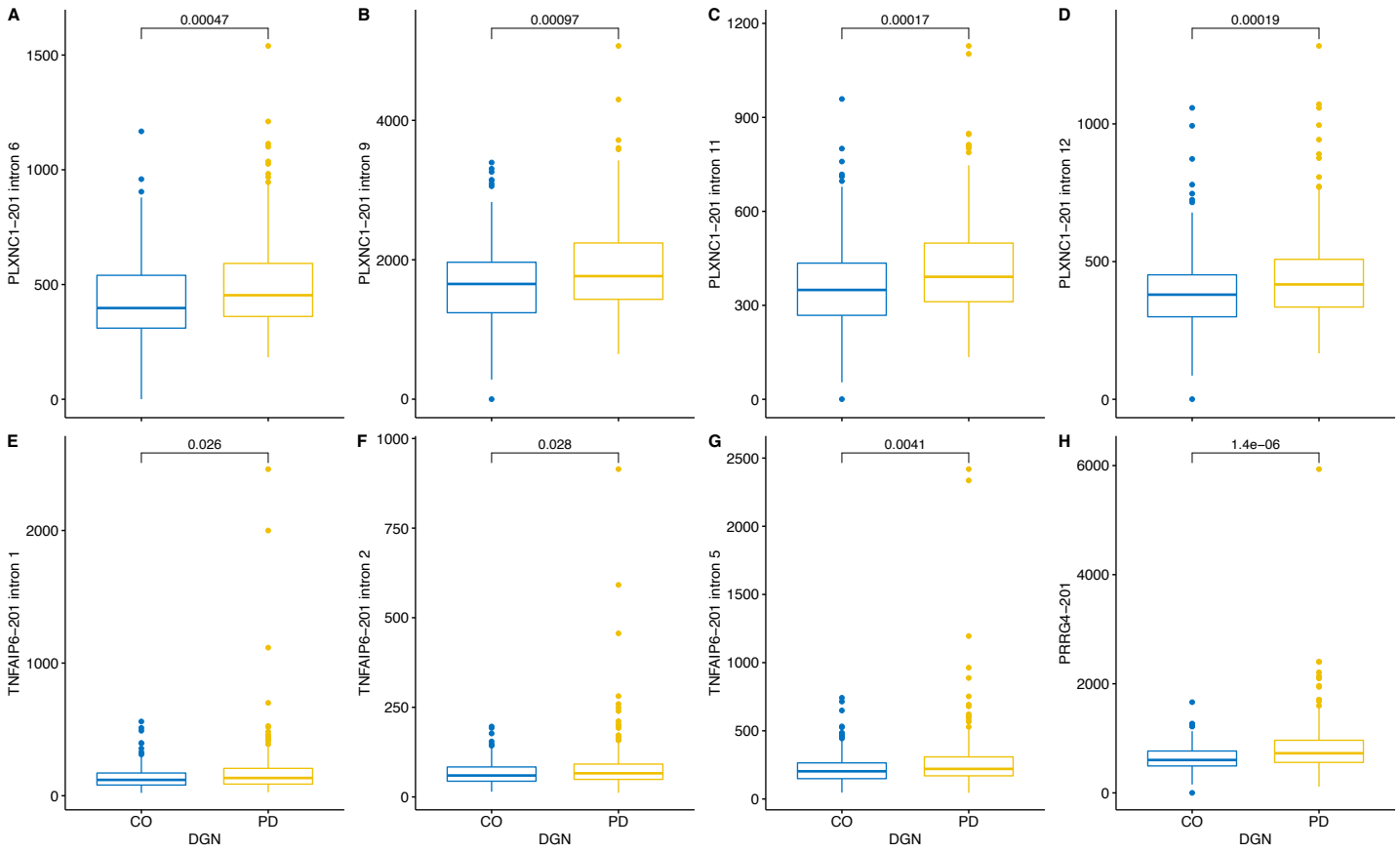


Figure 6. Cross-sectional analysis of the intronic transcripts at the baseline visit (BL), at the time of enrolment of the subjects. Different introns of the transcripts related to the immune response were up regulated in PD patients.

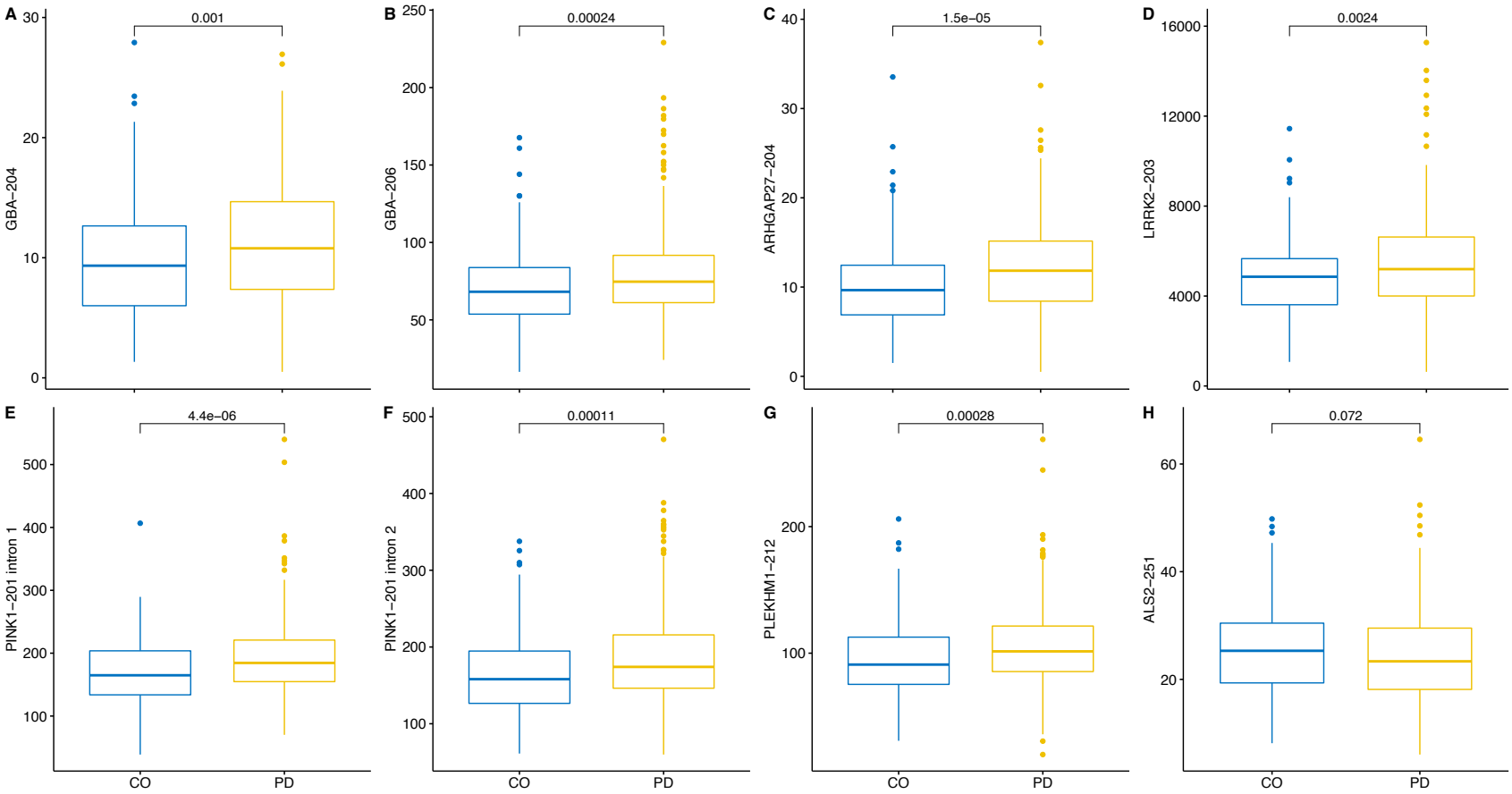


Figure 7. Cross-sectional analysis of the intronic transcripts at the V08 visit, three years after enrolment of the subjects. Two different transcripts of GBA, LRRK2-203, ARHGAP27-204 and PLEKHM1-212 were upregulated in PD patients. Intorns1 and 2 of the PINK1-201 transcript were also upregulated. ALS2-251 transcript was downregulated in PD patients.

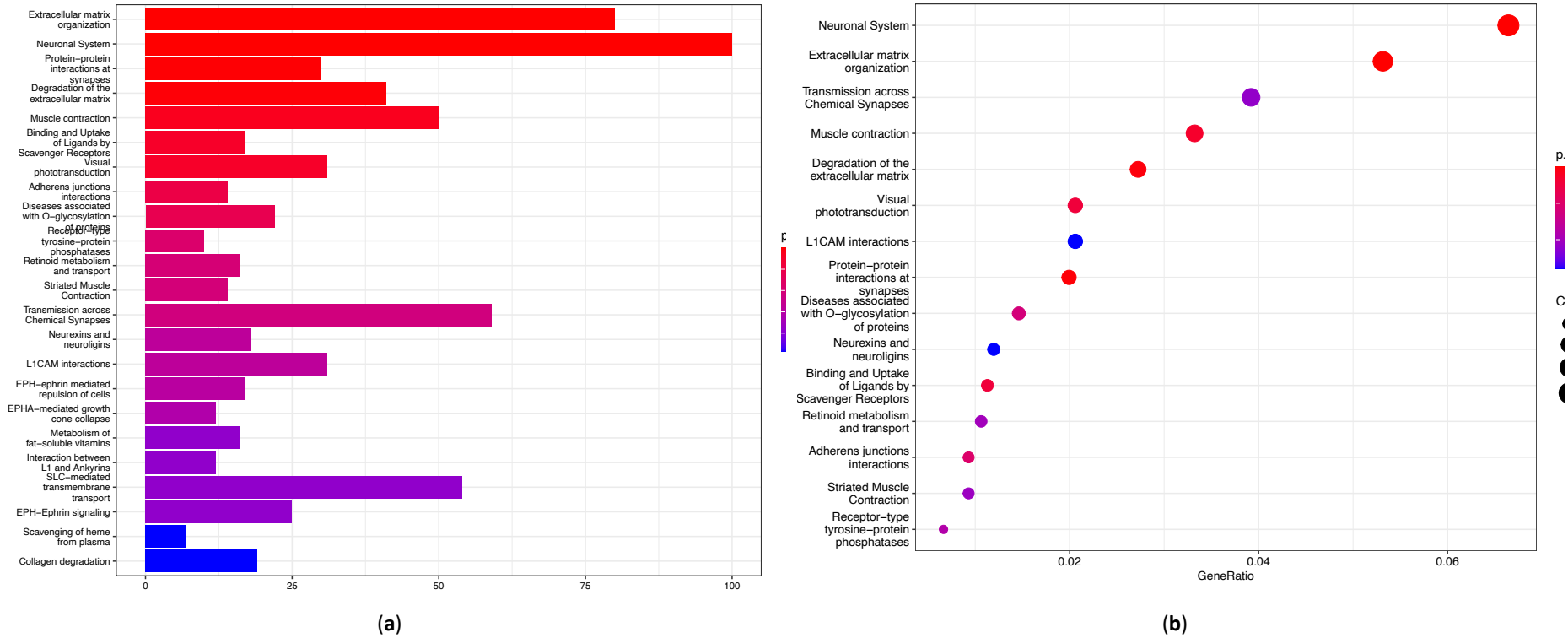


Figure 8. Reactome pathway analysis indicates longitudinal enrichment of the neuronal themes in the blood transcriptome of the PD patients. (**a**) Barplot of all statistically significant enriched pathways indicates the different number of genes for each pathway. (**b**) Dotplot of the enriched pathways shows the relation between the gene count, gene ratio and adjusted p-values. Key appears cut off on my screen

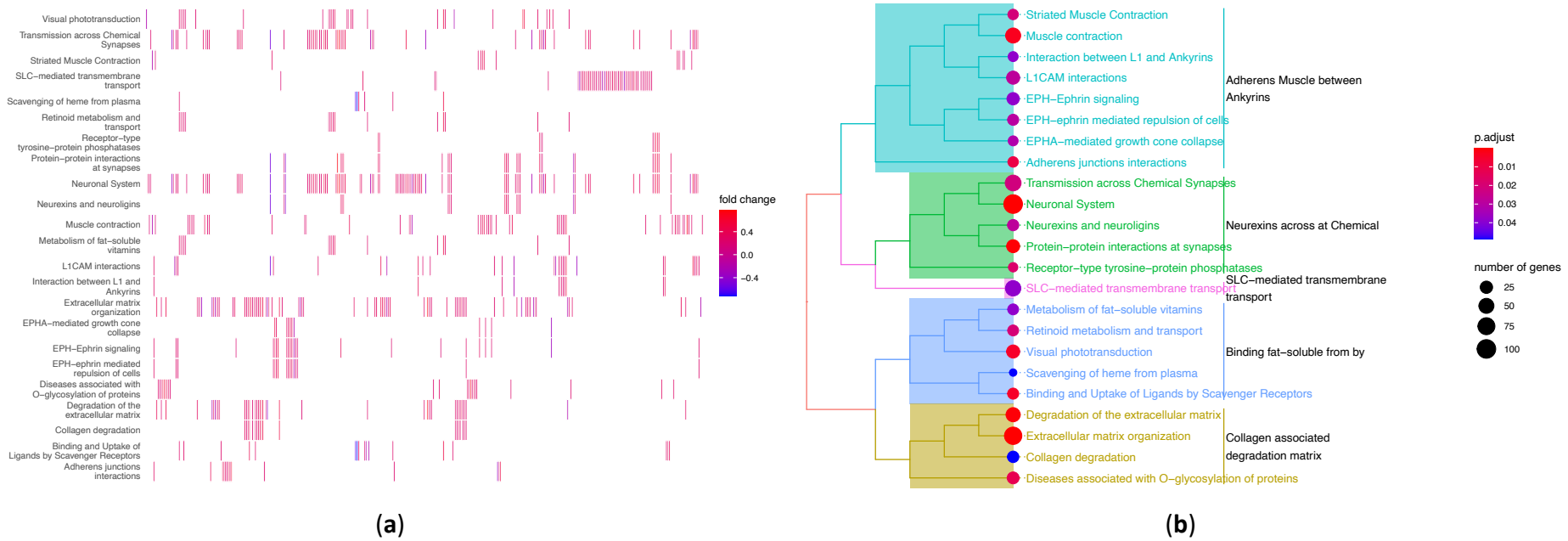


Figure 9. Reactome pathway analysis identified several neuronal pathways enriched longitudinally in PD blood transcriptome. (**a**) Heat plot showing individual genes and their mapping to pathways; (**b**) Tree plot illustrates clustering of the pathways, number of matched genes and adjusted p-values.

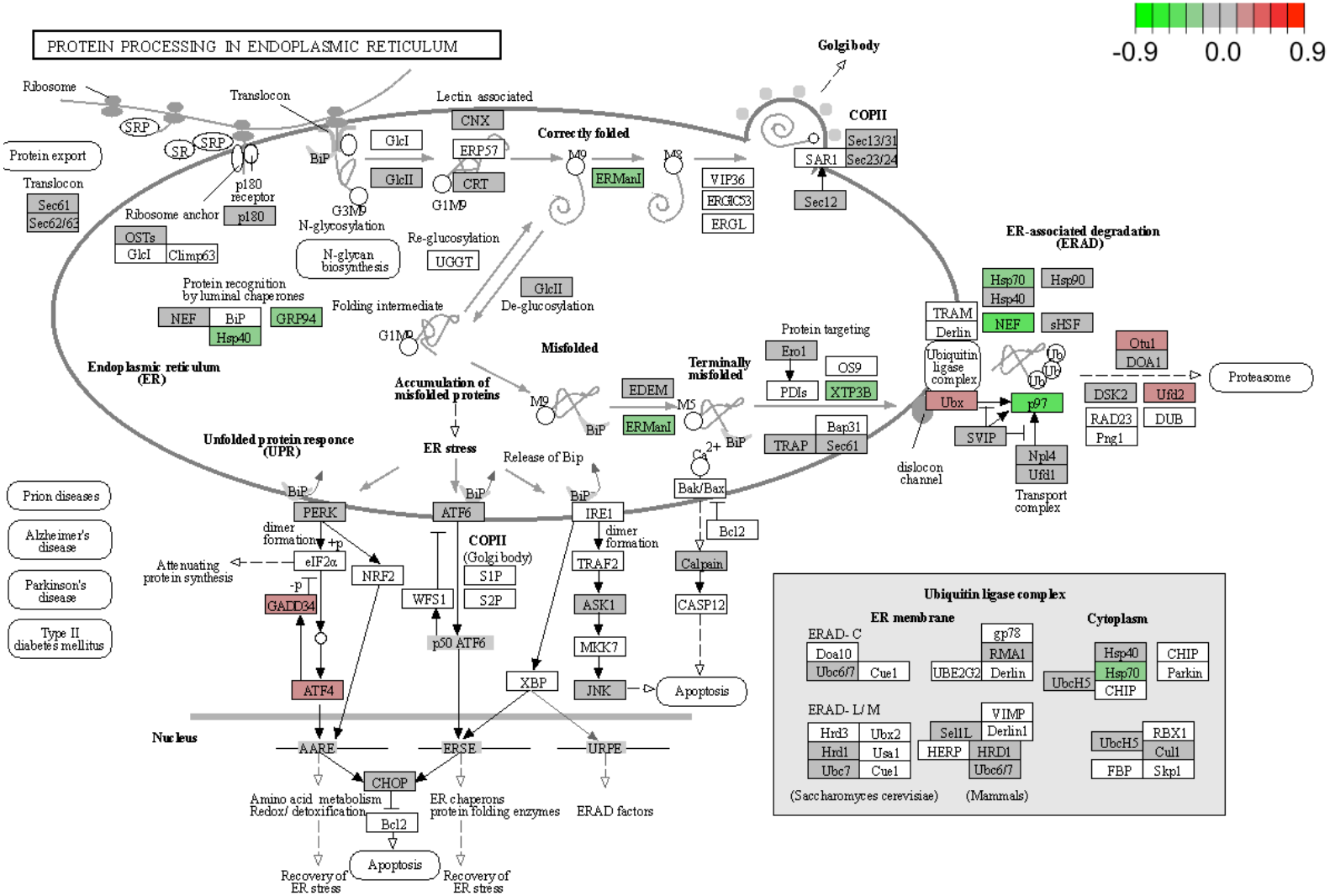


Figure 10. KEGG pathway of protein processing in the endoplasmic reticulum and the mapped genes from the intronic transcriptome. Mapped genes are in red, green, or grey colours. Indicate what colours correspond to – up /down/ unchanged

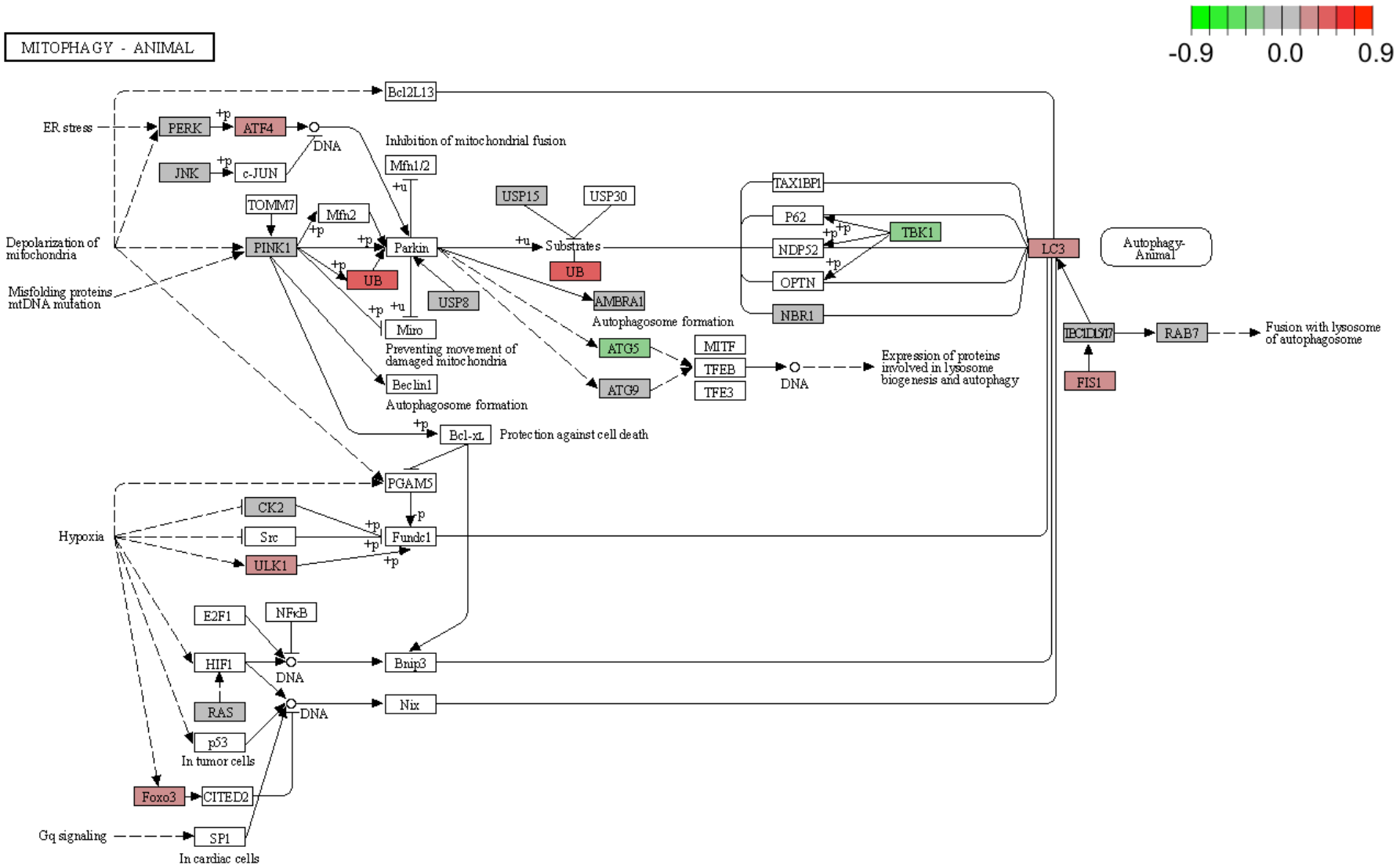


Figure 11. KEGG mitophagy pathway. Longitudinal blood intronic transcriptome profile of Parkinson disease mapped significantly to the mitophagy pathway. Mapped genes are in red, green, or grey colours.